Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Enhanced Backbone Dihedral Parameters for Protein Simulations

A Dissertation Presented

by

Carmenza Martinez

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

May 2014

Stony Brook University

The Graduate School

Carmenza Martinez

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree,

hereby recommend acceptance of this dissertation.

Carlos Simmerling – Dissertation Advisor
Professor, Department of Chemistry

David Green - Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Daniel Raleigh – Third Member
Professor, Department of Chemistry

Marivi Fernandez-Serra – Outside Member
Assistant Professor, Department of Physics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Enhanced Backbone Dihedral Parameters for Protein Simulations

by

Carmenza Martinez

Doctor of Philosophy

in

Chemistry

Stony Brook University

2014

Molecular Dynamics (MD) and Monte Carlo (MC) simulations of biomolecules have become a supportive tool to experimental methods for structure refinement, protein folding, drug discovery and many more applications. This is due to the increase in computational power and necessity of finding a method capable of describing the forces at work in chemical systems without the need to do extensive quantum calculations. All of these are possible because of the classical definition of the potential force field function.

The force field function is essential to perform MD simulations, since this equation carries the physical description of the molecules being studied. Therefore, it is imperative to have a realistic force field. One of the most popular force field functions for the study of biomolecules is AMBER ff99SB. According to web of science, as of 2013 it has been referenced more than a thousand times. This popularity results from ff99SB's substantial agreement with NMR and X-ray crystallographic experimental findings.

However, ff99SB is not perfect; there have been well-documented areas in which is in need of improvement. For example, helical propensities of transiently helical peptides are not being sampled nearly as much as suggested by experiments. The agreement between NMR observables for poly-alanine peptides and simulations needs to be further improved; as the results from simulations deviate from what is reported from experiment and agreement with

poly-alanine vicinal scalar couplings has become one of the qualities for which force fields are commonly being evaluated.

The backbone dihedral parameters were fitted to Quantum Mechanical (QM) energies of glycine and alanine tetra-peptides in AMBER ff99SB, as these tetra-peptides can helically hydrogen bond but isolate backbone effects from side-chains. Glycine and alanine populations better matched the Protein Data Bank (PDB) after this refinement. However, the dihedral maps did not match perfectly for each of the secondary structure regions of the Ramachandran map; indicating that there is still a gap between the results of QM calculations and what is observed in nature. This is because the secondary structure propensities of proteins in solution are governed by physical properties that cannot be completely reproduced by QM calculations of small homologous peptides in vacuum. Thus, it is necessary to empirically modify the energy function in order to overcome these limitations.

In this work, we present a set of modifications to ff99SB force field backbone dihedral parameters, in particular to the φ', ψ and ψ' corrections. Our goal was to further improve the agreement between results from simulations with backbone scalar couplings for poly-alanine peptides and helical propensities for transiently helical peptides. Furthermore, it would be ideal that these modifications do not alter significantly the propensities of the force field for other secondary structure regions in the dihedral map.

We generated thirty potential modifications that were tested for agreement with *J* scalar couplings for Ala_5 and Val_3. We tested whether the same force field could improve scalar couplings and helical propensities by simulating two transiently helical peptides. We additionally evaluated the behavior of the force field with β-like peptides. The stability of protein simulations was estimated by calculating NMR order parameters from proteins. One of the modifications denoted mod1φ had good agreement with most of the experimental standards tested; indicating that the changes to the energy function can be applied to simulations of systems with diverse secondary structure.

To my supportive family

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

I want to thank Dr. Carlos Simmerling for the opportunity that he gave me to join his group. It has been an honor to have him as my mentor and advisor. I have learned many things from him. I immensely appreciate his time, ideas, and funding in order to make my PhD degree a reality. His enthusiasm, patience, understanding, kindness and professionalism have inspired me, as well as many others that have had the pleasure to have him as a thesis advisor. I am also thankful for the example that he has provided me as a successful parent, professor, chemist and scientist.

I would also like to thank the members of the Simmerling group that have contributed tremendously to my professional and personal growth during my stay at Stony Brook. I could never thank them enough for all their useful discussions, advice and collaboration that allowed me to carry my work. I am especially grateful to James Maier, Dr. Lauren Wickstrom, Dr. Miranda Shang, Hai Nguyen and Dr. Amber Carr for all their contributions to my research. They were always helpful and kind to me. Their enthusiasm, intensity and willingness to help will always be greatly appreciated. I would also like to thank past members for the group such as Dr. Asim Okur and Dr. Daniel Roe for their contributions to the work presented here.

Many thanks to the members of my committee: Dr. David Green, Dr. Daniel Raleigh and Dr. Marivi Fernandez-Serra. Their time, patience, professionalism, valuable criticism and input that I have received from them is greatly appreciated. They have been kind and understanding and have helped me as much as they can to get to this point, which is also immensely valued. I would also like to mention my friends, professors, colleagues and the secretarial staff from the chemistry department for all their help, kindness and support throughout these years.

With regard to the computer resources used, I must not forget to acknowledge the computer resources at the New York Center for Computational Science (NYBLUE/L), the National Center for Supercomputing Applications (Mercury) and the National Institute of Computational Sciences (Athena and Kraken). Funding for the research presented here was

provided by the National Institutes of Health R01 Grants GM061678 and GM079383. Lastly, I would like to thank my family and my friends, in particular my beloved mother and husband for all their support, patience and help throughout the difficult times. I would not have completed all these without their immense patience, affection and encouragement. Thank you

Carmenza Martinez

Stony Brook University

May 2014

1. Introduction

1.1 Difficulties experimental techniques face when studying protein dynamics

Experimental techniques have evolved significantly to study structure and function. Recent advances in crystallography, Nuclear Magnetic Resonance (NMR), electron microscopy, neutron and fiber diffraction have elucidated more than ninety thousand protein structures deposited in the Protein Data Bank[1]. This vast set of information has been essential in the development of novel therapeutics by discovering potential drug targets as well as helping to elucidate the role and functional mechanisms of many proteins for which this information was unknown before. However, this information is static and is the average of an ensemble of structures at a given time. The information allows for inference of how the protein behaves, yet the events that preceded the given resolved structure as well as the events that followed, remain mysteries.

Methods such as Atomic Force Microscopy (AFM)[2] are capable of measuring atomistic dynamics as they occur, nevertheless the slow rate of scanning can contribute to thermal drifting making the measurement of accurate distances between images inappropriate[3]. NMR dynamic studies can provide information on the internal motions of the protein by measuring specific bond interactions. Yet the information obtained is limited to the flexibility of the regions being studied and no significant structural evidence can be extracted[4].

Other methods that measure bond dynamics, like Two Dimensional Infrared spectroscopy takes advantage of the vibrational frequencies between bonds and how they interact with each other. The method is capable of measuring dynamics in the picosecond time scale and it is also able to determine the mechanisms between states. However the information provided requires

interpretation in order to translate it into coordinates[5]. Hybrid methods such as fast relaxation imaging combine fluorescence microscopy and fast temperature jumps that allow observing how changes in temperature affect cellular dynamics. Although cellular dynamics can be observed, the detailed atomistic information cannot be obtained from these methods[6].

Despite the advances in experimental techniques to obtain protein dynamics, there is still a gap that needs to be filled between structure and atomistic mechanisms. As described previously, current experimental methods must compromise in one area or more. Therefore, alternative methods have evolved in order to study these processes that can "model" proteins' molecular mechanisms, though these modeling efforts often must use experimental structures as a "starting point" and must validate against available structural information to ensure reasonable predictions.

1.2 Challenges in the study of protein dynamics through quantum mechanical methods

Quantum mechanical (QM) methods to study protein dynamics are currently limited by the size of the system. Even though computers have evolved to handle rigorous calculations, still the amount of atomic interactions and conformations in a large system like a protein is considerably large and currently not possible. QM calculations are currently possible on small peptides with limited conformations. The information provided is useful to determine physical properties of bonds, angles, dihedral information and helpful to build force field models to study protein dynamics.

Other methods such as *ab initio* molecular dynamics were conceived in order to be able to study large molecules and to implement quantum mechanical calculations to molecules in the liquid phase[7-11]. The method combines finite temperature dynamics with forces obtained from electronic structure calculations performed "on the fly" as the MD simulation proceeds [12]. Many events such as bond breaking/forming and electronic polarization are calculated within the accuracy of the electronic structure method used. It can also be expanded to include nuclear quantum effects by using the path integral approach[13, 14]. *Ab initio* molecular dynamics and

ab initio path integral dynamics have been applied to a wide variety of problems such as liquid structure, acid-base chemistry, industrial/biological catalysis and materials[12].

Ab initio molecular dynamics have also been applied to bio-molecular simulations in order to calculate chemical shifts in drug-enzyme complexes[15], nucleic acids structure[16], design of bio-mimetics[17, 18] and dynamic studies of myoglobin[19, 20]. The results obtained from these studies previously listed were helpful to elucidate new mechanisms of function and clarified some of the experimental observations.

However, the method has limitations and one of them is the size of the system being simulated. In the case of *ab initio* molecular dynamics, the size of the system is around tens to hundreds of atoms in timescales of picoseconds [21-23]. The information obtained in these time scales is limited to bond vibrations which could be useful for comparison to Infrared (IR) spectroscopy. Nevertheless, events such as protein folding, ligand binding and domain dynamics happen in longer time scales. Furthermore, in *ab initio* simulations the accuracy of the calculation is defined by the electronic method used [12]. Currently the most popular electronic method is the Kohn-Sham formulation of Density Functional Theory (DFT) in which the electronic orbitals are extended in a plane-wave basis set[12]. This method provides a reasonable electronic depiction of chemical environments. It is not mandatory to use this electronic method for simulations, since it is possible to use more empirical or more accurate electronic structure methods [24, 25] as well as different basis sets for the calculation [26, 27]. Nevertheless, different electronic structure methods have different limitations, for instance the more accurate methods are more computationally expensive and slow. Other methods with lower levels of theory are more computationally feasible, however they are still slow; hence, the limitations given by the size of the system and the time scales obtained from the method make simulations of slow processes in proteins not practical at this moment. Other methods based on simpler classical approximations to atomistic mechanisms such as molecular dynamics simulations have been implemented for this purpose.

1.3 Molecular dynamics and Monte Carlo simulations

Proteins are one of the main components of cells and have crucial functions in biological processes and the sustenance of life. Interestingly they are only constituted of twenty amino-acids that assemble into very complex and specific structures with very distinct roles in cellular processes. Understanding how proteins assemble into these structures is imperative to determine how the association of the amino-acids present in the sequence interacts to give each protein its role and native fold. Lack of this appropriate native fold usually results in protein degradation in order to prevent damage. The type of damage that is caused by mis-folding ranges from protein aggregation to amyloid deposition and in certain cases even cell death[28]. Currently experimental methods have been developed in order to track folding intermediate states that reveal an intricate pathway to the native state from a vast range of potential alternatives[29].

Protein folding can be studied through a wide variety of experimental methods, some mentioned here such as circular dichroism, nuclear magnetic resonance, X-ray crystallography, fluorescence spectroscopy and laser induced temperature jump spectroscopy. However, these methods are unable to provide sequential trajectory of global folding events[30]. Other methods like atomic force microscopy[31] have disadvantages such as the time required to obtain an image which in principle would be an average image of the ensemble at that particular point in time. Scanning electron microscopy can be expensive and require a large area only dedicated to the instrument has the disadvantage that artifacts can be found in the preparation of the samples and these artifacts cannot always be identified or eliminated. The samples must also be frozen in order to be studied[32].

Molecular dynamic simulations can supplement the results obtained from experimental methods by presenting a sequential trajectory of motion of the ensembles. This is possible by approximating the atomic motions in molecules to Newtonian equations of motion and implementing these into computational algorithms that allow for these types of calculations[33].

Molecular dynamic studies have provided breakthrough in the field of protein folding such as the all-atom prediction of the folded state of a stable mini-protein[34]. They have also successfully reported a sequential course of protein motions that could not be obtained from experimental methods alone[35]. Other studies have been reported in detail about the reversible

folding and free energy surfaces of two designed β-sheets in implicit solvent based on accessible surface area[36, 37] . The reported simulation time was in the order of several tenths of microseconds and the mini-proteins were reported to satisfy most of the NOE distance restraints seen from experiment. Although both proteins appear to follow different folding paths, the authors concluded that the amino-acid sequence and the specific interactions between different side-chains determine the most plausible folding pathway [37]. MD simulations have also been implemented to study anomalous folding paths as in the case of amyloid formation where proteins mis-fold and aggregate. This effect in proteins and peptides alter their proper configuration and interactions causing formation of insoluble fibrils. These insoluble fibrils accumulate causing amyloidosis and have been linked to more than 20 human diseases among them type II diabetes. Neurodegenerative disorders such as Alzheimer's and Parkinson's disease have also been linked to amyloid formation. Among some of the recent MD simulations reported the work by Berhanu et al. [38] described explicit solvent simulations of cylindrin β-barrel amyloid, a common motif in amyloid fibril structures. The authors investigated the effect of mutations of the hydrophobic core by replacing Val_2 with other residues with different hydrophobic characteristics such as Leucine, Alanine and Glycine. The effect given by strong hydrophobic interactions from the Leucine mutation kept the strands in close proximity, maintaining the side-chain interactions, main-chain hydrogen bonds and the salt bridges along the weak out of register surface. The authors also identified water mediated interactions at the center of the hydrophobic cavity. Besides this handful of reports indicating contributions of MD simulations to the field of protein folding, there is a vast repertoire of publications in which folding and unfolding simulations of peptides and proteins have been reported [39-64]. These reports give proof that molecular dynamics can be a very useful tool in providing accurate statistical and analytical results that can help interpret some of the results obtained from traditional experimental methods for protein folding.

Molecular dynamics (MD) and Monte Carlo (MC) simulations have increased in popularity as computational power has increased. This is partially because other methods such as *ab initio* calculations require extensive quantum mechanical calculations of wave functions for each atom. As the size of molecules increases the amount of calculations increases as well, likewise requiring more computer power. Even with the increased speed and memory power in modern computers, *ab initio* calculations still remain computationally expensive and time

consuming.

On the other hand, the implementation of molecular mechanics to macromolecules like proteins and DNA has become routine. Proteins and DNA are of particular interest because of their role in cell biology and disease, making them preferred targets for drug development and novel therapeutics [65-68]. Thus, MD and MC simulations have increasingly played a significant role in the interpretation of experiments, likewise serving as a tool to complement the results obtained. This can be appreciated in the growing application of MD/MC dynamics to structure refinement of NMR [69]and X-ray crystallographic data [70] as well as structure prediction [71].

At the moment MD/MC simulations in the nanosecond time scale are quite feasible allowing for direct assessment of the quality of simulations against hetero-nuclear NMR spin relaxation [72]. In certain cases, other NMR observables such as chemical shifts, scalar couplings and residual dipolar couplings (RDC) can be detected at this time range as well [69]. Likewise, certain protein motions have longer time scales that require longer simulation time in order to be sampled [73].

As faster algorithms and hardware for computer simulations are being developed, μs/ms molecular dynamic simulations have become more achievable and customary [45, 50, 54, 73, 74]. Nevertheless, enhanced sampling methods are still a popular choice to study events such as protein folding and chemical processes at higher time scales [46, 64, 75, 76].

1.4 Challenges of molecular dynamic simulations

Although molecular dynamic simulations can be useful, they also have drawbacks like experimental methods. Comprehensive molecular dynamic simulations require high amounts of computational effort in order to perform all the required calculations between the atoms of the system. In order to overcome this, certain approximations have been implemented in order to obtain accurate results at a lower computational cost such as the treatment of protein-solvent interactions[9]. These approximations have been successful at accurately describing the folded state of a mini-protein[34] as well as sampling different folding events in another mini-protein system[77].

One of the core aspects of molecular dynamic simulations is the description of the potential energy of a system of particles in terms of its coordinates using classical mechanics. This description is also known as a force field function. This calculation is not as rigorous as quantum mechanical calculations; because the atomic interactions are decoupled into individual, classical bond stretches/bends, dihedrals, electrostatic and non-bonded terms versus complex atomic wave-functions. Computer power and resources have increased making it possible to carry out simulations of large biomolecules such as proteins and DNA that were not possible decades ago. Therefore, the use of molecular dynamic simulations of biomolecules has become increasingly popular in the fields of cell biology [78-81] and drug discovery [82-84].

Nevertheless, force fields have limitations, as in the case of fixed-charge models; where charges are described as static parameters and electronic polarization is not included. Although there have been attempts to overcome this problem [85-87] by introducing charges and dipoles that are allowed to change as the environment varies. However these models have been reported to have issues with transferability and standardization [88].

Force fields assume a pre-defined connectivity that limits its ability to accurately describe bond forming and breaking events and do not include spin properties. Besides, force fields are usually parameterized by empirically fitting physical properties of small hydrocarbons and peptides from quantum mechanical (QM) calculations. The peptides used for the fitting are small which in principle assumes the notion that the empirical corrections can be the same for both types of systems. Furthermore, parameters for the backbone dihedral terms are usually derived from QM calculations of small alanine peptides in order to make calculations more feasible since there are no side-chain effects on the backbone parameters. This method assumes that backbone properties for alanine are representative of the other amino-acids with the exception of glycine. Therefore, the parameterization of force fields can sometimes lead to undesirable biasing towards given states. Recently, it has become common to fit against NMR of peptides and full-length proteins such as in the case of ff99SB* [89], ff99SB-CS[90] and CHARMM36 [91] in order to overcome these issues.

1.5 Mini-proteins and peptides as model systems for protein folding

Mini-proteins and peptides can be quite useful to study protein folding because they can provide a simple overview of protein folding at a smaller scale that could be applied to larger systems [92, 93]. Sometimes they represent domains within larger proteins and usually they are fast folders [94, 95] in the μs time scale. Studying these small systems can provide insights in the conformational pathways required for protein folding. Some of these domains hold the main activity of the protein like in the case of the Villin headpiece (HP-67) which represents the smallest fragment of the enzyme that retains its F-actin binding activity[30]. A possible disadvantage of using mini-proteins and peptides as model systems is that sometimes their folding behavior undergoes intermediate and transition states defined as non-cooperative behavior, while is known that larger proteins fold in a cooperative manner[96]. However, there have also been reported cases of mini-proteins that display cooperative folding such as the Villin headpiece[97], the Trpcage[92] and the BBA motif[98]. Nevertheless, it is much easier to study protein folding through mini-protein and peptide models than through larger proteins, because the folding time for the proteins is much longer than for peptides/mini-proteins and also these small model systems provide the ability to quantitatively test simulation methods as well as serving like test models for folding theory.

In this work the fast-folding properties of peptides are implemented in order to benchmark the data for the proposed modifications. These folding mechanisms can provide insights as to how simulations of proteins would behave under the same conditions and how the modifications would affect their secondary structure propensities.

1.6 The importance of the force field function for molecular dynamic simulations

1.6.1 Background

One of the core aspects of molecular dynamic simulations is the description of the potential energy of a system of particles in terms of its coordinates using classical mechanics. This description is also known as a potential force field function. This calculation is not as

demanding as a quantum mechanical calculation; however as the size of the system increases the calculation between its atoms also increases. This is because the interactions between the atoms such as bonds, angles, dihedrals, electrostatic and van der Waals energies are computed classically, making it more intense for the computer to evaluate the force field function. Furthermore, computer power and resources have also increased making it possible to carry out simulations of large molecules such as proteins and DNA to extents that were not possible decades ago. Therefore, the use of molecular dynamic simulations of biomolecules has become increasingly popular in the fields of cell biology [35, 80, 99-102] and drug discovery [81-84, 101].

There is a wide variety of force fields available for simulations of molecules ranging from hydrocarbons and small organic molecules to macromolecules found in biological systems. Initially force fields were developed in order to simulate small molecules such as hydrocarbons, among these force fields we encounter Molecular Mechanics 2 (MM2) developed by Allinger et al. [103] mainly to study covalent bonds and equilibrium energies. Recent revised versions such as the MM3 force field [104] were designed to better reproduce heats of formation, conformational energies and rotational barriers; and the MM4 force field in which additional interaction terms were added to the force constant matrix to have better agreement in the calculation of physical properties of alcohols, ethers, carbohydrates and related compounds. Among the first force fields developed, the Consistent Force Field (CFF) developed by Warshel et al.[105] allowed for studying structure and vibrational energies of molecules and crystals. This method introduced the concept of Cartesian coordinates for the description of atoms in space, which was further developed and implemented to other subsequent force fields.

The work by Scheraga and co-workers known as the ECEPP[106-115] force field also provided novel parameters for the simulation of proteins at the time. This force field was the first one developed for simulations of biomolecules and some of the parameters were borrowed for the development of the most popular force fields being used today. It used fixed geometries for amino-acids in order to facilitate the potential energy calculation using energy minimization in the space of protein torsion angles. Similarly to the MM2 force field potential energy functions are included for hydrogen and single bond rotations.

Following on this initial work, other types of force fields followed, such as polarizable

force fields based on electronic structural theory, induce dipole, point charges, distributed multi-poles, density and Bond Polarization Theory[116]. The development of these force fields has been of particular interest due to the intrinsic polarity of biomolecules. Nevertheless, classical additive force fields still remain more popular at modeling biomolecules[116].

Other classical additive force fields such as CHARMM [117], GROMOS [118] and in particular AMBER [119] are popular for molecular dynamics simulations of biomolecules. These force fields are also called empirical because the derivation of the parameters presented is based on numerical approximations to *ab initio* calculations of small molecules and experimental results[116].

1.6.2 The AMBER Molecular Mechanics Force Field Function

$$E_{total} = E_{bonded} + E_{nonbonded}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

$$E_{nonbonded} = E_{electrostatics} + E_{vanderWaals}$$

$$V^{r^N} = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a\,(\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)]$$
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] \right\} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Equation 1 AMBER molecular mechanics force field function. The bond angles and torsion terms are considered the bonded terms, whereas the van der Waals and electrostatic terms are considered the non-bonded terms. The bonds and angles are described mathematically as harmonic potentials whereas the torsional potentials are described as a Fourier series of cosine functions. The torsional potential was implemented from the rotational torsional energy barrier function derived by Pitzer et al. for ethane [120]. In the case of the non-bonded terms the van der Waals terms are expressed as Lennard-Jones potentials [121, 122] and the electrostatic energies are expressed as Coulomb terms.

The family of AMBER force fields is primarily constituted by bonded and non-bonded terms. The bonded terms comprises the bond distances, bond angles and dihedral terms of the bonded atoms. The non-bonded term includes the Lennard-Jones potential and the Coulomb

term. The Lennard-Jones potential approximates the interactions between neutral atoms in a pairwise manner, and the Coulomb term for the non-neutral atoms in the same manner. Thus, the non-bonded term of the force field calculates the van der Waals and electrostatic interactions of the particles.

In order to obtain dynamics information from this equation, the potential energy is calculated from the relation:

$$F = -\frac{dU}{dx}$$

Equation 2 Force equation

From the force relationship we can obtain coordinates as a function of time:

$$\frac{d^2x}{dt^2} = \frac{F_x}{m}$$

Equation 3 Newton's second law of motion

The AMBER force field function includes improper dihedral terms within the torsions whereas other empirical force field functions such as CHARMM (Chemistry at Harvard Molecular Mechanics)[117] have separate terms for the improper angles as well an extra potential for the 1-3 interactions known as the Urey-Bradley potential. However the Urey-Bradley potential is only implemented at discretion of the user and not by default in the calculation. The GROMOS96 (Groningen Molecular simulation) [118] force field also has another term known as the Ryckaert Bellemans potential that is implemented for the simulation of long alkanes and polymers.

1.6.3 History of AMBER force field functions prior to the ff99SB era

AMBER stands for Assisted Model Building and Energy Refinement and covers a set of force field functions and a molecular dynamics package that have been developed over the last three decades. In the beginning the molecular dynamics package was designed to simulate small molecules as well as polymers[123, 124], however with the passage of time it has evolved to be

one of the most popular packages for simulations of biomolecules[125]. The initial AMBER force field functions were conceived partially due to the work published earlier by Momany /Nemethy et al. [109-113] on the ECEPP potentials and the work of Lifson et al. on the Consistent Force Field (CFF) [105, 126, 127]. Other previous efforts in the development of potential energy functions for simulations of organic molecules contributed as well[128].

At the initial stages of development, the parameterization of force fields was considerably limited to the computer power needed for the simulations. Taking into account every single atom in the biomolecules would have resulted into rigorous calculations, in particular in the case of hydrogen. Therefore the work by Weiner et al.[129] presented an alternative to this issue by treating the hydrogens bound carbons as united atoms with the forces at the centers and explicitly considering the polar hydrogens. The charges included in this work were taken from high level quantum chemistry calculations[116], the van der Waals parameters were borrowed from the work published by Lifson et al. [126, 127] and Jorgensen et al. [130]. The force constants, bond lengths and angles were taken from crystal structures and implemented to match normal mode frequencies of small peptides [116]. The torsion force constants were fit to quantum calculations or experimental data in order to better reproduce torsional barriers of small hydrocarbons. Unfortunately this approach was soon refuted due to the inability to represent effects present in certain chemical groups crucial for interactions and appropriate geometry[131, 132]. Mainly, accurate charge distribution in aromatic rings, the correct pseudo-rotational distribution for five-membered aliphatic rings and the direct comparison between calculated and experimental vibrational frequencies [131, 132]. Later on, the work published by Weiner et al.[129] was extended to an all atom force field based on gas phase simulations in 1986 [133]. As computer power increased so did the efforts to produce a more accurate force field for the simulations of biomolecules. Such efforts gave rise to AMBER ff94 force field from which most of the current AMBER force fields have evolved[134].

AMBER ff94[134] provided a novel approach to simulations by introducing parameters for all atom simulations of proteins, nucleic acids and organic molecules. Some of the features presented included a revision to the work described by Weiner et al. [129] for the bonded terms to better reproduce the vibrational frequencies and structures of molecules. The non-bonded terms were adjusted as well; the carbon and aliphatic hydrogen van der Waals terms were

adjusted to better reproduced the enthalpies of vaporization of liquid alkanes and benzene. Other terms such as the sp^2 and sp^3 hybridized nitrogen, sp^2 hybridized oxygen, ether/ester oxygen and hydroxyl groups were taken from the OPLS model [130, 135]. For the electrostatic terms, they introduced a revised version of the Restrained Electrostatic potential (RESP) charge model that corrected for the under-estimation of charges in buried atoms [136, 137]. The dihedral parameters were also refined by improving the fitting to gas phase quantum chemistry calculations of low energy conformations of glycine and alanine. This type of fitting is challenging because the model system of choice for the quantum chemistry calculations is dipeptides which lack the non-local electronic distributions that can be observed in larger biopolymers [138]. Furthermore, the gas phase calculations do not represent well the polarizing effects observed in solvent [139]. Nevertheless, these corrections made ff94 a popular choice for simulations even until the present day; however there have been reports that the force field over-stabilizes α-helical propensities [140-143].

Subsequent revisions to the AMBER ff94 force field such as the ff96 revision by Kollman et al. [144] attempted to improve the torsional parameters by modifying the φ and ψ terms to better match dihedral energy profiles for alanine dipeptide and tetra-peptide. One of the aims was to improve the agreement with the energy differences between the native and extended α-helical conformations, since this peptide has enough residues to match a helical turn. In fact the dihedral profiles between the two peptides were qualitatively different in comparison to the profiles obtained from ff94, although it was difficult to quantitatively determine the improvements between the two force fields[116]. Contrary to the effect in secondary structure propensities observed in ff994, ff96 appeared to bias β-conformations.

Another revision termed ff99 used the RESP charge model to generate conformations for the alanine tetra-peptide that were included in the fitting of the dihedral parameters along with the existing conformations for the alanine dipeptide. Similarly to the ff96 revision, a bias was introduced to the function; however, the bias favored the α-helical conformation instead of the β-conformations[140]. Following up on the work of ff96 and ff99, Garcia et al. released a force field function in which some of the φ/ψ dihedral torsional terms given by ff94 were zeroed. They observed better agreement with experimental helix-coil results, however the dihedral energy profiles for glycine and alanine proved worse than in the ff94 case[142]. Another correction to

the ff99 parameters known as ff99φ by Sorin et al.[145], modified the torsional φ parameters with the parameters of ff94. This modification showed dihedral energy surfaces comparable to its predecessor ff94[142]. Another force field in which the ff99 φ/ψ torsional parameters were modified was reported by Okur et al. [140]. They termed this force field as ffGA and it was applied to simulations of α-lac 101-111[146] and Tripzip2[147] peptides. They observed that the results from simulations had better agreement with experimental findings in comparison to the results obtained with ff99. Although the authors recommended caution when using this force field unless the fitting and testing of the parameters is done in a wider range of peptide structures with diverse secondary structure. This is because the fitting for this force field was done using a "decoy" database of mis-folded structures from Tripzip2 and α-lac 101-111. In principle the idea of using this database was to have wide variety of conformations from these two peptides where the fitting of the parameters would not be biased towards a particular secondary structure state [140]. Another modification to the ff99 φ/ψ torsional parameters was used in the simulations that predicted the native state of Trpcage [34]. These modifications were implemented in order to improve the agreement with quantum energies for the alanine tetra-peptide. Further simulations using this set of parameters have not been reported since then. However, the fitting to the quantum energies for the alanine tetra-peptide was used for the parameterization of ff99SB.

1.6.4 Definition of torsional terms in AMBER force fields and development of ff99SB torsional parameters

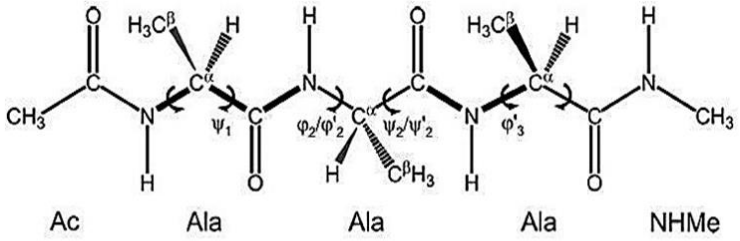
The dihedral torsional term for AMBER force fields is shown in equation, where the dihedral force constant or the amplitude of the cosine wave is the k or V_n term, n is the dihedral periodicity, γ_n is the phase of the dihedral angle (φ or ψ) and ϕ is the dihedral angle of interest φ or ψ.

$$\sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \gamma_n))$$

Equation 4 Dihedral Torsional Term for AMBER force fields, where V_n is the dihedral force constant (amplitude of the cosine wave), φ is the dihedral angle of interest and γ is the phase of the angle

When the torsional parameter terms for ff94 were conceived, an additional set of backbone dihedral terms were added to the existing terms for AMBER force fields. This set of terms was designed to address the torsional effects that amino-acids with β carbons have in the

backbone conformations, in the case of ff94 was alanine[142]. Originally, the dihedral term were constituted by φ/ψ backbone, side-chain and improper dihedral angles. Unlike other classical force fields such as CHARMM[117] and GROMOS [118], the improper dihedrals are part of the torsional term in AMBER force fields and are not excluded into a separate term. This additional set of backbone dihedrals includes the β-carbon present in most amino-acids with the exception of glycine as shown in figure 10:



All amino-acids
$\varphi = C\text{-}N\text{-}C^{\alpha}\text{-}C$
$\psi = N\text{-}C^{\alpha}\text{-}C\text{-}N$

Non-Glycine
$\varphi' = C\text{-}N\text{-}C^{\alpha}\text{-}C^{\beta}$
$\psi' = C^{\beta}\text{-}C^{\alpha}\text{-}C\text{-}N$

Figure 1 Scheme for the backbone dihedral definitions of AMBER force fields taken from Hornak et al. [142].

Since the dihedral energies in AMBER are additive, the total backbone dihedral energy is the sum of the general, specific and improper dihedral terms described in figure 10. Besides the difference in the four atoms definition, the φ/ψ and φ'/ψ' angle definitions differ by the "offset" ~120º between the dihedral planes. This definition for the backbone dihedrals around φ and ψ is exclusive to AMBER force fields. In other words, other force fields such as CHARMM and GROMOS do not have this "extra" set of dihedrals when defining φ and ψ, however the conformational effects of amino-acids with β carbons in the backbone dihedrals are not negligible and influence the propensities of the force field[142]. This effect is certainly not exclusive of AMBER force fields, other force fields such as CHARMM have accounted for this effect by adding corrective terms to its energy function in order to improve the dihedral maps of alanine[148]. In the case of ff99SB the φ/ψ and φ'/ψ' energy terms were refined to better match the dihedral profiles for glycine and alanine according the protein databank survey [149].

From the time in which ff94 was published until the development of the ff99SB[142] torsional parameters, all the revisions previously discussed were only applied to the general φ/ψ energy terms, leaving the φ'/ψ' terms identical to the ff94 parameters. As previously described, ff96, ff99 and the Simmerling et al. [150] revisions used quantum calculations to revise the torsional parameters based on energy profiles for alanine dipeptide and tetra-peptides. The other revisions were applied empirically to the existing parameters for ff99 and ff94 [140, 141, 145]. The ff99SB force field corrected the α-helical preference shown by ff94 by fitting the energies of numerous conformations of blocked tetra-glycine and tetra-alanine peptides. The local minima on the energy surface were optimized through random search in φ/ψ dihedral space followed by local minimization. The conformations for the local minima were optimized with high level of theory *ab initio* calculations obtaining 28 conformers for glycine tetra-peptide and 51 conformers for alanine tetra-peptide[142]. The energies were calculated using the approach of Beachy et al. [138] in the gas phase, as it was demonstrated that electron correlation calculations are necessary for the accurate description of relative energies for conformations in peptides [138]. These conformers were optimized using a 12-dimensional φ/ψ dihedral space search using the amplitudes (three amplitudes) and phases (three phases) of the torsional function. The amplitudes were systematically varied from 0 to 2 kcal/mol and the phases were varied between 0 to π yielding over 100 million grid points. As described by Hornak et al. [142] the zero energy point for the molecular mechanics (MM) energy function is arbitrary, therefore the optimization is best done using pair energy differences (QM/MM) for alternate conformations while using each conformer's energy set as the reference points. In ff99 they ignored this by setting the lowest energy as the "zero" energy point and calculated energy differences based on that "zero" energy point. However, this method is not ideal because the "zero" energy point is arbitrary and the results of the optimization will change depending upon the definition of the "zero" energy point. Therefore, Hornak et al. [142] described the function that was optimized as the average between the quantum mechanics (QM) and molecular mechanics (MM) pair energy differences with each conformer's energy set as a "reference". From this procedure the average absolute error (AAE) was obtained. Besides this definition, they also used another function for optimization where all individual QM and MM pair energy differences were calculated; again, with each conformers energy set as a "reference zero energy". From this procedure the maximum absolute error was obtained. As mentioned by Hornak et al. [142] the maximum absolute error is more sensitive to

different parameter sets than the average absolute error because there were several sets of dihedral parameters that had comparable average absolute error values. Therefore, the choice of parameters was based on the results obtained for the maximum absolute error calculation. In principle the maximum absolute energy errors correlate to the average absolute energy error, therefore low values for the maximum absolute error will also have low values for the average absolute error[142].

For the optimization of the glycine tetra-peptide, the first phase of the φ angle was zeroed in order to have rational dihedral functions around φ angle of zero since in this region there was not QM data available. This procedure was also applied to the 51 conformers of the alanine tetra-peptide from which the φ'/ψ' dihedral parameters were obtained. Although the fitting seemed to give reasonable results, after careful analysis it became known that one of the conformers contributed a significant amount of errors to the fitting, therefore it was removed from the optimization[142].

This correction improved the agreement between poly-alanines and the Protein Data Bank rotamer distribution as well as yielded better agreement with NMR order parameters for proteins like Ubiquitin and Lysozyme. Furthermore, this force field has become one of the most popular ones among the scientific community for its ability to quantitatively and qualitatively reproduce the results seen in experiments [69, 72, 151]. These features have made ff99SB a force field with major impact in the computational biology field contributing to the increase in popularity of the MD/MC simulations by experimentalists[69, 72]. Nevertheless, there have been many recent developments attempting to further improve the quality of the force field backbone dihedral parameters by several computational groups and experimentalists as described below.

1.6.5 Modifications to the ff99SB backbone dihedral function

Initially Robert Best et al.[152] reported that the force field could benefit from improvement in the agreement between experiments and simulations, in particular the description of poly-alanine ppII dihedral propensities. His claim was that ff99SB was one of the worst force fields at reproducing scalar coupling values for the Ala_5 peptide[152]. The deviations between simulations and experiments according to his claims were considerably significant in comparison to force fields like OPLS-AA/L, CHARMM-CMAP and some variants of the

GROMOS force fields[152]. After these claims, Robert Best issued a correction for his statement regarding the performance of ff99SB and ff03 in which he declared that the simulation protocol leading to the published results was faulty which prompted him to run another simulation for which he reported revised results that had comparable performance to the other force fields previously discussed [153]. Following up on this debate, Wickstrom et al.[151] revisited the simulations of Ala_5 with different water models for which better agreement than the one obtained by Best et al.[153] was reported and concluded that the results are dependent upon the Karplus parameters used. Nevertheless, both accounts agreed on the fact that AMBER force fields have had issues with helical preferences in the past and it appears that ff99SB while improved is not without error in this case. Both reports suggest that further investigation of the propensities of the force is required.

After these reports were published, optimizations to the ff99SB force field have been reported [89, 90, 154-158]. Some of these optimizations have been dedicated to the improvement of the backbone dihedral parameters while others have refreshed the side-chain torsion parameters[158] as well as the van der Waals interactions[157]. The initial goal of the backbone optimizations was to improve the conformational preference of the force field for helices since it had been reported that the force field had poor agreement with experiments at reproducing secondary structure for helix-forming peptides and proteins[89, 152]. One of these optimized parameter sets was reported by Best et al.[89] who presented modifications to the backbone dihedral potential for ff99SB and ff03 which they denominated ff99SB* and ff03*. Their modification was based on including a cosine correction term to the ψ dihedral term which is applicable for amino-acids without β carbons (glycine); these corrections have an angle offset; following a similar philosophy behind the definition of the ψ' dihedral angle [89] as indicated in figure 10.

Table 1 Optimized corrections to ff99SB and ff03 that originated ff99SB* and ff03*

Parameter	ff99SB	ff99SB*	ff03*
k_ψ (kcal/mol)	0.2	0.1788	0.3575
δ_ψ (deg)	180.0	105.4	285.5

Table 1 Optimized corrections to ff99SB and ff03 that gave rise to ff99SB* and ff03*. The parameters refer to a cosine correction to ψ given by $V_1(\psi; k_\psi, \delta_\psi) = k_\psi[1 + \cos(\psi - \delta_\psi)]$

Their recent publications point to the preference of the ff03* correction over the ff99SB* correction; according to their reports the helical propensities for helical systems significantly improve with both force fields. In other words, they increase with the ff99SB* revisions and decrease with the ff03* revisions; nevertheless, hairpins with the revised ff99SB* parameters de-stabilize the intrinsic native populations obtained with ff99SB whereas the populations obtained with the ff03* have better agreement than ff03 and ff99SB simulations [159-161].

Recently, Nerenberg et al. [154] presented their comparisons for the performance of ff99SB on TIP3P and TIP4Pew explicit solvent models. Primarily the aim was to show that the TIP4Pew explicit solvent model is more appropriate in order to obtain accurate conformational ensembles than TIP3P. Nevertheless, they derived a perturbation of the φ' backbone dihedral potential of ff99SB that adjusts the β-ppII secondary structure equilibrium[162]. As shown in table 1-2, their modification was based on decreasing the value for the second term in the amplitude of the φ'.

Table 2 Optimized corrections for ff99SB-opt φ'

Parameter V_2	**ff99SB**	**ff99SB-opt φ'**
k_φ(kcal/mol)	2.0	1.8

Table 2 Optimized corrections for ff99SB-opt φ'

This discrete modification showed improvement, with results showing better agreement between calculated scalar couplings from simulations and experiment for Ala_5 [154]. Although the results obtained for the NMR order parameter calculation was comparable to ff99SB[154], they did not test secondary structure conformational preferences for α-helix and β-sheet as Best et al. had done. Their justification was that this revision was designed to improve the accuracy of the intrinsic conformational preferences of single amino acids, and not the secondary structure propensities of the force field. Therefore, the exploration of the secondary structure behavior of this force field in helical peptides was not further pursued.

The quest for optimization of ff99SB has also drawn the attention of NMR experimentalists, mainly Li et al.[155] who initially modified the ff99SB backbone dihedral potential by introducing a Gaussian function that couples the dihedral functions of φ and ψ terms.

This modification was introduced based on a similar procedure done for the CHARMM-CMAP correction[163], however this iterative refinement was based on agreement with chemical shifts (CS); and later, another refinement was implemented following the same philosophy while including residual dipolar couplings (RDC) as a benchmark as well[90]. This revision showed marginal improvement for the agreement of folded proteins by showing lower RMSD values of chemical shifts and residual dipolar couplings when comparing to crystal and NMR structures; although, these improvements were ≤ 0.11 ppm. The scalar coupling values for Ala_3 were also calculated with this revision for ff99SB, ff99SB_CS and ff99SB_CS_RDC as shown in table 3.

Table 3 χ^2 values for Ala_3 for optimized force fields based on NMR observables

ff99SB	ff99SB_CS	ff99SB_CS_RDC
1.69	1.55	1.88

Table 3 χ^2 values for Ala_3 for optimized force fields based on NMR observables. The results obtained from ff99SB_CS_RDC force field were slightly worse than the ones obtained with ff99SB. The authors sustained that the *J* scalar coupling constants fall within the uncertainties from experiments and Karplus parameters used.

The performance of several dated and recent AMBER force fields along with variants of other force fields such as OPLS and CHARMM have been evaluated to determine its ability at reproducing NMR observables in small and medium size peptides, and their secondary structure propensities from folding simulations. Pande et al. [164] showed that overall several variants of AMBER force fields have the best performance when calculating NMR observables like scalar couplings. The results observed were within uncertainty of experiments. Other reviews as the one from Schulten et al.[165] using the CHARMM-CMAP correction show that after several microseconds of simulations this force field is not capable of folding the Pin 1WW domain, a β-sheet structure. The authors discuss that the CHARMM-CMAP correction has been known to bias helical conformations in small peptides and this issue could be resolved by introducing a better treatment of the hydrogen bond patterns to better match crystallographic structures[166]. One of the most recent reviews of the performance of these new versions of force fields was reported by Lindorff-Larsen et al. [73] where they showed that CHARMM-22*[45] and

AMBER99SB*-ILDN had the best agreement with experimental observables for peptide and protein systems such as CLN025 and $(AAQAA)_3$ 15-mer peptide, ubiquitin and GB3. The CHARMM-22* force field was obtained by modifying the CHARMM22 parameters in order to obtain better agreement for simulations of helical proteins[45]. AMBER99SB*-ILDN is a force field that was created by merging the parameters of AMBER 99SB* from Best et al. [89] and the ILDN optimized parameters for side-chain torsions of Isoleucine, Leucine, Aspartic acid and Asparagine from Lindorff-Larsen et al.[158]. Overall, AMBER force fields are the best force fields for simulations of biomolecules across the reviews discussed here.

The efforts of improving the backbone dihedral parameters were continued here, by introducing modest changes to the backbone dihedral energy function of ff99SB in particular to the φ', ψ and ψ' potentials. The ultimate goal was to improve the agreement for poly-alanines with NMR data and helical propensities of the force field without significantly altering other secondary structure regions. The most appropriate modifications to suit this purpose involved the φ', ψ and ψ' energy terms for which thirty force fields were derived, with small changes in the range of 0.2 to 0.4 kcal/mol. These modifications showed improvement in the agreement with scalar couplings for Ala_5 and higher helical propensities for transient helical peptides such as Baldwin-type peptide known as K19[167] and a smaller peptide derived from the Bak BH3 domains[168]. Furthermore, improved agreement with NMR order parameters for Lysozyme was also observed.

1.7 Goals of the thesis: To provide enhanced backbone dihedral parameters for protein simulations

The main aim of the thesis is to provide a new set of backbone dihedral parameters for protein simulations by introducing modest changes to the backbone dihedral energy function of AMBER ff99SB in particular to the φ', ψ and ψ' potentials. The idea of introducing these modifications was conceived to improve the accuracy of molecular dynamic simulations of transiently structured peptides. These modest changes to the potentials would be tested and

validated against a variety of peptide systems in order to discard the changes that worsen the accuracy of simulations or perform comparably to AMBER ff99SB.

AMBER ff99SB[142] was chosen as the force field of interest because it has a solid trajectory of being one of the best force fields for protein simulations[69, 72, 151, 169-171] with more than 1500 citations as per Google scholar. AMBER ff99SB has become popular among the scientific community for its ability to quantitatively and qualitatively reproduce the NMR observables [69, 72, 151] seen in experiments.

Nevertheless, force fields are not perfect because of the limitations in the methods used for the derivation of parameters obtained from quantum calculations such as fitting to secondary structure minima and the lack of solvent representation. Empirical fitting is necessary in order to close the gap between the results obtained from gas phase quantum calculations to the data obtained from protein experiments. Even though AMBER ff99SB has represented one of the highest standards for force field performance, the force field would benefit from fine tuning in a few areas.

1.7.1 Specific Aim 1: Improving the agreement between results from simulations and experiments for poly-alanine peptides

One of these areas is the agreement with poly-alanine peptides [151] as this has become one of the standards for evaluation of the force field quality [89, 151, 152, 164, 172, 173]. Results from previous simulations of poly-alanine peptides [151, 152] have indicated that agreement between experimental results and simulations could be improved. Furthermore, it appears that the largest discrepancies are due to the φ torsional potential [151]. In order to determine the φ dihedral populations of the force field, the penta-alanine peptide was simulated following the protocol described by Wickstrom et al. [151]. This peptide has been previously simulated in a comparative study performed by Best et al. [152] in which they used several variants of AMBER, CHARMM, GROMOS and OPLS force fields.

Figure 1 shows the comparison between the dihedral populations obtained from simulation with AMBER ff99SB (left panel) and the dihedral φ and ψ angles of alanine for 500 proteins obtained from the Protein Data Bank [149]. Although the information provided in the

rotamer library can be useful to determine the propensities of alanine in proteins, caution must be exercised when comparing the results from simulations to experimental data. Firstly, the data shown in the rotamer library is the collection of snapshots of protein structures that represent ensemble averages under specific experimental conditions that might not be applicable to the conditions described in the simulations. Lastly, simulations are carried in a small Ala_5 peptide not topologically comparable to proteins. The comparison done here serves as a guide to determine how the φ dihedral angles are behaving in simulation and to determine if this behavior is seen in proteins. As seen in figure 1 the comparison indicates that the sampling in the region of $-150° > \varphi > -120°$ is more pronounced in the simulation than in the rotamer library.

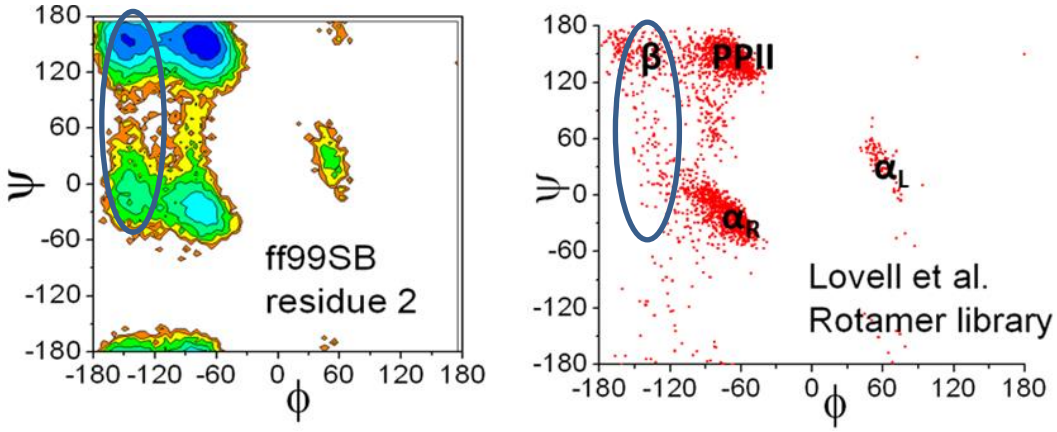


Figure 2 Comparison between the dihedral angle populations from simulations for Ala5 peptide (left) and the Protein Data Bank rotamer library of backbone dihedral angles for alanine [149].

In order to quantitatively benchmark these results, it was necessary to compare experimental results to simulation values directly. The method that would allow for direct comparison is scalar couplings. Recent studies by Graf et al. [172] measured NMR scalar couplings for backbone atomic interactions of several poly-alanine sequences (Ala_3 – Ala_7) and compared them to molecular dynamic simulations. Since this was published in 2007, the information provided has been used as a benchmark for performance of force fields [89, 162, 164, 173]. Scalar couplings have a direct relationship with dihedral angles that is described by the Karplus equation [174] presented below. This equation is particularly useful in order to calculate scalar coupling values in a straight-forward manner. The ^3J (H_N, H_α) scalar coupling

value was selected to obtain an initial estimation of the force field performance. This scalar coupling was chosen because it depends on the φ backbone dihedral angle and it has a wide range of values between 2 and 11 Hz depending on the Karplus parameters chosen for the fitting. This wide range allows for clear distinction of specific regions in φ dihedral space.

$$J = A\,cos^2\theta + Bcos\theta + C$$

Equation 5 The Karplus equation which establishes the relationship between scalar coupling values and dihedral angles. Below the function for the ^3J (H_N, H_α) scalar coupling is shown based on this equation.

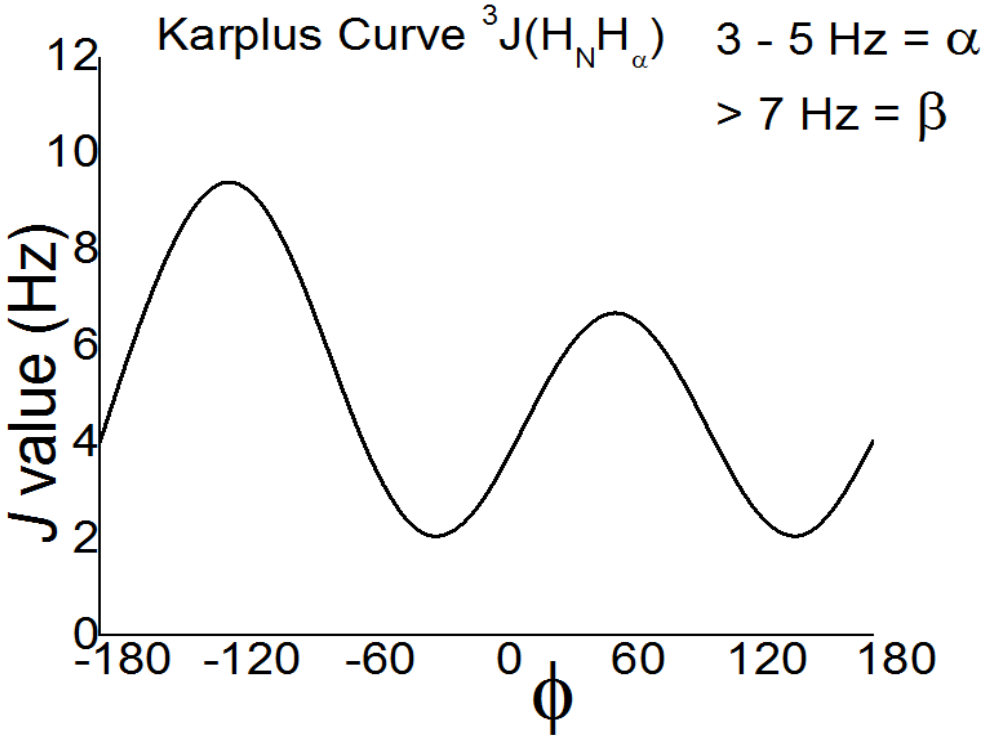


Figure 3 Karplus function for ^3J(H_N, H_α) scalar coupling based on the parameters published by Hu et al.[175]. The Parameters used for the function were A = 7.09, B = -1.42 C = 1.55 and θ = φ dihedral value – 60°.

The results from the simulations of AMBER ff99SB for residue 2 in Ala_5 indicated that the scalar coupling value was 6.99 Hz whereas the experimental value is 5.59 Hz. This result indicates that the simulations are considerably sampling high scalar coupling values. As shown in figure 3, the dihedral populations in the $-150° \leq \varphi \leq -120°$ region are contributing to this large value. In order to reduce the average value of the scalar coupling close to 5 Hz, it would be necessary to increase the sampling in the $\varphi = -60°$ region because this is the region where these values can be found according to the Karplus function (figure 2).

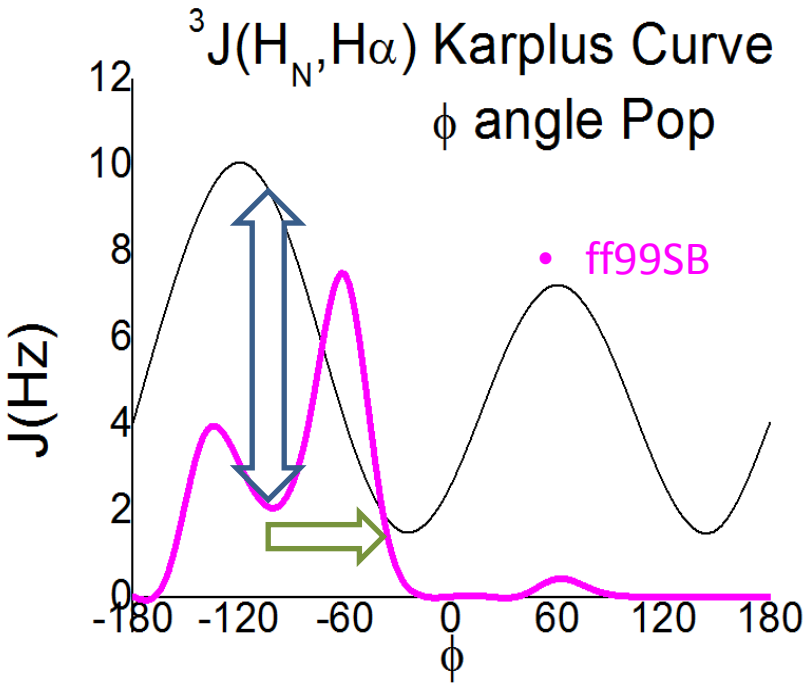


Figure 4 Histograms of dihedral populations from simulation with ff99SB shown in magenta. The populations were normalized and fitted to the range of the Karplus function shown in black for ^3J (H_N,H_α) scalar coupling. The function was obtained from the Hu et al. parameters described in figure 2.

1.7.2 Hypothesis

From figure 3, it appeared that the solution for the problem could be obtained by modifying the φ dihedral potential of the force field. In principal this modification should decrease the sampling in the $-150° > \varphi > -120°$ region. Furthermore, this modification would affect populations at different ψ values in dihedral space; it would affect the region around the $\psi = -150°$ and the region around $0 \geq \psi \geq -30°$ (figure 1).

This would narrow the sampling of the simulations for the ppII and the helical regions as

indicated in figure 4. However, it is apparent that this modification could also impact the β region in the Ramachandran map. Therefore, the proposed changes to the φ dihedral potential are small to not detrimentally affect the sampling of the β region in simulations. The proposed changes were in the range of 0.2 to 0.4 kcal/mol.

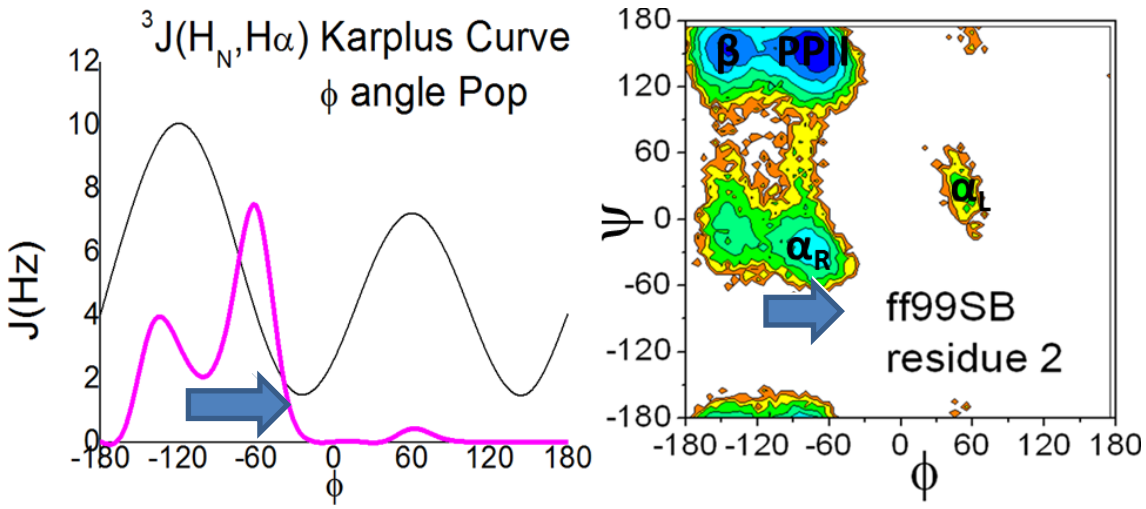


Figure 5 Comparison between the histograms of the dihedral populations for the φ angle (left, magenta) and the Ramachandran plot generated from simulations with AMBER ff99SB for residue 2 of Ala$_5$ (right). The Karplus function for the ^3J(H$_N$,Hα) is also included (left, black) [175].

These modifications were tested on Ala$_5$ as done in the case of AMBER ff99SB; other scalar coupling values that are dependent on the φ and ψ dihedral angles were compared against NMR *J* scalar coupling experimental data for all the residues in the peptide [172]. The scalar coupling values used for comparison were ^3J(H$_\alpha$,C), ^3J(H$_N$,C'), ^3J(H$_N$,C$_\beta$), ^3J(H$_N$,H$_\alpha$), ^3J(C,C') for the φ angle; ^1J(N,C$_\alpha$), ^2J(N,C$_\alpha$) for ψ angle and ^3J(H$_N$,C$_\alpha$) for the φ and ψ angles. The differences between the average obtained from simulations and the experimental values were calculated in the form of a χ^2 value using the definition provided by Best et al. [152].

1.7.3 Caveats of using calculated scalar couplings to benchmark the results from simulations

Although calculating scalar couplings from MD simulations is currently possible due to the Karplus function, the method has limitations. Karplus functions have been obtained by fitting

to X-ray crystallographic and NMR data of proteins such as ubiquitin, flavodoxin and others[176] [175, 177-181]. The parameters A, B and C are empirically derived and depend on the atoms and substituents used for the calculation. Uncertainties obtained from the X-ray and NMR data are intrinsically added to the derivations. Furthermore, scalar couplings measure spin/spin interactions, which depend on more than dihedral angles, bond lengths/angles, neighboring spin systems, hydrogen bonding etcetera.

Another caveat that impacts the calculations obtained from the Karplus function is precision. The Karplus function at the top of the hills and bottom wells have a wider range of possible dihedral angles that might represent a given scalar coupling value; whereas at the regions where is steep the range of dihedral angles is much narrower. For example, in figure 2, the region for the 8 Hz is at the top of the hill with a much wider range of dihedral space than the region between 3-4 Hz mainly dedicated to the -60° in φ dihedral space. This particular effect can contribute significantly to the results obtained because that is the region of interest for the modifications proposed.

Besides what has been mentioned, another aspect that requires discussion is the calculation of scalar couplings from parameters obtained experimentally; as discussed by Brüschweiler et al. [182] , when calculating scalar couplings with parameters derived from X-ray/NMR data for dihedrals from simulations; corrections should be implemented in order to avoid overemphasis of thermal fluctuations. Although this effect does not appear to overcome the intrinsic propensities of the force field choice in simulations [151].

1.7.4 Specific Aim 2: Increase the helical propensities for simulations of transient helical peptides

Following up on the work that was reported in 2008 [152], Best et al. published a comparative study between the performance of AMBER ff99SB and another contemporary AMBER force field known as ff03 [89]. Their work indicated that ff99SB had lower helical propensities than the NMR chemical shifts for the Ac-$(AAQAA)_3$-NH_2 peptide. In order to make the results from simulations and NMR comparable, the authors used the Lifson-Roig model to define helical propensities; their findings indicate that ff99SB has low helical propensities, whereas ff03 has high helical propensities compared to experiments.

In order to corroborate this, two other peptides known as Hydrogen Bond Surrogate (HBS) peptide [168] and K19 [167] were simulated with ff99SB. HBS peptide was chosen as a model because of its ten residue size makes it suitable for Replica Exchange Molecular Dynamics (REMD). Replica Exchange Molecular Dynamics is a type of Monte Carlo algorithm that allows obtaining enhanced sampling from simulations. This is particularly advantageous because more statistically robust data is obtained from this method. Another reason to choose this peptide is because there is NMR scalar coupling data available that can be useful to compare against the results obtained from the simulations. Although K19 is a larger peptide that cannot be studied using REMD simulations, it is small enough to perform long MD simulations that can be useful to study the differences in behavior between ff99SB and the modifications. This peptide has helical propensities obtained from NMR chemical shift deviations that can indicate folding behavior at room temperature.

As indicated in figure 5, the helical propensities from simulations eventually frail without attempting to reach a higher helical fraction value. The metric for helical propensities in this case was defined by the Define Secondary Structure in Proteins (DSSP) algorithm [183] which measures the intra-backbone hydrogen bonding pattern between the neighboring residues by means of electrostatic calculations. This metric cannot be compared directly to the metrics applied to experiments in order to determine helical fraction because in the case of the Hydrogen Bond Surrogate peptide [168] they used the ratio $[\theta]_{222}/\ [\theta]_{max}$ from the Circular Dichroism spectrum and for K19 they used Lifson-Roig theory from NMR Chemical Shift Deviations [184] to estimate fraction folded. In both cases the experiments were taken at room temperature as the simulations. Although the results cannot be directly compared, the results from experiments can provide insights for the behavior of the peptides at room temperature.

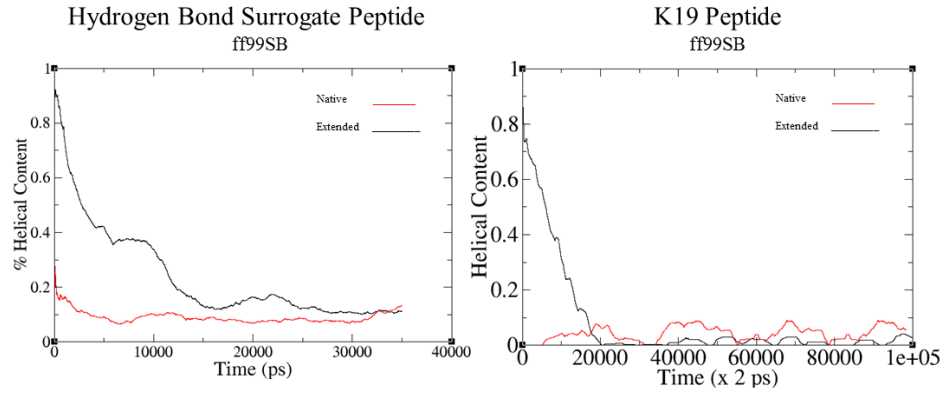


Figure 6 Helical Content vs. time for two helical peptides known as Hydrogen Bond Surrogate (HBS) peptide (left) and K19 peptide (right). The results for HBS peptide were obtained through Replica Exchange Molecular Dynamics (REMD) whereas the results for K19 were obtained with canonical Molecular Dynamics. The native structure for both peptides was obtained by imposing helical dihedral angles in all the amino-acid residues. The extended structure was obtained by imposing helical dihedral angles in the first five amino-acid residues of the peptides without imposing any other secondary structure angles for the rest of the residues. The structures were built using the tleap module in Amber tools 1.5.

1.7.5 Hypothesis

In order to correct this issue there is three possible solutions. The first one is to take advantage of the proposed solution postulated to improve the agreement with poly-alanines. As previously stated, this proposed modification in φ would have two effects in ψ dihedral space; it would narrow the sampling for the secondary structure region for ppII and it would also narrow the helical region. As shown in figure 4, modifying the dihedral potential to increase the sampling in the φ = -60° dihedral region could achieve this purpose. The sampling in the $-150 \leq \varphi \leq -120$ region would decrease allowing for more structures to simulate the helical region. However, this effect would also de-stabilize the β region and drive the populations to the ppII region. Another possibility would be to increase the helical sampling by modifying the potential in ψ dihedral space. This type of modification was previously reported in the work of Best et al. [89].

In the Best et al. work, the ψ potential was modified by including a cosine correction term for ff99SB and ff03. The authors recognized that including a cosine correction term goes against the AMBER notion of including separate definitions for the canonical φ /ψ backbone dihedral angles and the dihedrals dependent on the torsion around the β-carbon denominated φ'/ψ'. However, the cosine correction term takes advantage of the same philosophy behind the definition of ψ' ($\psi' = \psi + 120°$) because the authors included an angle offset in their correction and an adjustment of the amplitude of the cosine function.

In principle, modifying the ψ potential would drive the sampling from the ppII region to the helical region. This could be achieved by modestly introducing changes without significantly altering the propensities of ff99SB, similarly in size to the changes proposed to improve the agreement with poly-alanine systems.

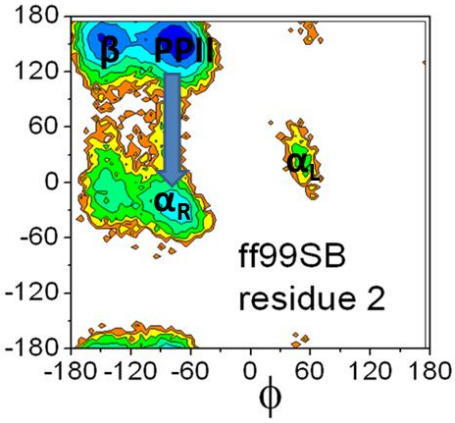


Figure 7 Dihedral populations obtained from simulation with ff99SB for Ala_5 residue 2. The arrows indicate the proposed modification to the ψ potential in order to drive the sampling from the ppII region to the helical region

Although the proposed solution would increase the sampling in the helical region, it would not solve the problem of the agreement with the poly-alanine experimental results. This proposed change might make the agreement with the poly-alanine data even worse.

In order to account for both problems, the modifications should couple both proposed solutions: modest changes to the φ and ψ dihedral potentials. The changes to the φ and ψ potentials alone would also be explored in order to determine their applicability.

The third possibility would be to introduce a modification that would couple the φ and ψ dihedral terms by introducing Gaussian modifications to the regions of interest. This type of modifications were introduced to the CHARMM [117] force field in 2004 by Mackerell et al. [148]. This modification is known as the CMAP correction. The rationale behind the correction was to improve the fitting to the dihedral maps from the protein Data bank. Although this proposed modification appeared to be the exact solution to the dihedral dilemma, the results of the simulations proved to be biased towards secondary structures such as helices; requiring the use of high temperatures and further refinement in order for simulations to obtain the appropriate

secondary structure populations [45]. It would also require the addition of an extra term to the force field function that would be related to the existent backbone dihedral term.

1.7.6 Challenges for benchmarking the results obtained from simulations of helical systems

HBS peptide has an un-natural covalent bond that holds the first five residues as in the turn of a helix. Experimentally, this turn was designed to promote helix nucleation in the peptide and it was introduced from the hydrogen of the amide bond in the first residue (cap) to the oxygen of the amide bond in the fifth residue (Alanine). The covalent bond is supposed to replace the naturally occurring hydrogen bond between *i i+4* residues. For our purposes, we cannot include this C-C covalent bond explicitly because besides the validation of the backbone parameters proposed here, it would also require the validation of the charges for the covalent bond. In order to overcome this, it was necessary to impose distance and angle restraints that would resemble the values from the C-C covalent bond introduced by experiments. This would limit our ability to directly compare the experimental results to the results obtained from simulations to certain extent. However, it is expected that qualitatively, the simulations would resemble the experimental results.

The experimental results from K19 NMR chemical shift deviation experiments were fitted to helical fractions using a Lifson-Roig model [184], whereas helical fractions in this case would be estimated using DSSP theory [183]. Besides, long MD simulations would be required in order to obtain qualitatively relevant data. Although the experimental results cannot be compared directly, the trends can help to describe how the modifications are affecting the helical propensities.

1.7.7 Specific Aim 3: Implement the proposed modification without significantly altering the force field propensities for other secondary structures

Overall the goal is to improve the agreement for poly-alanines with NMR data and helical propensities of the force field without significantly altering other secondary structure regions.

The modifications to the φ potential would intrinsically affect the β region but the changes should be designed to affect this region to the least extent. Although the proposed changes would be very small (~ 0.5 kcal/mol) or less, these small changes can affect the results from simulations significantly and should be carefully evaluated in order to validate their applicability to the proposed solution. For this purpose we would investigate a small peptide system known as Val_3 and four hairpins known as CLN025, chignolin, HP5F and GB1m3. These hairpins systems were chosen because of their size and because of their fast folding properties. Besides, they also have experimental information that can be useful to qualitatively compare the results from simulations.

Val_3 is another peptide system that has experimental *J* coupling information available that was published along with the poly-alanine peptides [172]. This peptide has been defined as having extended β conformation and its size makes it an ideal candidate for initial testing using REMD simulations. This peptide has been previously studied for the purpose of force field validation [154, 172]. The results obtained from simulations would be compared in a similar manner as proposed for Ala_5. The same type of scalar coupling information that is available for Ala_5 is available for Val_3 which allows us to use the same previously proposed methodology. Initially the agreement with the $^3J(H_N,H_\alpha)$ scalar coupling value was investigated then the rest of the scalar coupling values: $^3J(H_\alpha,C)$, $^3J(H_N,C')$, $^3J(H_N,C_\beta)$, $^3J(H_N,H_\alpha)$, $^3J(C,C')$ for the φ angle; $^1J(N,C_\alpha)$, $^2J(N,C_\alpha)$ for ψ angle and $^3J(H_N,C_\alpha)$ for the φ and ψ angles were investigated. Similarly to the Ala_5 case, the differences between the average obtained from simulations and the experimental values were calculated in the form of a χ^2 value using the definition provided by Best et al. [152].

Although the results obtained from Val_3 peptide can be helpful to evaluate the performance of force fields to an initial extent, further simulations on β-like systems are required to further assess the qualities of the force fields. Therefore, it is proposed to study larger hairpin peptides, in this case chignolin [185] and CLN025 [186]. Chignolin was chosen as a test case, because of its ten amino-acid residue size. This peptide has been previously studied by computational methods extensively [47, 53, 55, 59, 187-190]. Another property that makes it attractive for testing is its folded properties at room temperature (~50%) obtained from Chemical Shift Deviations from backbone/side-chain protons and CD spectrum. This transient nature is

ideal for this purpose because the modifications should be able to simulate systems like this in the same manner as it is proposed with the helical systems previously described. In the case of the CLN025 its fast folding properties were of interest. Hairpins fold at a slower rate than helices, requiring longer simulation times to obtain relevant statistical ensembles. A fast folding hairpin would be a good model system to obtain statistically relevant data in order to determine how the propensities are being affected.

Besides these two hairpin systems we also propose to study two other hairpins slightly larger in size (16 residues). These hairpins known as GB1m3 and HP5F were chosen for this purpose [191]. They were obtained from single point mutations to the GB1p sequence of Serrano et al. [192]. Both hairpins are almost identical, the only differences are at the 3^{rd} and 5^{th} amino-acid positions. In the case of HP5F, the 3^{rd} residue is Tyrosine and the 5^{th} residue is Tryptophan. In the case of GB1m3, the 3^{rd} residue is Tryptophan and the 5^{th} residue is Tyrosine. The rest of the amino-acid sequence is identical for both. GB1m3 has been previously simulated using Monte Carlo simulations [56] and HP5F was used as a test system for the validation of the revised GB Neck implicit solvent model by Nguyen et al. [193]. For our purposes, HP5F was simulated with REMD simulations and revised GB Neck implicit solvent model [193] whereas GB1m3 was simulated with canonical MD simulations and TIP3P explicit solvent model. Both systems have comparable melting temperatures and thermodynamics properties[194] . Although the simulations for HP5F would allow more statistical relevant information because of the enhanced sampling method, the simulations under explicit solvent could be useful to have a quick estimation of how the simulations differ between ff99SB and the modifications.

1.7.8 Hypothesis

The proposed modifications to the φ potential and the combinations of φ and ψ modifications would make the sampling of the β region in dihedral space more sparing because in order to increase the sampling in the φ = -60° the populations in the $-150 \leq \varphi \leq -120$ would be reduced and driven to the ppII/helical region in φ dihedral space. Therefore, it is expected that this would be resembled in the simulations; the β region would be less sampled. However the modifications would be designed so this effect is minimal. The most appropriate modifications to

suit this purpose involved the φ', ψ and ψ' energy terms for which thirty force fields were derived, with small changes in the range of 0.2 to 0.4 kcal/mol.

1.7.9 Challenges for benchmarking the results from simulations of hairpin systems

Calculating scalar coupling values from simulations for poly-alanine peptides can be challenging due to uncertainties in the fitting of the Karplus functions. One of the core aspects of the modifications presented here is to improve the agreement with experimental NMR scalar coupling values for the poly-alanine peptides in particular Ala_5. However the differences between the experimental values for Val_3 and Ala_5 are significant. Experimentally, Val_3 is described as an extended β peptide [195, 196] whereas Ala_5 is described as a poly-proline peptide according to experiments [172]. The proposed modification as described above would make the sampling of the β region less frequent, which in turn would reduce the agreement with the scalar values because one of the desired results is to improve the agreement of the backbone dihedral propensities from simulations with the rotamer library for alanine [149]. The assumption is that alanine behaves like the other amino-acid residues.

Another challenge is related to the currently available experimental data for the hairpin systems. The structure for chignolin was obtained from the PDB NMR structure access code (1UAO). The structure for CLN025 was obtained from the Biological Magnetic Resonance Data Bank access code 20009. Mostly the reported data for both hairpins was obtained from Chemical Shift Deviation (CSD) experiments and CD spectra. The CSD data was fitted to a two state model using the protons that moved more than 0.07 ppm with temperature in order to obtain the melting curve. For the CD spectra the melting curve was obtained from the $[\theta]_{222}/ [\theta]_{max}$ ratio [185]. As previously mentioned it is not possible to directly assess the data from simulations against experimental values because such metrics cannot be obtained from simulations. However, the trends observed from simulations can be compared to the trends observed from experiments in order to have a reasonable idea how the simulations are due to behave at room temperature.

1.7.10 Specific Aim 4: Evaluate backbone dynamics from NMR order parameter calculations for simulations of lysozyme and ubiquitin

Agreement between NMR order parameters calculations for proteins from simulations and experiments have become one of the standards for force field quality; several groups [69, 72, 89, 90, 142, 154, 197-199] have dedicated research efforts to compare the results obtained from experiments to those obtained from simulations. Therefore, it is necessary to investigate how the proposed modifications would perform with simulations of proteins, in particular lysozyme and ubiquitin for which there are several sources for information of this measurement [4] [200-205].

The calculation of order parameters is necessary in order to determine the stability of the backbone interactions from simulations. For this purpose, the dynamics of the N-H bond are measured for most of the amino-acids in the protein with the exception of proline and the termini residues. The dynamics of the ordered portions of the molecule such as helices, β strands and others would be almost rigid; that is slower; whereas the less ordered regions such as loops would be more flexible, meaning faster. In order to quantify these motions, the preferred method of choice is the Lipari-Szabo model free approach where the global motion of the protein is decoupled from the internal (bond) motions [206].

As described by the authors the method works really well because the internal motions are fast enough to be specified by the S^2 order parameter and the τ_c correlation time terms from the spectral density function. The spectral density equation is the probability function of finding motions at a given angular frequency (ω):

$$J_{(\omega)} = \left(\frac{2\tau_C}{1 + (\omega\tau_C)^2}\right)$$

Equation 6 Spectral density function that describes the probability of finding motions at a given angular frequency ω. T_c is the time needed for the root mean square deflection of the molecules to be ~1 radian (60°)

As described by Lipari-Szabo [206] the motions of the bond can be described by monitoring the bond vector over time starting at some random time *t* and any time *i* + *t*. The motions can be compared against each other. A number can be assigned to describe the

probability of finding the bond vector in the same orientation at $i + t$ that it had at t. At the beginning this probability would be very high and the larger t got, the probability would be smaller [207]. A sample of how the function would be like is shown in figure 7. In this case the independent variable is frequency.

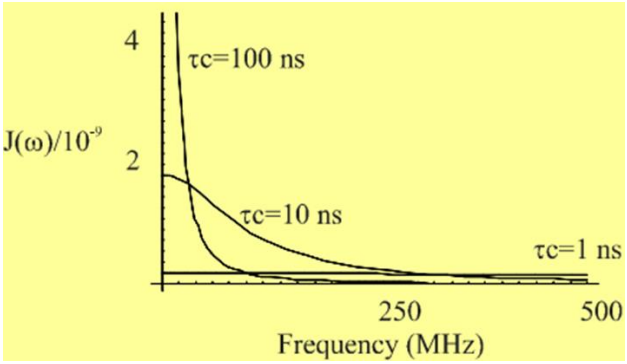


Figure 8 Spectral density function for the relaxation time. Molecules that tumble very rapidly can sample a wide range of frequencies. Molecules that tumble slowly with long correlation times only sample lower frequencies. Taken from http://www.ncbr.muni.cz/~lzidek/C6770/presentations/Dynamics_2.pdf

One more concept that is essential for the understanding of order parameters is the correlation function. This is called the correlation function, because it represents the correlation between the orientations at time i and the orientation at time $i + t$.

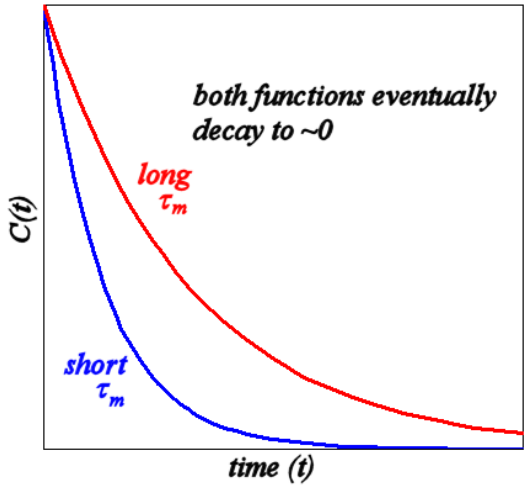


Figure 9 Correlation function for fast and slow motions of a bond vector. The correlation is represented between the orientation of the vector at time *I* and the orientation at time *i+t* taken from http://conflux.mwclarkson.com/2007/11/the-model-free-dynamics-formalism-of-lipari-and-szabo/

As seen from figure 8, the correlation function can be approximated to an exponential decay dependent on time t and the time constant τ_m based on the speed at which the bond

tumbles. Assuming that the motion of the molecule is isotropic, the function could be expressed as:

$$C_m(t) = \frac{1}{5} e^{\frac{-t}{\tau_m}}$$

Equation 7 Correlation function for the exponential decay of the motions of the bond vector studied. $C_M(t)$ is the correlation function, t is the time (x axis), and τ_m is the time constant that represents the speed at which the bond tumbles; also known as the rotational correlation time of the molecule.

Since the bond is part of a larger system such as a protein, an extra function needs to be introduced in order to account for bond constraints inside the protein. This function will also approximate an exponential decay, however the motions will not decay to zero as originally postulated, instead they will decay to a static value known as the order parameter S^2. If the bond is absolutely rigid, meaning there is 100% probability of finding the bond at the same place, the order parameter S^2 will be one and if is completely flexible this value will be zero. Figure 8 shows the dependence of the correlation time *C(t)* on time. Motions that dampen fast will have a high S^2 value whereas motions that dampen slower will have a low S^2. Nevertheless, the correlation function will never be lower than the S^2 value.

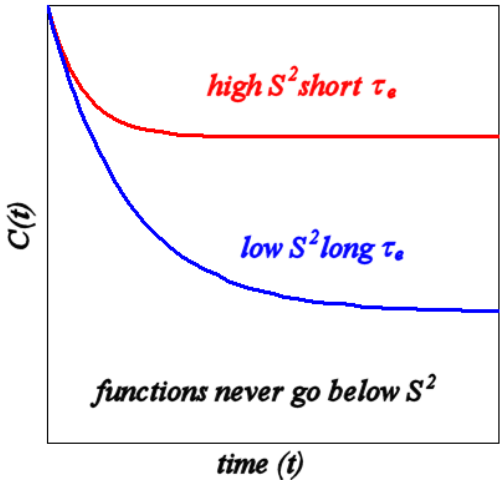


Figure 10 Correlation time function of motions for the bond vector inside a larger system such as a protein. This function accounts for the bond constraints inside such large system. Motions that dampen fast will have a high S^2 value whereas motions that dampen slower will have a low S^2. Nevertheless, the correlation function will never be lower than the S^2 value. Taken from http://conflux.mwclarkson.com/2007/11/the-model-free-dynamics-formalism-of-lipari-and-szabo/

Similarly to the first case, this "extra function" can also be expressed in terms of exponential decay; however the order parameter S^2 and the extra correlation time make the

function slightly different than the first equation. As shown in equation 1-4 the equation includes the S^2 value and the extra correlation time that will be denominated as $C_e\ (t)$. The extra time is denominated as τ_e.

$$C_e(t) = \frac{1}{5}\left(S^2 + (1-S)^2 e^{\frac{-t}{\tau_e}}\right)$$

Equation 8 Correlation time equation for the bond constraints that bond vector N-H experiences inside the protein. T_E is the extra correlation time that takes for the constrained motions to dampen to the static value S^2.

In the event of both functions being correlated, the calculation of the total correlation time would be very complex. However, as mentioned previously, the Lipari-Szabo method assumes that the motions of the bond vector and the protein are decoupled and makes the calculation simpler. Therefore in order to combine both the correlation time of the molecule $C_m(t)$ and the extra correlation time for the constraints of being inside the protein $C_e(t)$ it is possible to multiply both terms to obtain the total correlation time as described in equation 1-5:

$$C_T(t) = \frac{2}{5}\left(S^2 e^{\frac{-t}{-\tau_m}} + (1-S^2) e^{\frac{-t}{-\tau_T}}\right)$$

Where

$$\frac{\mathbf{1}}{\boldsymbol{\tau_T}} = \frac{\mathbf{1}}{\boldsymbol{\tau_m}} + \frac{\mathbf{1}}{\boldsymbol{\tau_e}}$$

Equation 9 Total correlation time function that incorporates the molecular correlation time and the extra correlation time.

Combining the spectral density function from equation 1-2 and the total correlation time for the molecule would give the spectral density function:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_m}{1+\omega^2{\tau_m}^2} + \frac{(1-S^2)\tau_T}{1+\omega^2{\tau_T}^2}\right)$$

Equation 10 Spectral density function that combines the molecular correlation function and the total correlation function in terms of the angular frequency ω.

The explanation provided here does not describe how the spectral density is related to S^2 and the relaxation parameters T_1 and T_2. So far, only the explanation for the correlation function and correlation time has been given. In order to describe it, the spectral density function needs to be defined in terms of the relaxation parameters.

As described by Lipari-Szabo [208], in the case of protonated carbons for ^{13}C NMR, nuclear relaxation is mainly dipolar, the dipole-dipole interaction between two nuclei is described by the correlation function [209, 210]:

$$C(t) = \langle D^2_{q0}\big(\Omega_{LF}(0)\big) D^2_{q0}(\Omega_{LF}(t))\rangle$$

Equation 11 Correlation function for dipolar nuclear relaxation, where $D_{q0}^{(2)}$ (Ω_{LF}) is a Wigner rotation matrix element [211] and the Euler angles, (Ω_{LF}). The Euler angles (Ω_{LF}) specify the orientation of the unit vector, ($\hat{u}_{LF}$) connecting the two nuclei in the coordinate system. This correlation function also describes the quadropolar and chemical shift anisotropy relaxation in the special case where the relevant tensors are axially symmetric.

For a system in solution, the correlation function does not depend on the index q and can be re-written by using the addition theorem for spherical harmonics [211] as:

$$C(t) = \frac{1}{5}\langle P_2(\hat{u}_{LF}(0)\cdot\hat{u}_{LF}(t))\rangle$$

Equation 12 Correlation function that implements the addition theorem for spherical harmonics. $P_2(x)$ is the second Legendre polynomial: $P_2(x) = \frac{1}{2}(3x^2 - 1)$

The spectral density, which determines the relaxation parameters T_1, T_2, taken from Lipari-Szabo's description [206] is given by:

$$J(\omega) = 2\int_0^{\infty} (\cos \omega t) C(t) dt$$

Equation 13 Spectral density function that determines the relaxation parameters T_1 and T_2 in terms of angular frequency ω.

At time *t = 0*, the correlation function from equation 1-8 becomes:

$$C_I(0) = \langle P_2(\hat{u}(0) \cdot \hat{u}(0)) \rangle = 1$$

At time $t = \infty$ the function becomes:

$$C_I(\infty) = \sum_{m=-2}^{2} |\langle C_{2m}(\Omega) \rangle|^2 = \iint d\Omega_1 \, d\Omega_2 P_{eq}(\Omega_1) P_2(\cos \theta_{12}) P_{eq}(\Omega_2)$$

$$\langle (\cdots) \rangle = \int d\Omega P_{eq}(\Omega)(\cdots) = \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta \, d\theta P_{eq}(\theta, \phi)(\cdots)$$

$$C_I(\infty) = S^2$$

Equation 14 Correlation time function that defines the generalized order parameter S^2. If the internal motion is isotropic (all orientations of û are equally probable), $P_{eq}(\Omega)=(4\pi\text{-}1)$ then $S=0$. If the motion is completely restricted $P_{eq}(\Omega)=\delta(\Omega-\Omega_0)$ then $S=1$. When the motion is azimuthally symmetric about an axis $P_{eq}(\Omega)=P_{eq}(\theta)$, independently of ϕ, where θ is the angle between û and the symmetry axis, $S = \langle P_2(\cos\theta) \rangle = S$.

As indicated from equation 1-10, the correlation function information is essential to obtain the spectral density information and the generalized order parameter S^2. From the spectral density information, the relaxation times T_1, T_2 and the NOE parameters can be derived from the following equations [206].

For ^{13}C NMR where relaxation is primarily dipolar, the relaxation times and the Nuclear Overhauser Enhancement (NOE) are given by:

$$T_1^{-1} = \frac{\hbar^2 \gamma_C^2 \gamma_H^2}{4 r_{CH}^6} \left(J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_C - \omega_H) \right)$$

$$T_2^{-1} = \frac{\hbar^2 \gamma_C^2 \gamma_H^2}{8 r_{CH}^6} \left(4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H) + 6J(\omega_H - \omega_C) \right)$$

$$NOE = 1 + \frac{\gamma_H \left(6J(\omega_C + \omega_H) - J(\omega_H - \omega_C) \right)}{\gamma_C \left(J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_C + \omega_H) \right)}$$

Equation 15 Relaxation times T_1, T_2 and Nuclear Overhauser effect equations based on the spectral density function for 13C NMR.

Where $\hbar$ is Planck's constant, γ is the gyromagnetic radio of ^{13}C and H, r is their inter-nuclear distance, ω_I, I = C, H are the Larmor frequencies [212, 213]. For ^{2}H NMR of deuterated carbons, the quadrupolar tensor is axially symmetric about the C-^{2}H bond, and the relaxation times are given by [210]:

$$T_1^{-1} = \frac{3}{16} \left(\frac{e_q^2 Q}{\hbar} \right)^2 \left(J(\omega_D) + 4J(2\omega_D) \right)$$

$$T_2^{-1} = \frac{1}{32} \left(\frac{e_q^2 Q}{\hbar} \right)^2 \left(9J(0) + 15J(\omega_D) + 6J(2\omega_D) \right)$$

Equation 16 Relaxation times for ^{2}H NMR of deuterated carbons, the quadropolar tensor is axially symmetric about the C-^{2}H bond. The correlation function has the same form for both ^{13}C NMR and ^{2}H NMR because the $\hat{u}_{LF}$ vector points along the ^{13}C-H bond in ^{13}C NMR, while it points along the C-^{2}H bond in ^{2}H NMR.

Nevertheless, this is all possible assuming the simplest of the cases, that is, the bond has isotropic motion. This means that the bond has a single correlation time τ_m because the motions are equal in all x, y and z directions. However this is not always the case, when the contrary happens, meaning the molecule has anisotropic motion, then the method is slightly revised and the correlation time function becomes as described by the Lipari-Szabo method [206]:

$$C(t) = \frac{1}{5}\left(Ae^{-\frac{t}{\tau_1}} + (1-A)e^{-\frac{t}{\tau_2}}\right)\left(S^2 + (1-S^2)e^{-\frac{t}{\tau_e}}\right)$$

Equation 17 Correlation function for anisotropic relaxation. A, τ_1 and τ_2 are the adjustable parameters that can be determined by fitting the relaxation data of a nucleus that is attached to the macromolecular backbone. This equation has the same form as equation 1-5 that describes the correlation time for isotropic relaxation ($A \rightarrow S^2$, $\tau_m \rightarrow \tau_1$, and $\tau \rightarrow \tau_2$).

As described by the authors [206] the method is accurate even in cases where there are several correlation times involved as in the case of random-coil polymers, as long as the correlation times decay in an exponential-like manner. The dynamics of the local bonds must be much faster than the overall tumbling of the protein in order to be de-coupled.

1.7.11 Challenges of NMR order parameter calculations using the Lipari-Szabo method

The Lipari-Szabo model is applicable when the motions of the protein can be decoupled from the bond vector motions. However, this is not always the case; for instance, in certain systems where there is significant flexibility such as RNA loops, the local motions cannot be decoupled from the overall motions of the system [214] [215]. Other cases such as in protein loops where there is a significant amount of flexibility, bond vector reorientation is correlated even after overall protein tumbling is subtracted [216].

Another instance in which the method was not able to reproduce relaxation parameters was described by Clore et al. [217] where they compared T_1, T_2 and Nuclear Overhauser Enhancement (NOE) parameters for staphylococcal nuclease [218] and interleukin-1β [219] against estimates obtained from the Lipari-Szabo model free approach. They mentioned that the method appears to reproduce the data within the experimental error, however for some residues in both proteins the internal re-orientational correlation function that is probed by NMR is not exponential with slow correlation times that are not in the extreme narrow limit as described by Lipari-Szabo [206]. This effect was particularly seen when comparing the ^{1}H-^{15}N NOE data from experiment to the fitted data using equation 6. The fitted values were significantly small or negative whereas the experimental values were positive. The issue is due to fast and slow correlation times for the relaxation motion of these bond vectors, when these differ by more than

one order of magnitude the function does not decay exponentially. Instead, the function decays to a plateau intermediate before reaching the S^2 value. In order to account for this, Clore et al. [217] to revise the definition of the S^2 order parameter for the amino-acid residues where this effect was observed to:

$$S^2 = S_f^2 S_s^2$$

Equation 18 S^2 order parameter equation that decomposes the term for fast and slow correlation times when the difference between them is greater than one order of magnitude.

Where S_f and S_s are the correlation function values obtained from the fast (S_f) and slow (S_s) relaxation motions of the bond vectors.

Although this method is considered an extension of the Lipari-Szabo method, it shows instances where the model-free approach falls short and requires additional parameters for the estimation. Other methods used for the interpretation of hetero-nuclear spin relaxation data of proteins include spectral density mapping [220], and analytical motional models [221, 222]. These methods are not as popular as the model-free Lipari-Szabo approach but they are used for the estimation of NMR order parameters as well.

Recently, other methods based on re-orientational intra-molecular motions have been reported [216, 223, 224] that describe the correlation dynamics that affect spin relaxation parameters. These methods implement covariance matrices of spherical harmonics of rank 2 for inter-nuclear vector motions estimated from MD simulations. Covariance matrices are used because nuclear spin relaxation parameters for ^{15}N and ^{13}C spins are obtained from the deviations of the angular parts of the lattice functions of spin-relaxation interactions such as dipole-dipole and chemical anisotropy interactions. The lattice functions are often represented in terms of spherical harmonics of rank 2, Y_{2M} (θ,φ), where θ,φ are the polar angles describing the orientation of an inter-nuclear vector in a stationary snapshot [225]. As described by Bloch, Wangness and Redfield [226-228] relaxation is a second order perturbation and their three-dimensional properties are described by variance and covariance properties of the Y_{2M} (θ,φ) lattice functions.

In principle the re-orientational approach offers a more statistically relevant method because relaxation parameters are obtained from many conformations obtained from trajectory snapshots, whereas the experimental results are obtained from a single snapshot that poses as the ensemble average at a given time. The results obtained from MD simulations are compared against experimental data in order to determine its applicability [229]. This approach was applied because the correlation of spin interactions responsible for NMR relaxation is present at short distances between nuclei. Since atoms do not move individually from each other, but simultaneously involving groups of variable sizes, the direct identification and description of correlation times from experiments alone can be challenging; making it necessary for the supplementation with results from MD simulations [229].

1.7.12 Isotropic Re-orientational Eigenmode Analysis (iRED)

This method relies on principal component analysis of an isotropically averaged covariance matrix of the spin contacts in control of relaxation. This covariance matrix is generated from the snapshots of a molecular dynamics trajectory and diagonalized to obtain the eigenfunctions and eigenvalues that describe the correlated dynamics of the system being studied. The information provided by the eigenvalues determines if the overall and internal motions of the protein are separable, while the eigenfunction information provides the correlation time needed to calculate the relaxation parameters T_1, T_2 and NOE's described in equations 1-8 through 1-11. In principle the iRED method and the other re-orientational (RED) methods use covariance matrices of the lattice functions to describe the re-orientation of the N-H bonds. For molecules in which the internal and overall motions can be decoupled, the covariance matrix of the RED approach is described in terms of a static reference frame. The eigenmodes and eigenvalues are obtained by diagonalizing the covariance matrix where each eigenmode has its own correlation time [229].

However, in the case of more flexible systems, large-amplitude re-orientational dynamics of the backbone N-H vectors can be observed from MD simulations; where the covariance

matrix cannot be represented in terms of a single static frame. The iRED method solves these issues by analytically integrating each snapshot of an MD trajectory over an equally scattered distribution of orientations in spherical coordinate's space. The following derivations were taken from Prompers et al. which applied the method to the calculation of NMR order parameters for ubiquitin [230]. For each snapshot of an MD trajectory a *n(2L+1)* dimensional vector $|Y(t)\rangle$ is constructed from spherical coordinates Y_{LM} (Ω_j) evaluated at the *n* directions $\Omega_j(t)=(\theta_j(t), \varphi_j(t))$, $j=1, .., n$, of the inter-nuclear vectors N-H

$$|Y(t)\rangle = |Y_{L,-M}(\Omega_1), Y_{L,-M+1}(\Omega_1), .., Y_{L,M-1}(\Omega_1), Y_{LM}(\Omega_1), \dots\ Y_{L,-M}(\Omega_n), Y_{L,-M+1}(\Omega_n), \dots, Y_{L,M-1}(\Omega_n), Y_{LM}(\Omega_n)\rangle$$

Equation 19 Dimensional vector $|Y(t)\rangle$ built from spherical harmonics of rank 2 Y_{LM}. The notation described here is slightly different than the notation provided in equation 1-8 through equation 1-10, however the correlation times can be obtained from this definition as well in a similar manner.

The *n(2L+1)* × *n(2L+1)* covariance matrix *P* is constructed as described below:

$$P = \overline{\left(|Y\rangle - \overline{|Y\rangle}\right)\left(\langle Y| - \overline{\langle Y|}\right)} = \overline{|Y\rangle\langle Y|} - \overline{|Y\rangle}\,\overline{\langle Y|}$$

Equation 20 Isotropically averaged covariance matrix P generated from the dimensional vectors $|Y(t)\rangle$ built from spherical harmonics of rank 2 Y_{LM} as described in equation 1-15.

Where $|Y\rangle$ is the complex-conjugate row vector of column vector $|Y\rangle$. The horizontal bar indicates an ensemble average over the N conformations or a time average over a trajectory [230].

In order to account for the equally scattered distribution of orientations, in other words the isotropic effect, the MD snapshots are integrated analytically following the expression:

$$\boldsymbol{P}:\boldsymbol{Q} = \langle \boldsymbol{P}_{iso} \rangle$$

Equation 21 Analytical expression for the isotropically averaged matrix P: The term $\overline{|Y\rangle}\,\overline{\langle Y|}$ of P averages to zero for $L\neq 0$, since $\langle Y_{LM}(\Omega)\rangle_{iso} = \int d\alpha \sin\beta\ d\beta\, d\gamma\, Y_{LM}(\Omega) = 0$. The averaging of the term can be done using individual matrix elements $P_{M'k,M''l}$ along with the transformation properties in spherical coordinates R (α, β, γ) where α, β, γ are Euler angles [211]. The 3D rotation would be $R(\alpha,\beta,\gamma)Y_{LM}(\Omega) = \sum_{M'} D^{L}_{M'M}(\alpha,\beta,\gamma)Y_{LM'}(\Omega).D^{L}_{M'M}(\alpha,\beta,\gamma)$ are the Wigner matrix elements [211].

Using the concepts described in equation 1-17 the expression for the isotropically averaged matrix P becomes:

$$
\begin{aligned}
Q_{M'k,M''l} = \langle P_{M'k,M''l}\rangle_{iso} &= \overline{\langle Y_{LM'}(\Omega_k) Y_{LM''} * (\Omega_l)\rangle}_{iso} \\
&= \langle \overline{\boldsymbol{R}(\alpha,\beta,\gamma) Y_{LM'}(\Omega_k) \boldsymbol{R}^{\dagger}(\alpha,\beta,\gamma) Y_{LM''} * (\Omega_{\boldsymbol{l}})}\rangle_{\alpha\beta\gamma} \\
&= \frac{1}{8\pi^2} \sum_{N',N''} \int d\alpha \sin\beta d\beta d\gamma \times \overline{D^L_{N'M''}(\alpha,\beta,\gamma) Y_{LN'}(\Omega_k) D^L_{N'M''} * (\alpha,\beta,\gamma) Y_{LN''}(\Omega_l)} \\
&= \frac{1}{8\pi^2} \sum_{N',N''} \overline{Y_{LN'}(\Omega_k) Y_{LN''} * (\Omega_l)} \frac{8\pi^2}{2L+1} \delta_{M',M''} \delta_{N',N''} \\
&= \frac{\delta_{M',M''}}{2L+1} \sum_{N'=-L}^{L} \overline{Y_{LN''}(\Omega_k) Y_{LN'} * (\Omega_l)} = \frac{\delta_{M',M''}}{4\pi} \overline{P_L(\cos(\Omega_k - \Omega_l))}
\end{aligned}
$$

Equation 22 Isotropically averaged covariance matrix obtained from equation 1-17. As shown in equation 1-8 $P_L(x)$ is the Legendre polynomial of order L, and $\Omega_k - \Omega_l$ denotes the angle between directions Ω_k and Ω_l; $\delta_{M',M''}$ denotes the Kronecker δ. The Wigner matrix elements are given by: $\int d\alpha \sin\beta d\beta d\gamma D^L_{M'N'}(\alpha,\beta,\gamma) D^L_{M''N''} * (\alpha,\beta,\gamma) = \frac{8\pi^2}{2L+1} \delta_{M'M''} \delta_{N'N''}$ and the addition theorem for spherical harmonics is $P_L(\cos(\Omega_k - \Omega_l)) = \frac{4\pi}{2L+1} \sum_{N'=-L}^{L} Y_{LN'}(\Omega_k) Y_{LN'} * (\Omega_l)$

Since the matrix is isotropically averaged, the elements of the matrix **Q** depend only on the intra-molecular angles $\Omega_k - \Omega_l$ and no longer in the orientation. Matrix **Q** can be described as the matrix product $\mathbf{Q} = \frac{\mathbf{1}}{4\pi} \mathbf{M} \otimes \mathbf{1}$ where **1** is a *2L* + 1 dimensional unity matrix, and **M** is a *n* × *n* matrix with elements $M_{kl} = \overline{P_L(\cos(\Omega_k - \Omega_l))}$. The eigenvalues of **Q** are at least *2L* + 1 times degenerate, and all information enclosed in **Q** is also enclosed in **M,** this leads to simplification due to spherical symmetry for the covariance matrix **Q** [230]. Matrixes **Q** and **M** can be diagonalized $\boldsymbol{Q}|q\rangle = \lambda_q |q\rangle$ and $\boldsymbol{M}|m\rangle = \lambda_m |m\rangle$ and the eigenvectors of **Q** can be constructed from the eigenvectors of **M:** $|q\rangle = |m\rangle \otimes |e_j\rangle$**.** $|e_j\rangle$ are the *2L* +1 –dimensional column vectors $|e_1\rangle = (1,0,\cdots,0)^T, |e_2\rangle = (0,1,\cdots,0)^T, \cdots, |e_{2L+1}\rangle = (0,0,\cdots,1)^T$**.** The diagonal elements of **Q** are $Q_{ij} = {}^{1}/_{(4\pi)}$, and the diagonal elements of **M** are M_{ij} = 1. The traces of **Q** and **M**, which are

the sum of their respective eigenvalues (mode amplitudes), are proportional to the number of vectors n: $Tr\{\boldsymbol{Q}\} = \frac{n(2L+1)}{(4\pi)}, Tr\{\boldsymbol{M}\} = n$

The correlation times can be reconstructed by projecting the vector $|Y(t)\rangle$ constructed from the snapshot at time t onto eigenvector $|q\rangle, a_q(t) = \langle q|Y(t)\rangle$ in order to obtain the correlation function:

$$C_q(t) = C_{m,l}(t) = \langle a_{m,l} * (\tau + t) \cdot a_{m,l}(\tau)\rangle_\tau$$

Equation 23 Correlation time function for individual modes obtained by projecting vector $|Y(t)\rangle$ onto eigenvector $|q\rangle$ at time t. $q = (2L+1)m + l - L$, and $l = -L, \ldots, L$. The average covers snapshots sampled at times $\tau = 0 \rightarrow T - t$. The correlation functions can be obtained from matrix M by summing the $2L + 1$ correlation functions $C_{m,l}(t)$ to $C_m(t) = \sum_{l=-L}^{L} C_{m,l}(t)$ from the degenerate modes q.

Given the condition that the correlation function has exponential decay, the correlation time τ_m discussed in equation 1-5 of section 1.5.10 can be described as:

$$\tau_m \cong \frac{1}{C_m(0) - C_m(t \to T)} \int_0^T (C_m(t) - C_m(t \to T) dT)$$

Equation 24 Correlation time τ_m associated with mode $|m\rangle$. $C_m(t \rightarrow T)$ represents the plateau value of $C_m(t)$: $C_m(t \to T) = \sum_{l=-L}^{L} C_{m,l}(t \to T) \cong \sum_{l=-L}^{L} |\langle a_{m,l}(\tau)\rangle_\tau|^2$. For systems where internal and overall tumbling motions are separable, $C_{m,l}(t)$ and $C_m(t)$ and their correlation times reveal both internal and overall motions. Isotropic tumbling's correlation time is given by τ_c, the effective correlation time τ_m of an internal mode with an internal correlation time τ_{m}' is given by $\frac{1}{\tau_m} = \frac{1}{\tau_c} + \frac{1}{\tau_{m'}}$. This equation is equivalent to equation 1-5b described in the Lipari-Szabo model free method.

Having described how the correlation times and functions is obtained from the isotropically averaged matrix calculation, it is necessary to define how the order parameter components are obtained. In the case of Prompers et al. [230] they considered the scenario of individual interactions where the contribution of mode m to the decay of the correlation function of interaction j is the principal order parameter component:

$$\delta S_{j,m}^2 = \lambda_m(|m\rangle\langle m|)_{ij}$$

Equation 25 Principal order parameter component obtained from the contribution of mode m to the decay of the correlation function. $\delta S_{j,m}^2 \geq 0$, and $\sum_m \delta S_{j,m}^2 = 1$

The correlation function $C_j(t)$ can be expressed as a weighted sum of the normalized correlation functions $C_m(t)$, that is $C_m(0) = 1$, belonging to iRED modes m

$$C_j(t) = \sum_m \delta S^2_{j,m} C_m(t)$$

Equation 26 Correlation function expressed as a weighted sum of the normalized correlation functions $C_m(t)$, where m numbers all re-orientational eigenmodes, including overall and internal motions, leading to a complete de-correlation at long times $C_j(t \rightarrow \infty) = 0$. If $C_m(t)$ is mono-exponential, $C_m(t) = e^{\frac{-t}{\tau_m}}$, this is the case of isotropic re-orientation and it was found to be the case in good approximation for nearly all modes

From the correlation function described here the spectral density function $J_j(\omega)$ is obtained from analytical cosine-transformation based on the relation described in equation 1-9:

$$J_j(\omega) = \int_{-\infty}^{\infty} C_j(t) \cos \omega t \, dt = \sum_m \delta S^2_{j,m} \frac{2\tau_m}{1 + \omega^2 \tau_m^2}$$

Equation 27 Spectral density function expressed in terms of principal order parameter components $\delta S^2_{j,m}$ of the covariance matrix M and the correlation time τ_m .

Initially this approach was implemented to study NMR relaxation in ubiquitin [230] and recently it has been applied to study the dynamics of RNA loops [214, 215, 231]. The method was applied because the dynamics of RNA loops have large range of motions where the overall tumbling of the protein cannot be decoupled from the local bond vector dynamics. In the case of the simulations of lysozyme and ubiquitin presented here, the iRED method was implemented in order to account for loop regions where there is significant flexibility. Furthermore, the iRED method does not require the *a priori* assumption that the overall tumbling motions of the protein can be decoupled from the local N-H vector motions. This is because the correlation functions and ultimately the order parameter values are obtained from vector operations of an isotropically averaged covariance matrix of spherical coordinates. In other words, the dynamics of the N-H vectors are calculated with respect to all conformations sampled by the protein in spherical coordinate space. Although the decoupling assumption is not required for the calculation, it is possible that the correlation function information obtained decays to approximately a single exponential function (isotropic re-orientation) or several exponential functions (anisotropic re-orientation) as described by the Lipari-Szabo model. Furthermore, the correlation function may not decay in an exponential-like manner at all, but in another manner that approximates more

complex functions than exponential decay [217]. Therefore, the iRED method implements the concepts described by the model-free approach in a more robust way for the treatment of order parameters obtained from MD simulations; due to this robustness, this method was applied to the simulations described here.

2. Methods

2.1 Changes to the backbone dihedral function of ff99SB

In ff99SB the backbone dihedral term is expressed as a Fourier series of periodicity 4:

$$\sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \gamma_n))$$

Equation 28 Dihedral Torsional Term for AMBER force fields, where V_n is the dihedral force constant (amplitude of the cosine wave), φ is the dihedral angle of interest and γ is the phase of the angle

Where V_n is the dihedral force constant (amplitude of the cosine wave), n is the dihedral periodicity and γ_n is the phase of the dihedral angle θ. The backbone dihedral angle θ can be any of the canonical dihedral angles φ/ψ also known as general, the specific dihedral angles φ'/ψ' or the improper φ''/ψ''. The definition for each of these dihedral angles is given in figure 11 borrowed from the work of Hornak et al. [142].

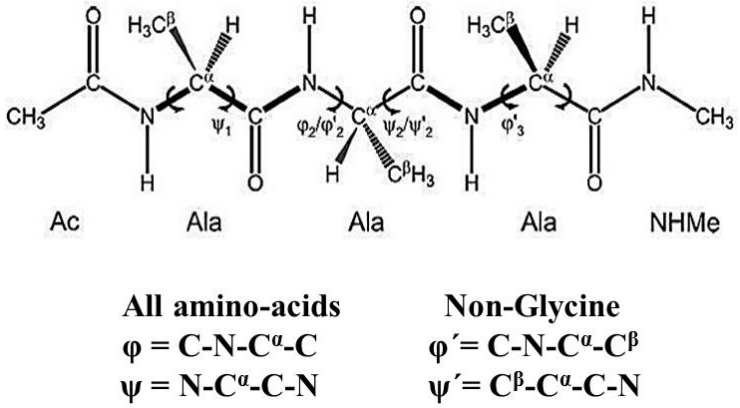


Figure 11 Backbone dihedral angle definitions for AMBER force fields. The scheme was taken from Hornak et al. used for the dihedral angle definition in the alanine tetra-peptide.

It should be noted that the φ'/ψ' angles have an offset of 120° between the dihedral planes in comparison to the canonical φ/ψ angles. The notion of this "extra" set of dihedrals is exclusive to AMBER force fields and was re-visited by Hornak et al. [142] when they applied the dihedral changes to ff99SB.

In order to describe how these changes are being applied to the backbone dihedrals of ff99SB we have plotted the functions for the φ' ψ and ψ' terms as this is crucial for the understanding of how things got modified, figure 13 illustrates the energy function for the φ' term of ff99SB.

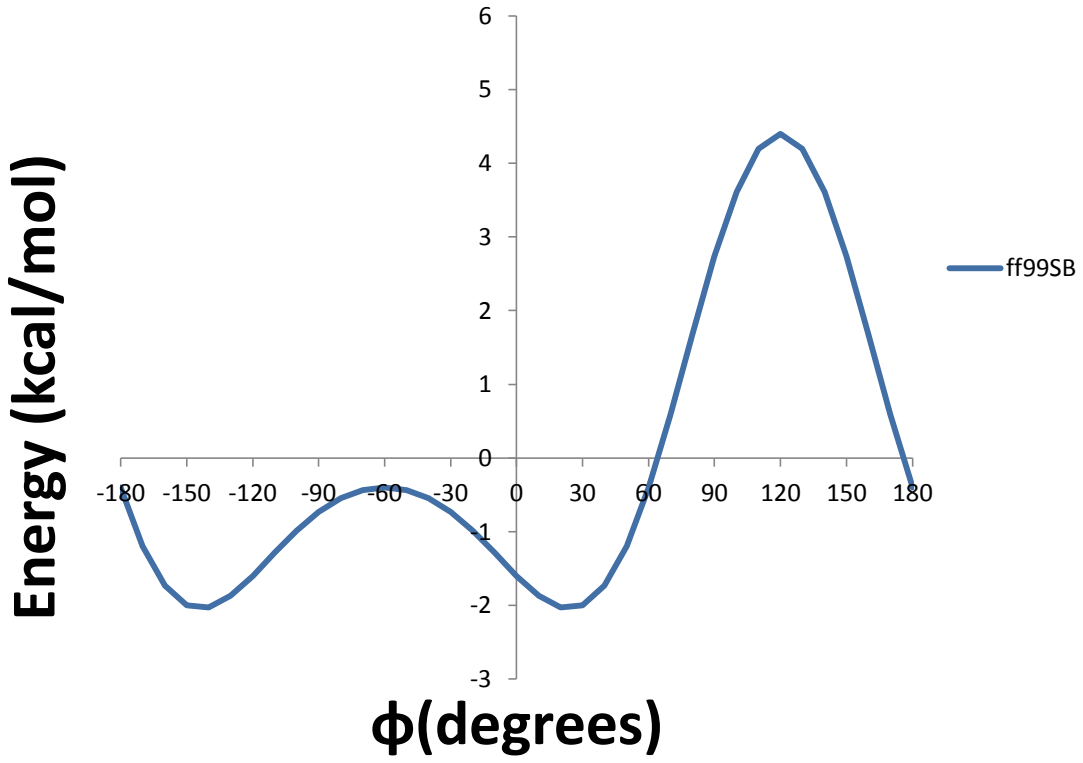


Figure 12 Dihedral energy functions for ff99SB force field

As previously described, the energy function for φ' is constituted by the sum of three cosine functions described in figure 13. Since the first term has the most contribution to the energy function, we began by applying the changes to the second and third terms by introducing small changes to the amplitudes of the cosine waves in order to make it more favorable for the sampling of the φ = -60º region.

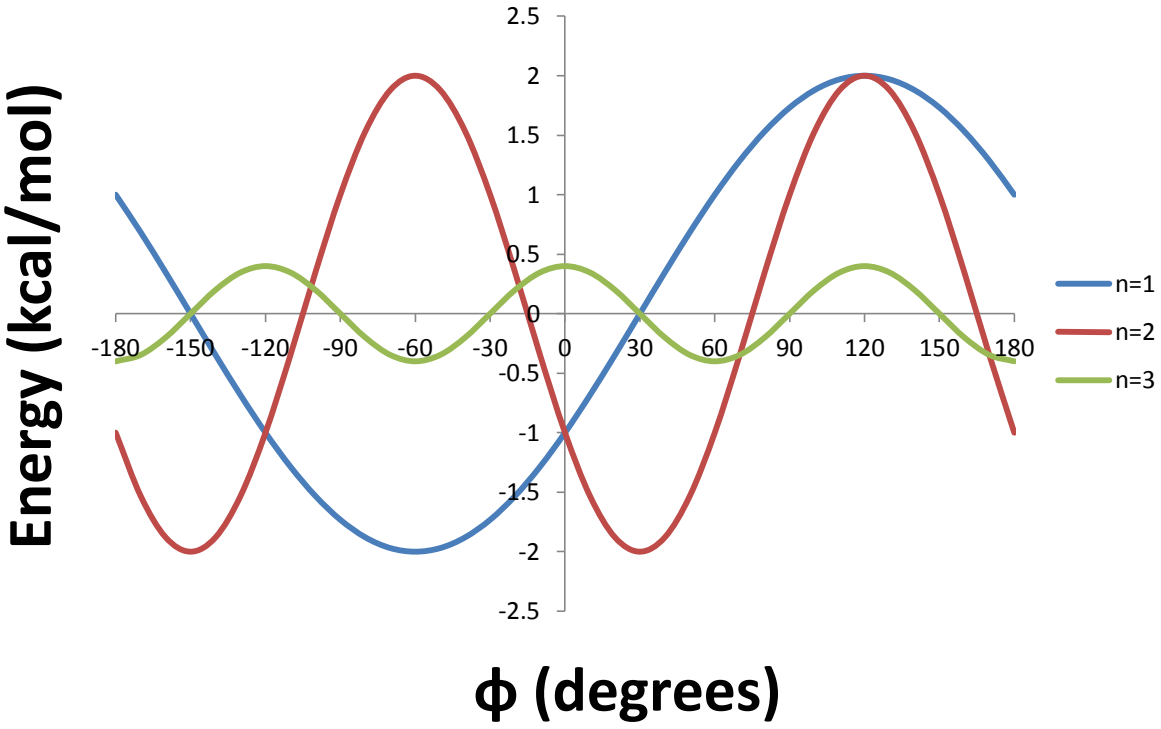


Figure 13 Cosine terms that constitute the φ' backbone dihedral term of ff99SB

The modifications were empirically derived by manually adjusting the φ', ψ and ψ' dihedral energy terms. Initially, an ensemble of structures was obtained from a simulation of ff99SB. Then, the dihedral angles and dihedral energies were measured and fitted to the backbone dihedral energy function. The dihedral angle information was used to calculate the vicinal ^{3}J (H_N, H_α) scalar couplings, in order to know how much the scalar coupling value differed from experiments. Lastly, the energy functions (φ', ψ and ψ') were empirically perturbed to yield a closer value. The dihedral energy term was modified by making small adjustments to the periodicities in the series .As previously described the first set of changes was applied to the φ' dihedral term, one of these first modifications was mod1φ. The amplitude of the cosine wave was changed by 0.2 and 0.4 kcal/mol in the second and third terms of the energy function in comparison to ff99SB (see figure 14-16).

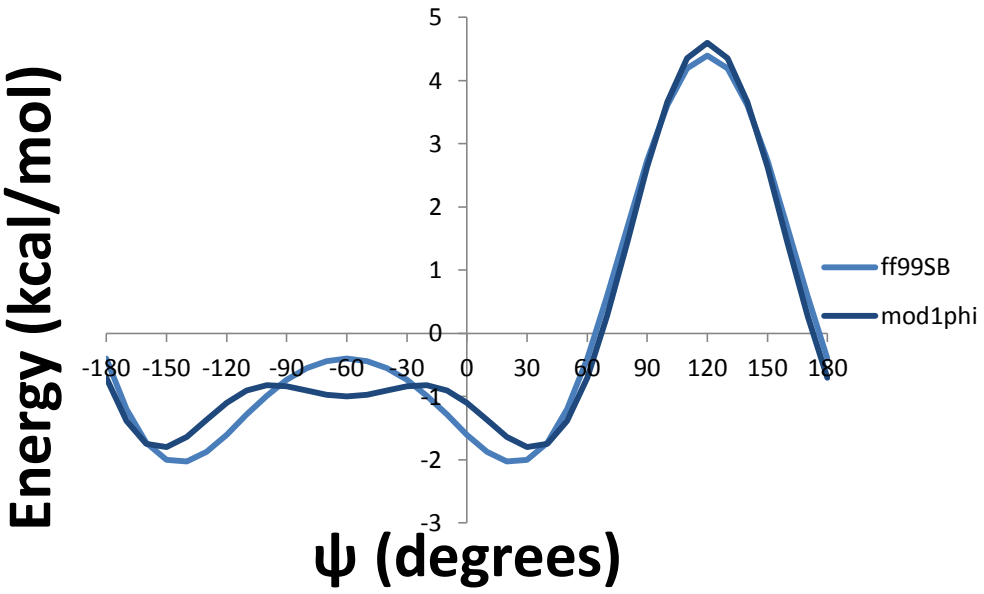


Figure 13 Dihedral energy function for ff99SB (blue) and mod1phi (dark blue) force fields

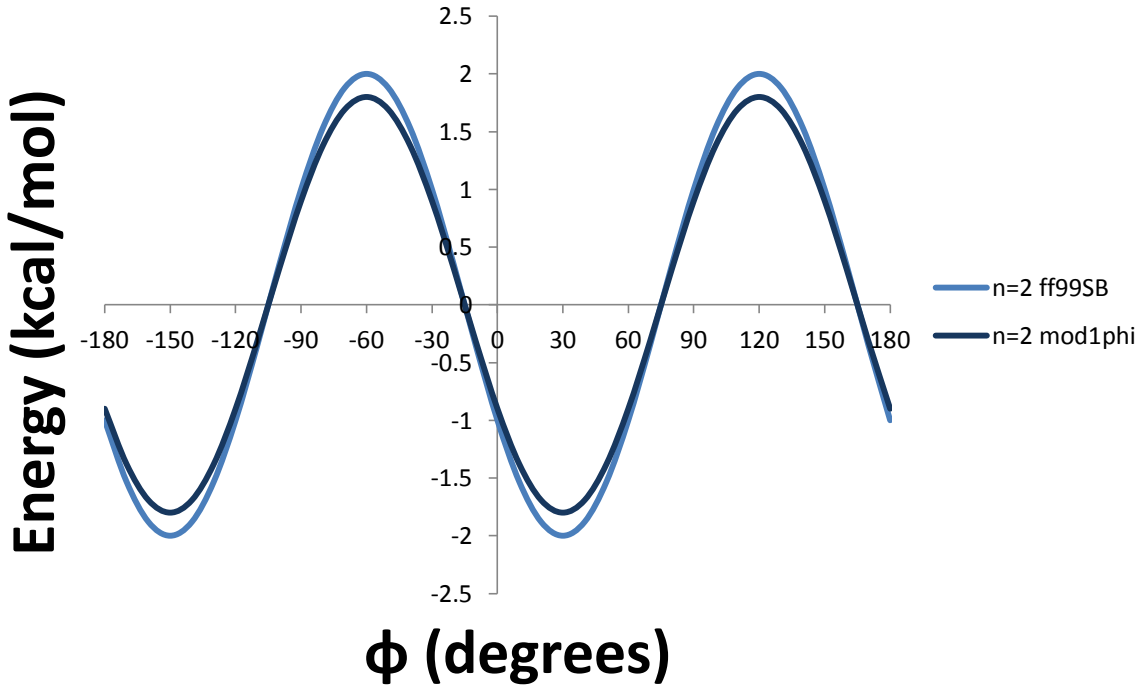


Figure 14 Applied changes to the second term of the φ' energy function for mod1φ

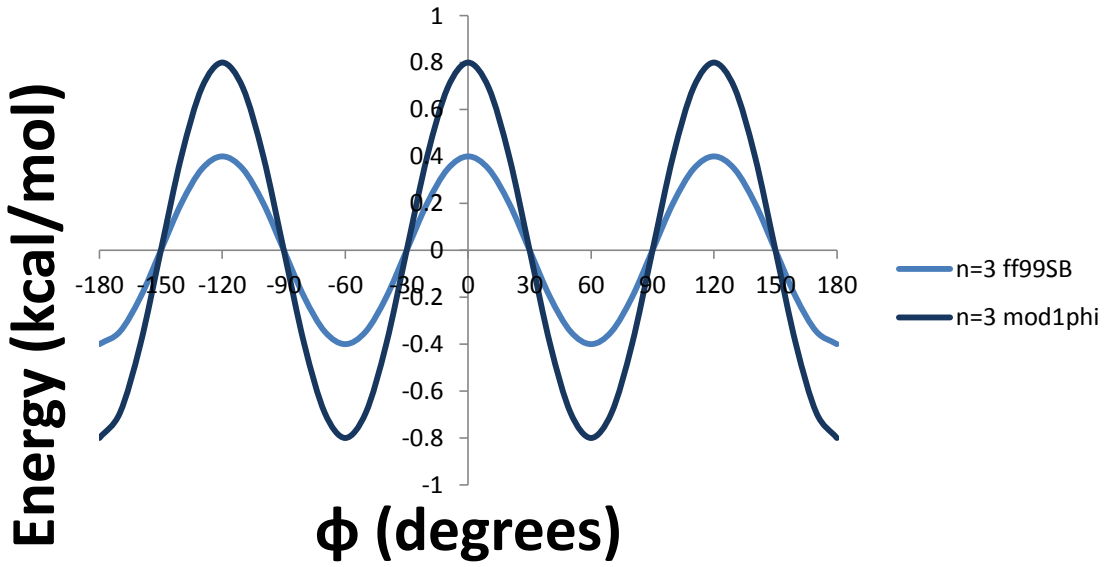


Figure 15 Applied changes to the third term of the *φ´* energy function for mod1*φ*

The energy differences for the different terms between the modifications were plotted for easy reference (see figure 17). Since the changes for the mod1*φ* and mod2*φ* for the second term of the cosine function were the same, we only plotted the energy differences for mod1*φ*; there were no changes in the second cosine term for mod3*φ*, mod4*φ* and mod5*φ*.

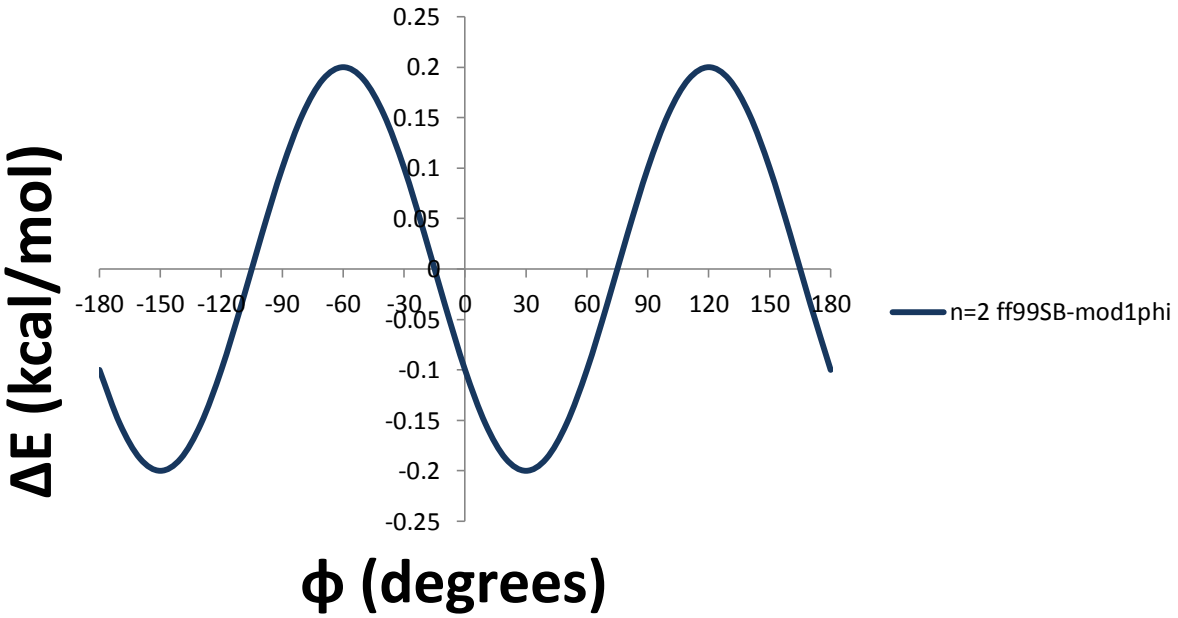


Figure 16 Energy differences for the second cosine term of the *φ´* energy function of ff99SB and mod1*φ*. The changes for mod2*φ* were the same, whereas the other modifications did not have any changes (mod3*φ*, mod4*φ* and mod5*φ*)

In the case of the third cosine term all the force fields had changes and are shown in figure 18. The differences between the curves clearly show how the φ = -60º is stabilized by each of the modifications, especially mod1φ.

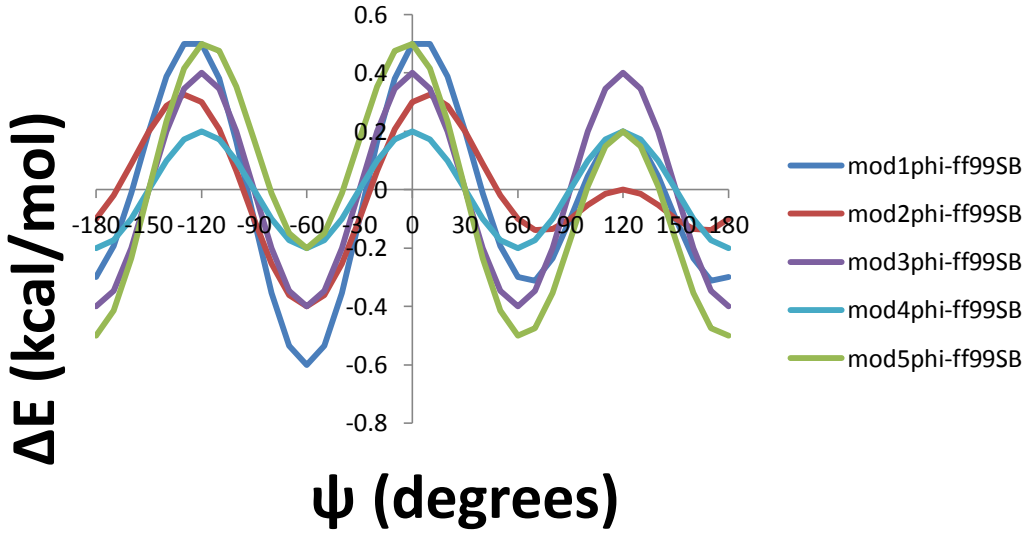


Figure 17 Energy differences between the modifications for φ' and ff99SB. The force field that stabilizes the φ = -60º region the most is mod1φ.

Besides these changes the mod5φ modification changed the first term of the cosine function from 2.0 to 1.8 kcal/mol. All the changes to the energy function are shown in figure 19.

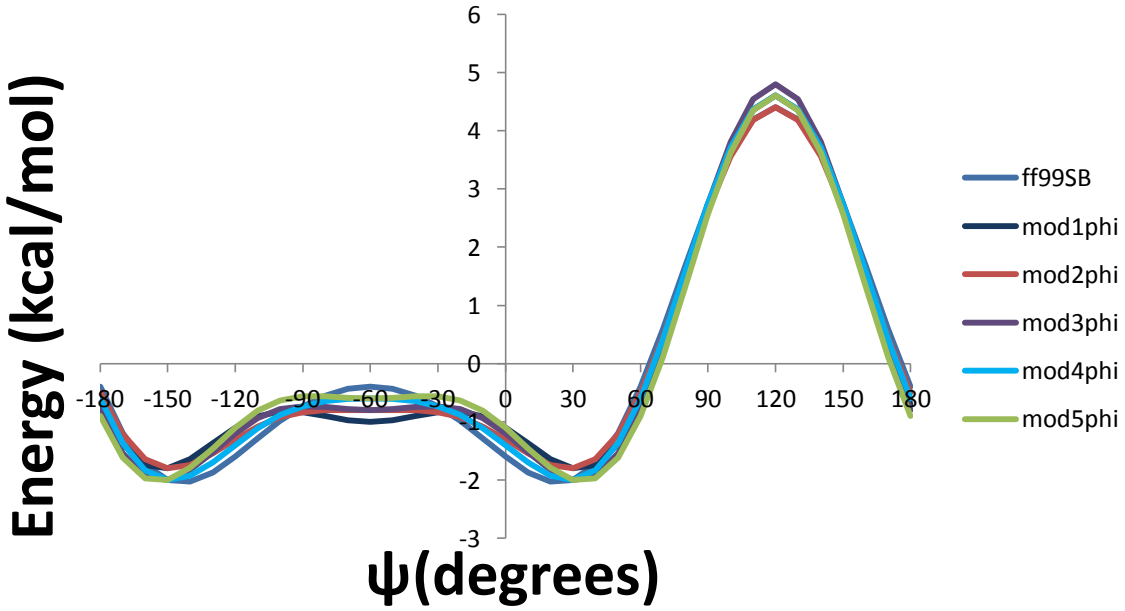


Figure 18 Changes to the dihedral energy function for φ'

Initially the modifications to the force field were only introduced to the φ' term, however when work by Best et al. was published [89] changes to the ψ and ψ' were tested as well. Their claim indicated that ff99SB was stabilizing β sheets; therefore the most reasonable adjustment would require modifying the ψ term in order to increase helical propensities.

In the case of the ψ angle, modifications were introduced to both the ψ and the ψ' terms of the function: mod1ψ, mod2ψ and mod3ψ had changes to the ψ and ψ' terms whereas mod4ψ only had changes to the ψ' term. The changes applied to mod1ψ, mod2ψ and mod3ψ were only to the first term of the ψ cosine function, where mod2ψ and mod3ψ had the same changes (from 0.45 to 0.6 kcal/mol). For mod1ψ the changes were from 0.45 to 0.7 kcal/mol. In the case of the ψ' term, the changes were applied to the first, second and fourth terms of the cosine function (see table 4). A fourth term of the cosine function was added for mod1ψ and mod2ψ (from 0 to 0.4 kcal/mol). Mod2ψ did not have any other changes to the ψ' term, while mod1ψ had changes to the second cosine term (from 0.2 to 0.5 kcal/mol). Mod4ψ only had changes in the first term of the cosine function (from 0.2 to 0.3 kcal/mol with an offset of -60º).

In order to clarify this we are plotting the cosine functions for ff99SB and modifications for ψ and ψ' independently (figures 20-23) and jointly (figure 24); we also plotted the energy differences between ff99SB and the modifications to ψ (see figure 25).

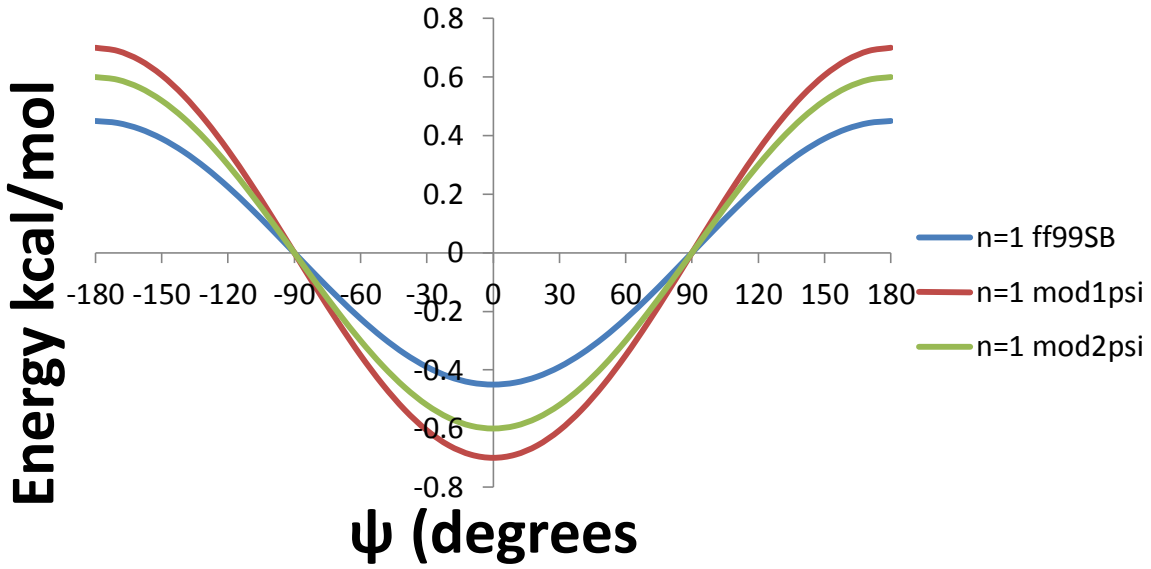


Figure 19 First cosine term of the ψ energy function of ff99SB, mod1ψ and mod2ψ. Mod3ψ has the same values as mod2ψ and mod4ψ has the same values as ff99SB.

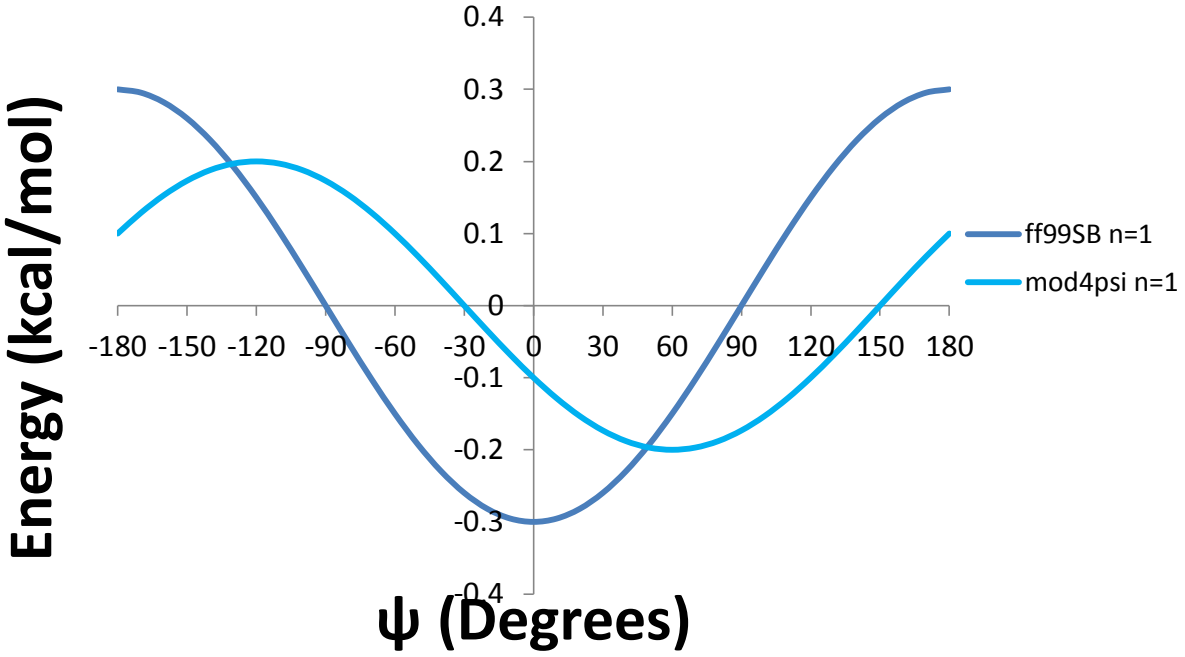


Figure 20 First cosine term of the ψ' energy function for mod4ψ. Mod1ψ, mod2ψ and mod3ψ have the same values for ff99SB.

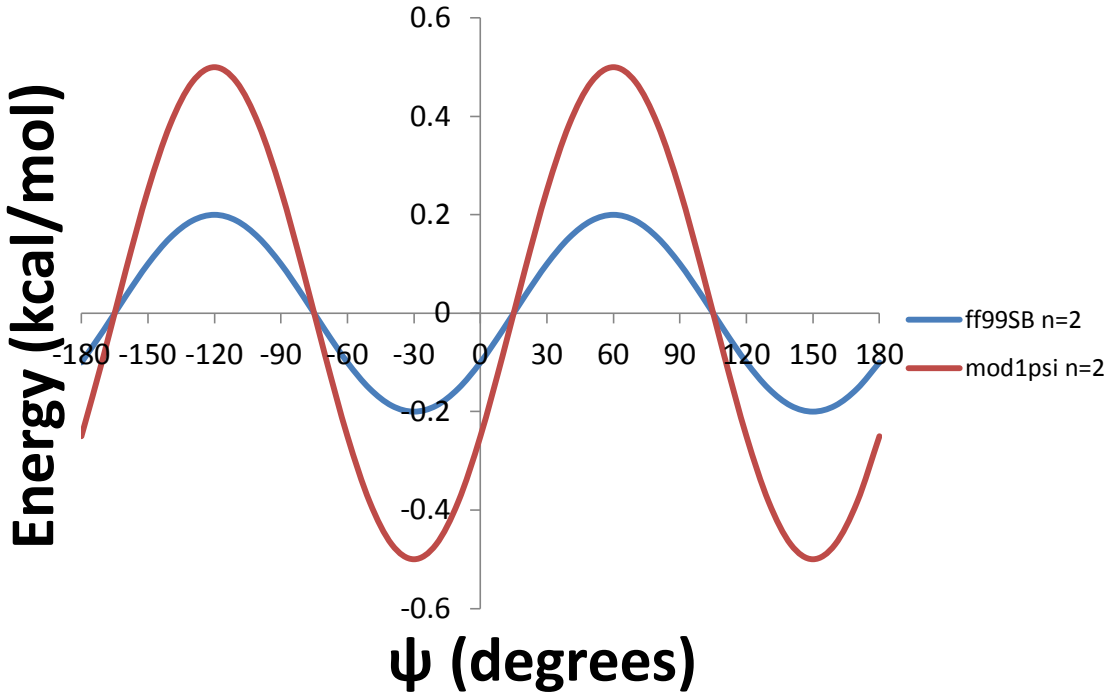


Figure 21 Second cosine term of the ψ' energy function for mod1ψ. Mod2ψ, mod3ψ and mod4ψ have the same values as ff99SB.

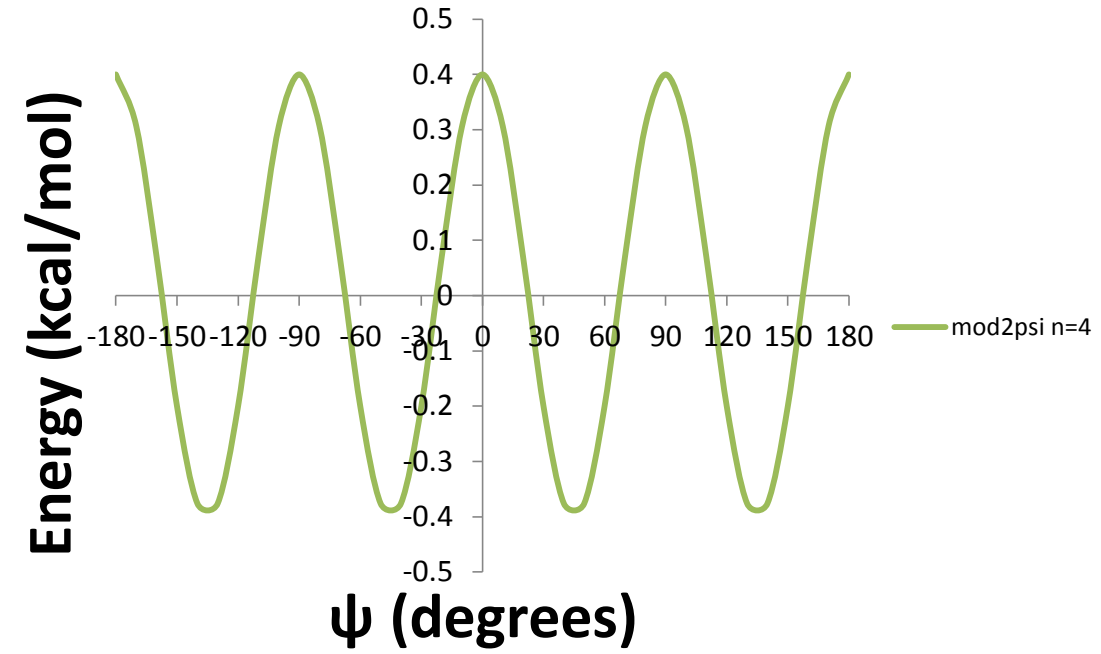


Figure 22 Fourth cosine term of the *ψ´* energy function for mod2*ψ*. Mod2*ψ* and mod3*ψ* have the same values, ff99SB does not have a fourth cosine term.

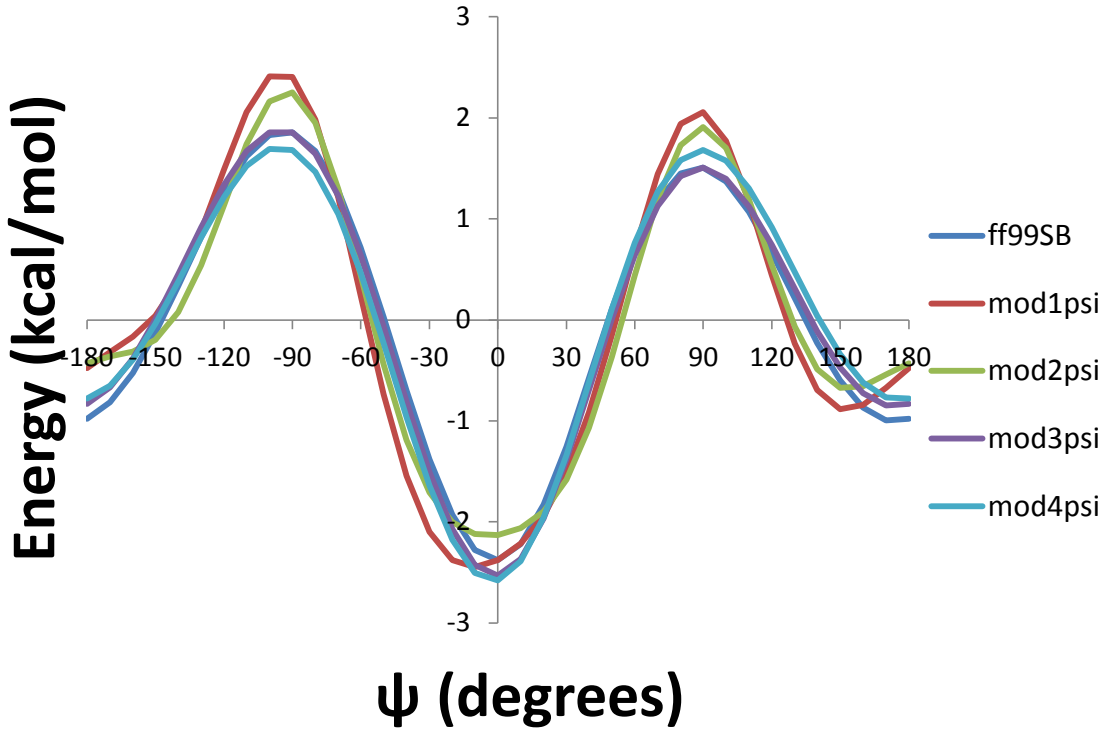


Figure 23 Dihedral energy functions of ψ for modifications to the ψ and ψ´ and dihedral parameters added

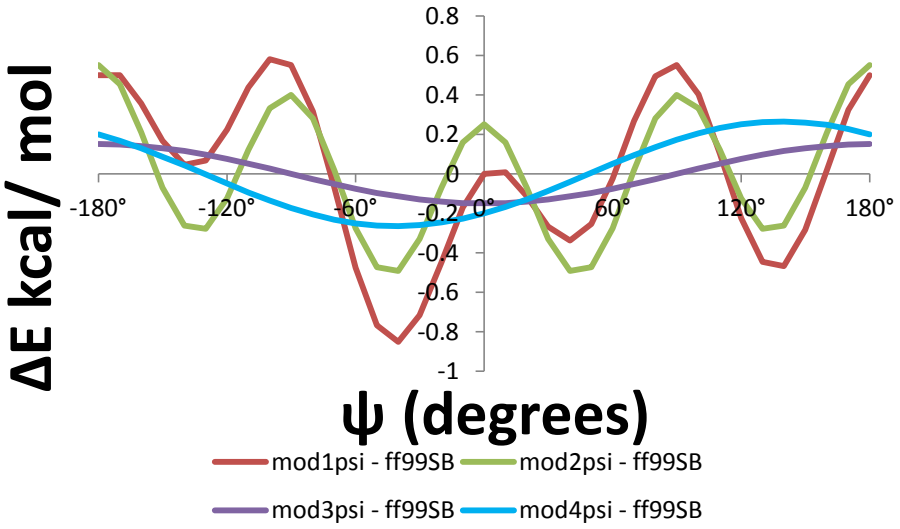


Figure 24 Energy differences between modifications to ψ/ψ′ and ff99SB

These modifications to the $\varphi'\psi$ and ψ' dihedral terms primarily focused on changing the amplitude of the functions whereas the changes to the ψ' dihedral term included phase changes as well. Initially, we tested thirty different force field combinations (see table 3) in which $\varphi'\psi$ and ψ' were modified in parallel with simulations of Ala_5 and Val_3. This initial testing was designed to choose the parameters that had the results with best agreement between experimental and simulation vicinal scalar couplings. We selected the force fields that had the best results for both systems. Then, we proceeded to test helical peptides such as Hydrogen Bond Surrogate Peptide (HBS) and K19. Again the force fields that had the best results were chosen for testing using NMR order parameters of Lysozyme and Ubiquitin. From this test we proceeded to test β secondary structure propensities for hairpins in implicit solvent. After this testing, there was one force field chosen for simulations of hairpins in explicit solvent, mainly mod1φ**.**

The modifications to the φ' term known as mod1φ, mod2φ and mod5φ were the most extensively tested throughout the systems because they had the best results through the study. This is because these modifications were initially conceived in order to increase the sampling of the ppII region in the Ramachandran plot. As previously described, our hypothesis was that these changes would inherently increase the sampling of the helical region as well, as the energy function was shaped to have more dihedral populations in the φ = -60º region.

The goal of the modifications to the ψ' term was to increase the sampling of the α-helical

region by stabilizing the $0 \geq \psi \geq -60^{o}$. The modifications that accomplished this in comparison to ff99SB were mod1ψ and mod2ψ. Stabilizing this region would further increase the helical region in the Ramachandran map which might have undesired effects on the rest of the secondary structure regions such as destabilizing the ppII region. This effect would not be desired because one of the goals of the work presented here is to increase the sampling of the ppII region in order to improve the agreement between experimental and simulated vicinal scalar couplings for poly-alanines which are considered to have ppII secondary structure.

Table 4 Modifications to the φ´, ψ and ψ´ terms of the force field

	φ'	C-N-C_α-C_β			ψ	N-C_α-C-N		ψ'	N-C-C_α-C_β	
ff	n	V_n	γ	ff	n	V_n	γ	n	V_n	γ
99SB	1	2	0	99SB	1	0.45	180	1	0.2	0
	2	2	0		2	1.58	180	2	0.2	0
	3	0.4	0		3	0.55	180	3	0.4	0
Mod1φ	1	2	0	Mod1ψ	1	**0.7**	180	1	0.2	0
	2	**1.8**	0		2	1.58	180	**2**	**0.5**	**0**
	3	**0.8**	0		3	0.55	180	3	0.4	0
								4	**0.4**	**120**
Mod2φ	1	2	0	Mod2ψ	1	**0.6**	180	1	0.2	0
	2	**1.8**	0		2	1.58	180	2	0.2	0
	3	**0.6**	0		3	0.55	180	3	0.4	0
								4	**0.4**	**120**
Mod3φ	1	2	0	Mod3ψ	1	**0.6**	180	1	0.2	0
	2	2	0		2	1.58	180	2	0.2	0
	3	**0.8**	0		3	0.55	180	3	0.4	0
Mod4φ	1	2	0	Mod4ψ	1	0.45	180	**1**	**0.3**	**-60**
	2	2	0		2	1.58	180	2	0.2	0
	3	**0.6**	0		3	0.55	180	3	0.4	0
Mod5φ	1	**1.8**	0							
	2	2	0							
	3	**0.8**	0							

Table 4 Modifications to AMBER ff99SB dihedral parameters. Changes highlighted in bold

2.2 Peptide/protein systems used to benchmark the data

In order to validate the modifications proposed we tested these parameters in peptides and proteins with diverse secondary structure in order to determine how these changes were affecting the dihedral propensities of the force field.

2.2.1 Testing the agreement between simulations and experiments for homologous peptides. Using Ala_5 as a model system

This peptide has been used as a model system for testing the ability of simulations to reproduce experimental NMR scalar coupling data [89, 151, 152, 172] for poly-alanine systems. Others have used analog peptides of glycine[172, 173], valine [162, 172, 173] and alanine[90, 172] residues for this purpose as these homologous peptides have become popular to benchmark the performance of force fields at reproducing NMR observables. One of the advantages of these systems is that its size makes them suitable for long molecular dynamics/Monte Carlo simulations in explicit water without the need of extensive computer power. Another reason is that in 2007 Graf et al.[172] measured vicinal scalar couplings that probe the φ and ψ backbone dihedral angle propensities and other NMR observables for homologous peptides of glycine (Gly_3), valine (Val_3) and Alanine (Ala_3-Ala_7). They also compared simulations with the GROMOS96 force field 43a149[232] with their experimental data.

At the time when the ff99SB parameters were published, the experimental data by Graf el al. [172] was not available for comparison, therefore the force field could not probed in this manner. Following up on this, Robert Best et al. used this information to benchmark the data of obtained from simulations of ff99SB/other variants of AMBER force fields, CHARMM-CMAP, OPLS-AA/L and some variants of the GROMOS force fields[152] using Ala_5 peptide as their test model. Their results portrayed ff99SB as one of the worst performers in the list, although later, they issued a correction with new results that indicated that the previous results were faulty and that ff99SB was one of the best performers[152]. Wickstrom et al. [151] also used this model system and its analogous Ala_3 to investigate the vicinal coupling results for ff99SB in TIP3P and TIP4PEw[233] water models. After this, Best et al. published revised parameters for ff99SB and ff03 which he denoted ff99SB* and ff03*; again using Ala_5 as one of its model peptides to benchmark the data. Later on, Nerenberg et al. [154, 162] also used the information for Ala_3 and Val_3 scalar values to probe their modified version of the ff99SB backbone parameters, as well as Li et al.[90] who used Ala_3 for the same purpose.

For the work described here, Ala_5 is being used because it has enough residues to model

the turn of a helix[89]. Since helical propensities for small peptides are primarily governed by dipole interactions, having the hydrogen bonding of a helical turn would define the dihedral populations for this secondary structure region more accurately than for shorter homologous peptides. Furthermore, it is known that poly-alanine peptides have mainly ppII character; although as the number of residues increases, the helical content increases as well[234]. This system was used to evaluate the quality of simulations against experimental data, in particular vicinal scalar couplings and dihedral populations. This information was used to discriminate between force fields that were not comparable to experimental information. Figure 26 shows the structure of the protonated Ala_5 peptide used for the simulations described here. This peptide was protonated in order to reproduce the low pH conditions described by Graf et al. in their publication [172]. The parameters for the protonated C-terminus were obtained from Wickstrom et al.[151]. This peptide was simulated using the Replica Exchange Molecular Dynamics method in order to generate converged statistics between the native and unfolded simulations.

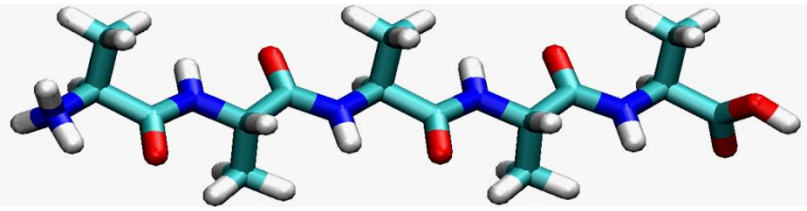


Figure 25 Protonated Ala_5 simulated structure

2.2.2 Ramachandran plots for residue 2 of Ala_5

We calculated the dihedral angle populations for residue 2 from the trajectories of Ala_5 at 299K and built free energy Ramachandran plots using equation 28. Where N_0 is the population of the most occupied bin calculated from the simulation and N_i is the population of bin i. Therefore the free energy of the most populated bin by definition would be zero.

$$G_i = -RT \ln\left(\frac{N_i}{N_0}\right)$$

Equation 29 Free energy differences equation for each bin of dihedral populations obtained from simulations with Ala5

2.2.3 Comparing vicinal scalar couplings for simulations and experiments of Ala_5

Experimental vicinal scalar couplings measure the spin-spin interaction effects of neighboring nuclei in molecules. This is very useful to measure structure in bio-molecules, in particular, secondary structure of peptides in order to determine if they are random coiled or ordered[168]. However vicinal scalar coupling values obtained from NMR data alone are not indicative of any ordered or disordered state unless they are interpreted with the help of Karplus functions[174]. The functions couple the scalar values to dihedral angles in a two dimensional manner, due to this relation it is possible for NMR experimentalists to convert their results to secondary structure dihedral maps. The relation is shown in equation 29:

$$J = A\,cos^2\theta + Bcos\theta + C$$

Equation 30 Karplus equation that establishes the relationship between scalar coupling values and dihedral angles where A, B and C are specific parameters for the specific scalar value measured and θ is the dihedral angle of choice plus a given offset.

Similarly, with the help of this function it is possible to calculate scalar coupling values from dihedral angle populations from simulations, by using the Karplus parameters of choice, the equation will give the scalar value populations. There are many Karplus parameters that probe for backbone dihedral angles [176, 180, 181, 235], although for purposes of the work described here, only the parameters derived by Case et al.[235] and Hu et al. are being discussed[178].

The use of vicinal scalar couplings to evaluate the performance of force fields is not new, Graf et al. [172] pioneered this type of analysis with their work, followed by Best et al.[152], Wickstrom et al.[151], Nerenberg et al. [173], Lindorff-Larsen et al. [158], Li et al.[90] and more recently Pande et al. [164]. Primarily this type of analysis has been applied to study secondary structure propensities and dihedral propensities of side-chains for emerging force fields against some of the older versions.

Although the calculation seems straightforward between angles and scalar couplings, the method has innate deficiencies. For example, the derivation of the parameters can be empirical, based on experimental data with uncertainties [235]. Furthermore, direct comparison from vicinal scalar couplings of experiments to simulation can also be difficult because of the nature of simulations vs. experiments. In other words, simulations generate trajectories of molecular

snapshots over periods of time, whereas NMR experiments capture one molecular snapshot at a given point in time where effects such as motional averaging are present. These must be taken into account when comparing directly to simulations where this effect is not intrinsically present[235].

2.2.4 Karplus parameters from density functional theory (DFT) quantum calculations

The parameters derived by Case et al. [235] were obtained from Density Functional Theory (DFT) *ab initio* calculations of two variants of alanine dipeptide, mainly Ace-Ala-Nme (DFT1) and Ala-Ala-NH_2 (DFT2). Parameters were derived for vicinal scalar couplings that probe for the backbone's φ/ψ angle; mainly ^3J (H_N, H_α), ^3J (H_N, C_β), ^3J (H_N, C'), ^3J (C', H_α), ^3J (C', C_β) and ^3J (C', C') for the φ angle and ^3J (H_α, N), ^3J (N, N) that probe for the ψ angle. The detail of the parameters derived is given in tables 4 and 5. The experimental work published by Graf et al.[172] measured mostly the same scalar values as the ones described by Case et al. with the exception of ^3J (C', C_β) for the φ angle and ^3J (H_α, N), ^3J (N, N) for the ψ angle. Instead Graf et al. [172] measured ^1J (N, C_α) and ^2J (N, C_α) for ψ.

Table 5 Karplus coefficients for DFT1 functions

Scalar Couplings	A	B	C	Offset
$^3J(H_N,H_\alpha)$	9.44	-1.53	-0.07	-60
$^3J(H_N,C')$	5.58	-1.06	-0.30	180
$^3J(H_\alpha,C')$	4.38	-1.87	0.56	120
$^3J(C,C)$	2.39	-1.25	0.26	0
$^3J(H_N,C_\beta)$	5.15	0.01	-0.32	60

Table 5 Scalar Coupling parameters for Karplus functions obtained from Density Functional Theory Calculations of Ace-Ala-Nme dipeptide by Case et al. The parameters shown probe for the φ angle.

Table 6 Karplus coefficients for DFT2 functions

Scalar Couplings	A	B	C	Offset
$^3J(H_N,H_\alpha)$	9.14	-2.28	-0.29	-64.51
$^3J(H_N,C')$	5.34	-1.46	-0.29	172.49
$^3J(H_\alpha,C')$	4.77	-1.85	0.49	118.61
$^3J(C,C)$	2.71	-0.91	0.21	-2.56
$^3J(H_N,C_\beta)$	4.58	-0.36	-0.31	58.18

Table 6 Scalar Coupling parameters for Karplus functions obtained from Density Functional Theory Calculations of Ala-Ala-NH_2. The parameters shown probe for the φ angle.

Case et al. [235] reported that most of the quantum-calculated Karplus functions (DFT1 and DFT2) agree well with empirically derived Karplus functions with the exception of $^3J(H_N,C_\beta)$ that couples φ at -60°. The deviations between the empirical and quantum Karplus functions can be as high as 3 Hz as shown in figure 27. The authors mentioned that these discrepancies are due to motional averaging effects and to the size/identity of the side-chain used for the quantum calculation. Calculations with serine and higher basis set gave better agreement between the empirical, DFT1 and DFT2 parameters.

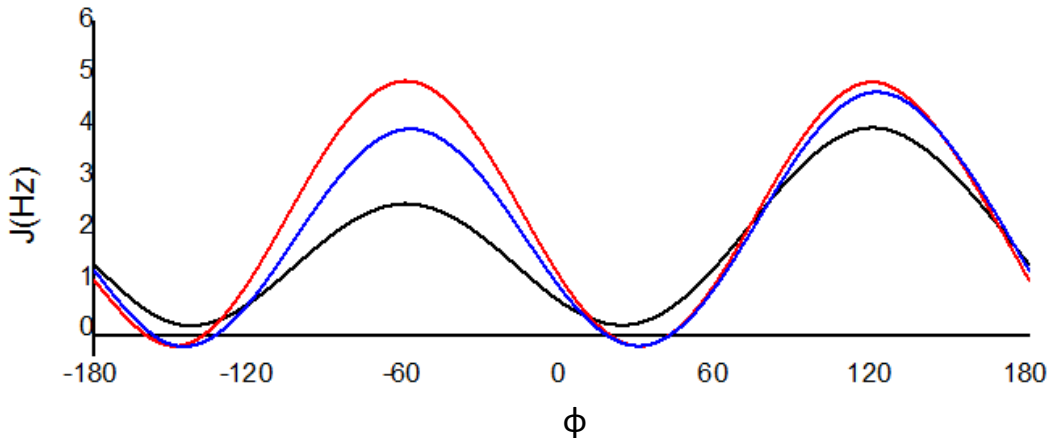


Figure 26 $^3J(H_N,C_\beta)$ Karplus Function. The black curve corresponds to the empirically derived parameters by Hu et al. [175] from ubiquitin, the blue curve corresponds to the parameters obtained from DFT calculations of Ala-Ala-NH_2 [235] dipeptide and the red curve corresponds to the parameters obtained from DFT quantum calculations of Ace-Ala-NMe dipeptide [235].

2.2.5 Karplus parameters from ubiquitin

These parameters were reported by Hu et al. [175] and Wirmer et al. [236]. They were obtained by empirically fitting the vicinal scalar coupling values from NMR spectra with dihedral angle populations from crystal structures of ubiquitin and staphylococcal nuclease. Hu et al. [175] reported parameters for $^3J(H_N, H_\alpha)$, $^3J(H_N, C_\beta)$, $^3J(H_N, C')$, $^3J(C', H_\alpha)$ and $^3J(C', C')$ that probe for the φ angle, whereas Wirmer et al. [236] reported the parameters for $^1J(N, C_\alpha)$ and $^2J(N, C_\alpha)$ that probe for ψ. Contrary to the case of the parameters obtained from the DFT calculations that only used dihedral profiles from alanine, these parameters are obtained from fitting to dihedral profiles of all the amino-acids present in ubiquitin and staphylococcal nuclease. Intrinsically, the fitting is more robust since effects of branched amino-acids on the backbone dihedrals are accounted for. Such is the case of the $^3J(H_N, C_\beta)$ value that couples φ at -60° and is not well described by the Case et al. [235] parameters from DFT calculations. However, these fittings are "empirical" and have "uncertainties" within that need to be accounted for in the calculations.

For the work described here, these parameters will be termed "Original" parameters as Best et al. [152] and Wickstrom et al. [151] had denominated them from their published work. The details of the parameters are given in table 7.

Table 7 Karplus coefficients for calculations of scalar values also known as original parameters with uncertainties (σ_i) from experiments

Scalar Couplings	A	B	C	Offset	σ_i
$^3J(H_N,H_\alpha)$ [175]	7.09	-1.42	1.55	-60	0.91
$^3J(H_N,C')$ [175]	4.29	-1.01	0.0	180	0.59
$^3J(H_\alpha,C')$ [175]	3.72	-2.18	1.28	120	0.38
$^3J(C,C)$ [175]	1.36	-0.93	0.60	0	0.22
$^3J(H_N,C_\beta)$ [175]	3.06	-0.74	0.13	60	0.39
$^1J(N,C_\alpha)$ [236]	1.70	-0.98	9.51	0	0.59
$^2J(N,C_\alpha)$ [236]	-0.66	-1.52	7.85	0	0.50

Table 7 Karplus coefficients parameters reported by Hu et al. [175] and Wirmer et al. [236] obtained from NMR and crystallographic data of Ubiquitin and Staphylococcal Nuclease. $^1J(N,C_\alpha)$ and $^2J(N,C_\alpha)$ are the only scalar couplings that probe for ψ angle. The other scalar couplings probe for φ angle

2.2.6 χ^2 value calculation for vicinal scalar couplings

Best et al. [89] used this calculation in order to determine how well AMBER, OPLS and GROMOS force fields were reproducing experimental scalar coupling values for Ala_5. He also used this calculation to validate his modifications of ff99SB and ff03 named ff99SB* and ff03*. Wickstrom et al. [151] used as well for her validation of ff99SB simulations with TIP3P and TIP4PEw water models; measuring scalar couplings for Ala_3 and Ala_5 peptides. Nerenberg et al. [154] did the same as Best et al., although they included calculations for Ala_3, Gly_3 and Val_3 peptides in their report and Li et al. [90] used this calculation for validation of their force field including only calculations for Ala_3 peptide. As previously done we used this calculation to evaluate the quality of the modifications proposed here as previously reported. The details of the χ^2 value calculation are included in equation 30. Since the DFT1 and DFT2 parameters do not have Karplus coefficients for vicinal scalar couplings that probe the ψ angle, in order to keep the calculations even for all the cases of parameters (DFT1, DFT2 and Original) the original coefficients that probe for the ψ angle were used to calculate χ^2 values for all parameters. In other words, χ^2 values were generated for each of the Karplus parameters (DFT1, DFT2 and Original) independently, using the original coefficients for the calculations of the ${}^1J(N,C_\alpha)$ and ${}^2J(N,C_\alpha)$ scalar couplings that probe for ψ in all cases. Besides these scalar couplings mentioned here, there was an additional vicinal scalar coupling value that was included in the calculation known as ${}^3J(H_N,C_\alpha)$. This scalar coupling couples the φ and ψ angles and its definition is two dimensional. The coefficients for this scalar coupling were borrowed from the work of Hennig et al. [237], who used NMR data for the fitting. Since Graf et al. included it in their report, Best et al. and Wickstrom et al. included it in their work as well. The details for the parameters of this "special" Karplus function are explained in table 8.

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{(\langle J_i \rangle_{sim} - J_{i,\mathrm{expt}})^2}{\sigma_i^2}$$

Equation 31 χ^2 calculation to evaluate the agreement between results obtained from simulations and experimental vicinal scalar couplings. N is the number of vicinal scalar couplings measured (7 in this case). $\langle J_i \rangle_{sim}$ is the average of scalar value obtained from the simulation trajectory. $J_{i,\,expt}$ is the scalar value from experiments and σ_i is the uncertainty for the scalar values from experiments

Table 8 Karplus coefficients for $^3J(H_N,C_\alpha)$ scalar coupling that probes for φ and ψ angles

Scalar Coupling	A	B	C	D	E	F	G	H	I
$^3J(H_N,C_\alpha)$	-0.23	-0.20	0.07	0.08	0.07	0.12	-0.08	-0.14	0.54

Table 8 Karplus coefficients for the $^3J(H_N,C_\alpha)$ vicinal scalar coupling that couples φ and ψ. The equation for the Karplus function is given by $^3J = A\cos(\varphi) + B\cos(\psi_{i-1}) + C\sin(\varphi) + D\sin(\psi_{i-1}) + E\cos(\varphi)\cos(\psi_{i-1}) + F\cos(\varphi)\sin(\psi_{i-1}) + G\sin(\varphi)\cos(\psi_{i-1}) + H\sin(\varphi)\sin(\psi_{i-1}) + I$

2.2.7 Using ^3J (H_N,H_α) scalar coupling for the initial assessment of the quality of modifications

The $^3J(H_N,H_\alpha)$ vicinal scalar coupling is one of the few vicinal scalar couplings that has a wide range. This has made it useful for the initial assessment of secondary structure propensities because it makes it easier to distinguish between helical/ppII conformations and β conformations while probing the φ angle distributions. It has a wide range of values between 2 and 11 Hz depending on the Karplus parameters chosen for the fitting. This wide range allows for clear distinction of specific regions in φ dihedral space.

As previously discussed, the results from the simulations of AMBER ff99SB for residue 2 in Ala_5 indicated that the scalar coupling value was 6.99 Hz whereas the experimental value is 5.59 Hz. This result is indicative that the simulations are considerably sampling high scalar coupling values. As shown in figure 3, the dihedral populations in the $-150° \leq \varphi \leq -120°$ region are contributing to this large value. In order to reduce the average value of the scalar coupling close to 5 Hz, it was necessary to increase the sampling in the φ = -60° region. From the $^3J(H_N,H_\alpha)$ values it is possible to roughly infer how the other scalar coupling values will behave for the rest of the modifications.

2.2.8 Secondary structure basin definitions

In order to determine how the secondary structure propensities were being affected by the modifications, we calculated the basin populations from simulations according to the definition previously reported by Wickstrom et al. [151]. Similar calculations were reported by Graf et al.

[172] and Best et al. [89] for their work. Although the definitions of their secondary structure basins were different because Graf et al. used the scalar coupling values obtained from simulations to estimate the populations whereas Best et al. used almost the same definition described here with the exception of the extended β-strand conformation, (-180° to -90°, 50° to 240°; or 160° to 180°, 110° to 180°) The definitions of the four principle regions for the work described here were as follows: right handed helix (α_R), (φ,ψ) ~ (-160° to -20°, -120° to +50°); extended β-strand conformation, (-180° to -110°, 50° to 240°; or 160° to 180°, 110° to 180°); and poly-proline II, (-90° to -20°, 50° to 240°). The number of structures in individual regions were summed and divided by the total number of structures. Then, they were multiplied by 100 to get the percentages in each basin. Error bars were taken from independent runs.

2.2.8.1 Secondary structure free energy maps for alanine dipeptide

Following up on the information provided by the secondary structure basin information, free energy maps were calculated for some of the modifications using umbrella sampling simulations for alanine dipeptide (N-acetyl alanyl-N-methyl-amide). Small peptides such as this have been previously used for parameterization of force fields as in the case of Cornell et al. ff94 [134] and Kollman et al. ff96 [144].

In order to quantify in terms of energy the results for the secondary structure basins from the Ala_5 simulations, potential mean force maps were calculated for some of the best modifications and ff99SB. This information was useful to determine how the dihedral maps were changing with the different parameters. These results were compared against the results of ff99SB to see how the different regions were changing between modifications.

2.2.9 Using Val_3 peptide to compare vicinal scalar couplings between simulations and experiments

In order to further investigate the performance of the force fields in another small system analog to Ala_5, we tested all modifications in a small peptide known as Val_3 (figure 28) for which experimental data was also published in the work of Graf et al. [172]. Simulations for this peptide were previously reported by Nerenberg et al. [154] intended to validate their

modification to the AMBER ff99SB backbone dihedral parameters. This peptide is useful to investigate how the secondary structure propensities are being affected by the modifications proposed here because this peptide is considered to have extended β character according to previously documented experimental findings [195] [196]. Since the proposed modifications are designed to increase the sampling in the ppII region, the simulations from this peptide will help us to indicate how the modifications are affecting the β region.

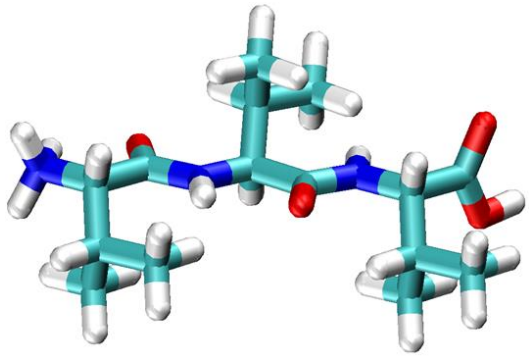


Figure 27 Val_3 protonated structure used for simulations in order to compare to the experimental data published by Graf et al. [172].

One of the assumptions for the modifications of the φ' torsional potential based on scalar coupling values of Ala_5 is that all amino-acid residues have similar φ angle dihedral behavior. In other words, fitting the energy function in order for the simulation results to match the dihedral behavior of poly-alanines should also improve the agreement of the dihedral behavior for other amino-acids; in this case, valine. Although this is partially true, the effect of side-chain dynamics in the backbone conformations cannot be neglected and is not addressed in the fitting presented here. Therefore, the results obtained might be indicative of this behavior. The vicinal scalar couplings studied and the methods used for this peptide are the same as in the case of Ala_5 discussed previously [172]. Like Ala_5, the peptide is protonated in order to reproduce the experimental conditions described by Graf et al. [172].

2.3 Evaluating helical propensities for modifications

One of the goals of this thesis is to determine if the proposed modifications are appropriate to address the issues of helical propensities for AMBER ff99SB. In order to do this, the modifications that have the best results for the poly-alanine peptides were tested against two model helical peptides. To our knowledge these model peptides have not been used for this

purpose before, however they have experimental data that was used to estimate how the helical propensities were behaving. Initially, a small helical peptide known as the Hydrogen Bond Surrogate peptide [168] was evaluated, then the testing was extended to a longer helical peptide known as K19 [167].

2.3.1 Hydrogen Bond Surrogate (HBS) peptide

The Bcl-2 protein family has an essential role in the regulation of cell apoptosis. Some members of the family, in particular Bcl-2 and Bcl-x(L) inhibit cellular apoptosis, whereas others such as Bak and Bax stimulate cell death. It is known that the BH_3 domain of the Bak protein is responsible for regulating binding to other pro-apoptotic members of the family such as Bax, Bik and Bad [238]. Furthermore, truncated sequences of the BH_3 domain have been reported to be enough to induce apoptosis in transfected cells [239].

Following up on this notion, Wang et al. [168] generated truncated sequences of the BH_3 domain consisting of only ten amino-acid residues. However, stabilization of short peptides like this can be challenging due to the energy penalty of helix nucleation [240, 241]. Short peptides that are made out of 10 residues or less are in essence unstable due to low helix nucleation probability [242, 243]. In order to overcome this, the peptide sequences were stabilized by introducing a 1C=1C′--5C-5N covalent bond between residues 1 (cap) and 5 (alanine) at the N terminus as seen in figure 29 and figure 30. The idea behind this approach was to generate a stable helical domain by promoting a pre-organized helical turn that would increase the nucleation probability and mimic the function of Bak-BH_3 domain which interacts with the Bcl-x(L) target protein.

In fact it was demonstrated that this was the case when experimental binding assays were performed [244]. This approach has been applied to other protein drug targets successfully as well [245, 246] [247, 248].

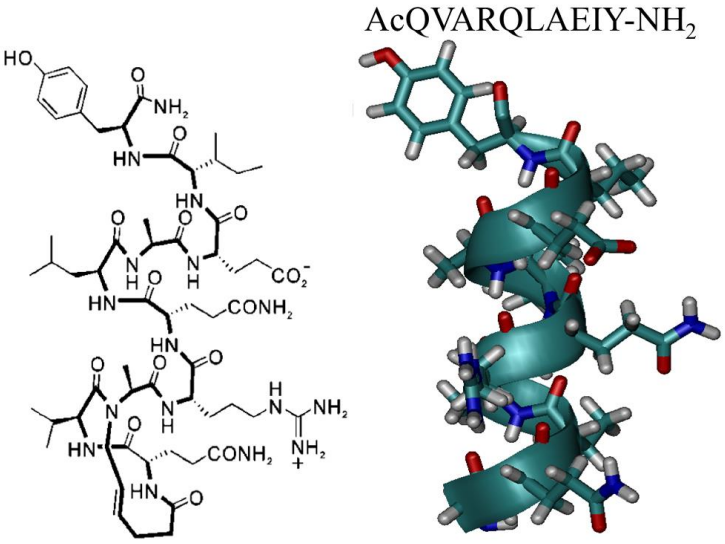


Figure 28 Left panel corresponds to Hydrogen Bond Surrogate (HBS) Peptide as described by Wang et al. [168]. Left panel corresponds to the experimental sequence and the right panel corresponds to the simulated sequence AcQVARQLAEIY-NH_2.

For the purpose of the work described here this peptide was used to evaluate helical propensities for the modifications proposed. This peptide was chosen because of its size that makes it suitable for Replica Exchange Molecular Dynamics and because it has ^3J (H_N-H_α) vicinal scalar coupling data available. Replica Exchange Molecular Dynamics was chosen for the simulations because it is a method that enhances conformational sampling [249-251].

The covalent bond described by experiments was not implemented in the simulations because it would require validation of the charges along with the backbone parameters involved in the C-C covalent bond described by Wang et al.[168] as seen in figure 30. This would add more unknowns to the fitting. In order to overcome this, bond angles and bond distances were restrained harmonically to covalent bond parameters of AMBER force fields. The length of the bond between the cap and Alanine 5 (1O-5H) was restrained to 1.522Å with a weight of 317 kcal $mol^{-1}Å^{-2}$. The 1C=1O-5H angle was restrained to 120° with a weight of 80 kcal mol^{-1} rad^{-2} and the 1O-5H-5N angle was restrained to 110.1° with a weight of 63 kcal/mol/rad^2.

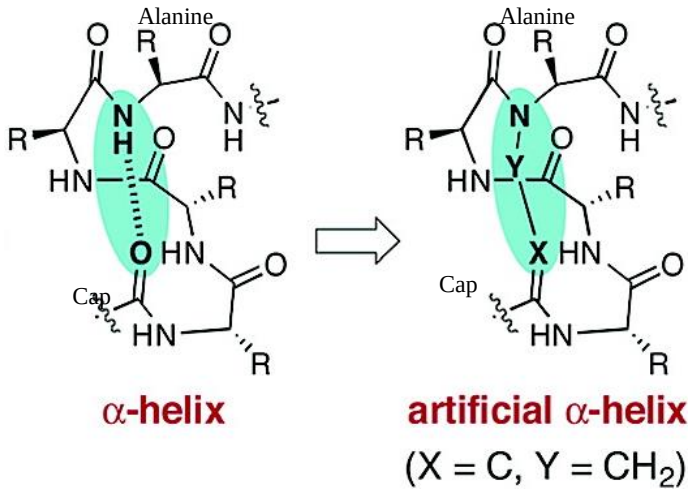


Figure 29 Left panel: Canonical hydrogen bond pattern in alpha-helices. Right panel: Hydrogen Bond surrogate pattern. Image taken from [252]

Besides motional averaging effects intrinsic to the nature of results obtained from experiments, direct comparison between experiments and results from simulations can be challenging; especially in this case, where the covalent bond description from experiments was not implemented in the simulations. Nevertheless, the helical propensities for the other residues that are not part of the Hydrogen Bond Surrogate (HBS) region should follow the same trends as in the experiments.

Previously there was a molecular dynamics study dedicated to investigate the binding mechanism of HBS helix to Bcl-x_L in comparison to the binding mechanism of wild type Bak to Bcl-x_L [253]. However, protein folding studies have not been reported for this peptide to our knowledge. Furthermore, this peptide has not been used for the validation of force fields parameters recently, and no previous simulation information is available that is comparable to what is proposed here. Previous work such as the one published by Best et al.[89] used Ac-$(AAQAA)_3$-NH_2 peptide, a small fragment of lysozyme known as 19-HEWL, GB1 hairpin, Trpcage, Villin HP35 and Pin WW to validate their force fields[89, 159-161]. Pande et al. [164] used dipeptides, tri-peptides, tetra-alanine and ubiquitin to investigate how the force fields reproduced NMR observables. Nerenberg et al. [154] also used small tri-peptides for the validation of their force field along with ubiquitin. The Ac-$(AAQAA)_3$-NH_2 could have been used for the validation of the helical propensities as Best et al. had done[89], however this

peptide is quite homologous and does not have scalar coupling information available for comparison. The HBS peptide has this information available and it also has more diversity of amino-acids since it would be highly desirable to evaluate the helical propensities of the modifications for more amino-acids other than alanine.

2.3.2 K19 peptide

In order to further validate the helical propensities of the modifications proposed, it was necessary to study another helical peptide that would be longer than the HBS peptide and that would not require bond constraints as previously described for this peptide system. Previously the work by Graf et al.[172] and Best et al. [89] used the 19-mer of Hen Egg White Lysozyme (19-HEWL) for the validation of their force field. This peptide was particularly useful for them because Graf et al. [172] measured scalar coupling information available that they could compare against their simulation results. They also used the Ac-$(AAQAA)_3$-NH_2 peptide as a benchmark for their validation of force fields. Furthermore, designed alanine rich peptides in which polar amino-acid residues are introduced into the sequence are popular to study helix formation experimentally [242, 254-259] and computationally [141, 259-261].

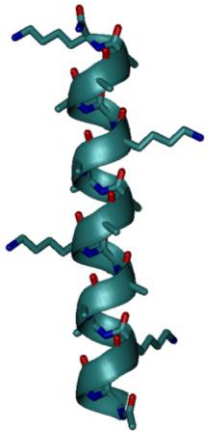
These two peptides used by Graf et al. and Best et al. could have been used for the validation of the parameters described here as well, however only the information provided by Graf et al. [172] was available for comparison and validation. It was decided to use another peptide known as K19 for which more experimental information (other than NMR) and simulation data was available[167].

K19 is an alanine rich peptide with transient helical nature, in which Lysine residues have been introduced in order to make it more soluble[167]. Furthermore it was reported that this type of peptides benefit from glycine N-capping by enhancing fractional helicities greater than 40% at any position in the sequence [167]. This effect was not as marked when C-capping was introduced based on the data obtained from experiments however the effect was more noticeable from simulation results [167]. Apparently when the cap at the N-terminus is removed, helix fraying appears due to the lack of hydrogen bonding that supports helix formation. The

introduction of polar amino-acids, in particular positive side-chains at the C-terminus was also investigated; it showed enhanced helicity for this peptide due to favorable interactions with the helix macro-dipole. Seemingly charged side-chains protect the backbone's hydrogen bonds from contact with water [141].

This peptide was chosen in order to evaluate the helical propensities in a larger system for which experimental information is available (CD and NMR), without the need to impose distance and angle restraints as in the case of the Hydrogen Bond Surrogate Peptide (HBS). This peptide has been studied previously with implicit solvent simulations [167]. However in the work described here, the simulations were done with explicit TIP3P solvent model.

The structure of the K19 peptide is described in figure 31



$Ac\text{-}GGG(KAAAA)_3K\text{-}NH_2$

Figure 30 K19 Peptide

2.2.2.1 Definition of helical content for HBS Peptide

Wang et al [168] defines helical content as percent helicity which was measured from CD experiments as the ratio $[\theta]_{222}/[\theta]_{max}$, where $[\theta]_{max}$ = -23,400. The concentration was determined by monitoring the absorbance of the tyrosine residue under denaturing conditions. In the case of K19, the experimentalists measured Chemical Shift Deviations (CSD) in order to calculate fractional helicities. Although it is possible to use chemical shift libraries for prediction measurements from simulations, the assignments for the peaks are usually done with software

packages that are based on empirical assignment from these libraries. Furthermore, these software packages are trained to predict chemical shifts which are taken from NMR ensemble averages; therefore the method can become less sensitive to conformational variations that contribute to the experimentally measured average. In other words, the software package can prefer the static experimental conformation and incorrectly predict chemical shifts from conformational dynamics from simulations [262]. Since we cannot measure how the backbone of the protein absorbs polarizable light through its amide bonds, or the measurement of chemical shift deviations from simulations is currently a method that has not been extensively validated, we ought to use another way to quantify helical content in a matter measurable from simulations.

In order to solve these issues, we used an alternative helical content definition that was defined as the amount of residues that adopt helical hydrogen bonding as defined by Dictionary of Protein Secondary Structure (DSSP) analysis [183]. Only the residues listed as α-helical (H) or 3_{10} helical (G) were included in the calculation of the average. DSSP defines the intra-backbone hydrogen bonds of the protein through electrostatics, assuming partial charges of -0.42 *e* and +0.20 *e* to the carbonyl oxygen and amide hydrogen, their opposites assigned to the carbonyl carbon and amide nitrogen. A hydrogen bond is identified if *E* in the following equation is less than -0.5 kcal/mol:

$$E = 0.084\left\{\frac{1}{r_{ON}} + \frac{1}{r_{CH}} + \frac{1}{r_{OH}} + \frac{1}{r_{CN}}\right\}.332\ kcal/mol$$

Equation 32 Intra-backbone hydrogen bond energy definition given by DSSP analysis.

One of the basic patterns for secondary structure definition is the turn. A residue is assigned as turn if it has a hydrogen bond from the oxygen of the C=O of amino-acid *i* to the hydrogen of the H-N of amino-acid *i*+*n*; where n = 3, 4, 5. There are three types of turns: turn 1 where the backbone dihedral angles of amino-acid *i* are (-60 = φ, -30 = ψ) and the dihedral angles for amino-acid *i*+1 and *i*+2 are (-90 = φ, 0 = ψ). Turn 2 where the backbone dihedral

angles of amino-acid *i* are (-60 = φ, 120 = ψ) and the dihedral angles for amino-acid *i*+1 and *i*+2 are (80 = φ, 0 = ψ), and turn 3 where the backbone dihedral angles of amino-acid acid *i* are (-60 = φ, -30 = ψ) and the dihedral angles for amino-acid *i*+1 and *i*+2 are (-60 = φ, -30 = ψ). Turn 3 is considered a single unit of a helix. A minimal helix is defined by two consecutive type 3 turns [183]. Longer helices are defined as overlaps of minimal helices, in the case described here, only the overlaps of 3-helices (3_{10}) and 4-helices (α) are considered for the calculation.

The secondary structure type was calculated for every residue per picosecond, with the exception of the 1st, 2nd and last residue of the peptide. These residues are excluded from the calculation because in the case of the 1st and 2nd residues they don't have preceding residues (*i*-1 and *i*-2) in order to form turns as previously described. In the case of the last residue, it does not have following residues (*i*+1 and *i*+2) that would form turns. The average was calculated by dividing the total number of residues in the (H) or (G) conformations by the total number of possible helical residues (7 for HBS peptide and 18 for K19 peptide). The simulation temperature used for this analysis was 300K as the experimental information for both peptides was obtained at this temperature.

2.3 Comparison of backbone NMR order parameters for Lysozyme and Ubiquitin

One of the standards for which force fields are evaluated is internal backbone dynamics in the form of order parameters. The comparison between NMR order parameters obtained from MD simulations and spin relaxation data has become customary for lysozyme and ubiquitin [4, 69, 72, 90, 154, 263-268] and necessary for this purpose. As previously described (section 1.5.10) the implementation of the Lipari-Szabo method as described by Prompers et al. [230], was applied to MD simulations of lysozyme and ubiquitin.

Lysozyme is a well-known and characterized protein with vast amounts of experimental information available such as NMR nuclear relaxation data [4, 69, 72, 198, 265] as well as

ubiquitin [269]. The validation of force field quality against these experimental observables has become customary and necessary; most of the time when new modifications to the existent force fields are implemented, simulations comparing experiments in simulations are reported as in the case of GROMOS [263, 264], CHARMM [266, 270] and AMBER [89, 90, 154, 158, 268]. Following up on this validation we also calculated NMR order parameters for the postulated modifications and compared to the experimental data available following the conditions described by Koller et al. [198]. The simulated structures of lysozyme and ubiquitin are shown in figures 32 and 33

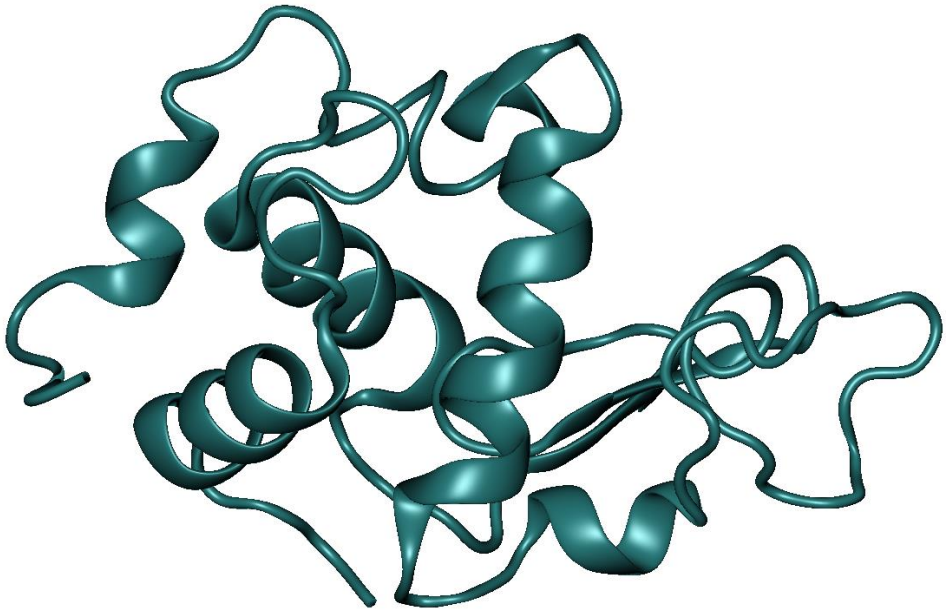


Figure 31 simulated lysozyme structure. PDB access code 1IEE

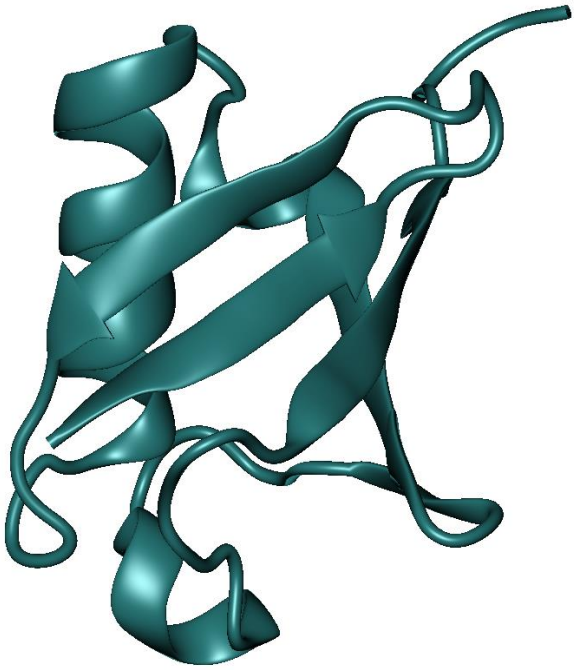


Figure 32 simulated structure of ubiquitin. PDB access code 1UBQ

2.4 Evaluating β secondary structure propensities for modifications

One of the goals of this thesis is to increase the helical propensities for simulations of transient helical peptides, however since the dihedral parameters for classical AMBER force fields are decoupled; small changes to the energy function as the ones proposed here will influence secondary structure propensities elsewhere. The proposed modifications to the backbone dihedral energy functions are minimal; nevertheless it is necessary to determine how these changes will affect β structures for simulations. Ideally the force field parameters should be robust enough in order to simulate diverse secondary structures without preferring specific ones. In order to corroborate this, β-hairpins ranging from 10 to 16 amino-acid residues were simulated for comparison to the secondary structure propensities to ff99SB to investigate how the newly developed parameters were performing with β structures for both hairpins. The metric for this test was agreement with backbone root mean square deviation to the native structure. For this particular testing we used four hairpins chignolin [185], CLN025 [186], GB1m3 and HP5F [191].

2.4.1 Testing CLN025 and HP5F hairpins in implicit solvent simulations

2.4.1.1 CLN025

CLN025 [186] is a hairpin that originated from the chignolin peptide [185]. This peptide was designed to be more stable than its predecessor chignolin by mutating the glycine residues at the termini to tyrosine residues. It is stabilized by electrostatic contacts between the charged termini, cross-strand hydrogen bonds. Furthermore, for this hairpin, almost 50% of the amino-acid sequence is aromatic; which also suggests a strong influence of aromatic interactions that stabilize its fold. Other interactions such as cation-π interactions have been described for this peptide [271].

One of the features that makes CLN025 useful for protein folding simulations is its fast folding properties previously described by Davis et al. [272] in the nanosecond time scale. Previous experimental/MD simulation work published by Hatfield et al. [271, 273], Lindorff-Larsen et al. [73] reported folding mechanisms for this small peptide as well. Its size of only 10 residues also makes it attractive for experimental and MD simulations, hence it has become a popular test model [73, 271, 273-277] for protein folding. Details for the peptide are specified in figure 34.

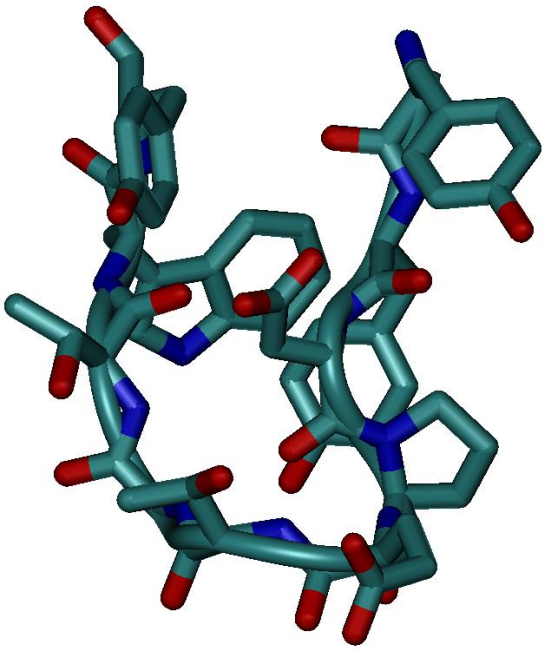


Figure 33 CLN structure, amino-acid sequence = YYDPETGTWY. Bio-magnetic Resonance Bank (BMRB) access entry 20009

Simulations for this hairpin were ran with Replica Exchange Molecular Dynamics (REMD) [249, 250] in implicit solvent in order to obtain converged results between native and extended conformations. This solvent model was used for these simulations because it simplifies the calculations by representing the solvent around the peptide as a continuum. Only the internal atomic interactions of the molecule are computed in the simulation which makes the Replica Exchange method feasible for this system. Subsequently one of the limitations of the Replica Exchange method is that the number of replicas required for the simulation increases as the size of the system increases; hence, including the solvent explicitly will increase its size, requiring large number of replicas for the simulation. Furthermore, replica exchange molecular dynamics of large systems have been reported as problematic [278, 279]

One of the challenges of using implicit solvent models such as Generalized Born is that they have limitations that have been previously reported [9, 193, 280-283] such as secondary structure preferences for different models and over-stabilization of salt bridges. Nevertheless, the method is still used for simulations of large biomolecules [284-286] and its results reproduce the trends observed in explicit solvent simulations [261, 286, 287].

2.4.1.2 HP5F

Further testing with a different hairpin was necessary in order to determine if the results obtained with CLN025 were transferable to other hairpins. We used HP5F because it is a longer hairpin unrelated to CLN025 and it has been simulated previously for the parameterization of Generalized Born solvent models using ff99SB as the chosen force field [193]. The previously obtained information for ff99SB served as a measurement of how the modifications were performing in comparison to ff99SB and to determine if the results obtained from simulation had reasonable agreement.

HP5F (figure 35) is a 16 residue hairpin that was obtained from mutational experiments designed to stabilize the second hairpin of the protein GB1 domain. The mutations were specifically designed to enhance the turn region and to favor Coulombic interactions at the termini. The optimization of the turn region was done based on database search of amino-acids that commonly occur in turn regions of hairpins like this [191]. The termini residues were mutated to Lysine in order to favor hydrogen bond interactions between the two strands in the peptide [191]. Along with HP5F, Fesinmeyer et al. [191] reported other hairpin systems where stability was increased as much as 7 kJ/mol. The reported HP5F sequence was one of the hairpins with higher stability where it was reported to be 82 ± 4% as per chemical shift deviation experiments [191].

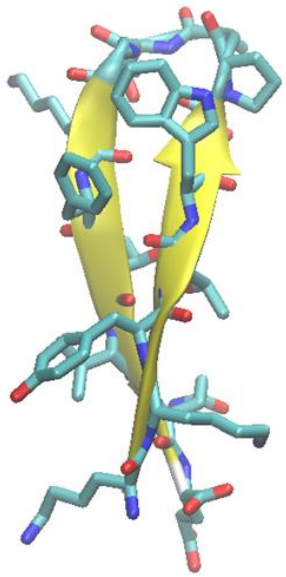


Figure 34 Structure for HP5F hairpin obtained from mutations to the GB1p[288] hairpin. The amino-acid sequence for this hairpin is KKYTWNPATGKFTVQE

2.4.2 Simulations of Hairpins in Explicit TIP3P waters

2.4.2.1 Chignolin

In order to further validate the results obtained from the implicit solvent simulations for hairpins, testing continued in explicit solvent. The goal of this testing was to determine if the modifications would cause the simulated conformations to frail in a given period of time. In other words, the conformations would be destabilized throughout the simulation which would indicate the fallacies of the modifications proposed. Chignolin was chosen for this type of simulation because is small like CLN025 and has transient β-hairpin structure contrary to CLN025 which has higher thermodynamic stability. The chosen peptide ought to be able to sample both folded and folded states for the given amount of time that the simulation took place (100 ns). The intention was not to run folding simulations of this peptide, instead it was necessary to observe if the peptide could fold and unfold without preferring a specific state.

Chignolin is a 10 residue peptide (figure 36) with melting temperature at 315K[185] designed from the G-peptide scaffold reported by Honda et al. in 2000 [289]. G-peptide is constituted by 16 residues taken from the section of amino-acids 41-56 of the protein G B1 domain. Originally Blanco et al. reported this peptide to fold into a hairpin in aqueous solution in 1994 [192] without the need to introduce any mutations to enhance its stability. Honda et al.

[185] selected 8 residues from this peptide that would rationally form a stable β-hairpin using secondary structure propensity information for all amino-acids from databases of proteins of the Protein Databank; this information was obtained from more than 100 non-homologous proteins.

Similar to CLN025, this peptide has been used for experimental [187, 188, 290] and protein folding simulation studies [53, 55, 59, 190] making it popular as a model system. Recently Best et al. used this peptide for force field validation of ff99SB, ff03*, ff03w and CHARMM22/CMAP [59] against NMR observables.

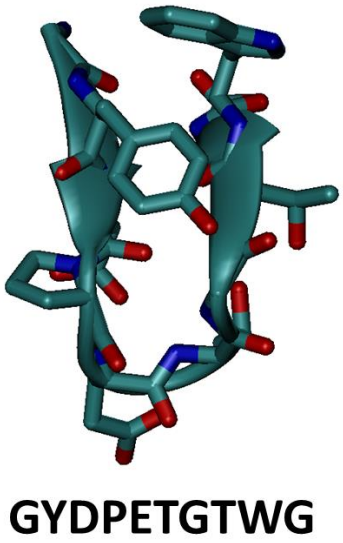


Figure 35 Structure of chignolin. PDB access code 1UAO

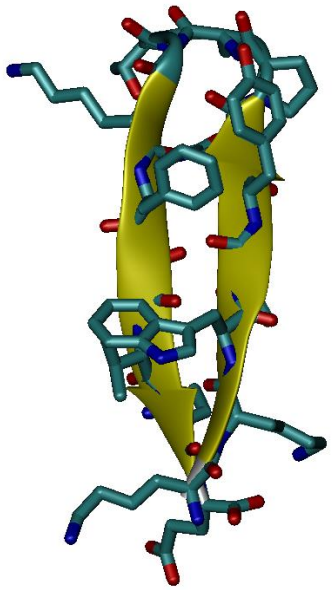
Native and extended conformations of this peptide were simulated in explicit solvent for approximately 100 ns in order to determine if it would unfold in this time as a measure of the effect of the modifications in the stability of the peptide. The results were compared against the results obtained from simulations of ff99SB.

2.4.2.2 GB1m3

Similarly to the case of implicit solvent simulations, it was necessary to test the hairpin stability for a peptide longer than chignolin. As previously described the goal of this testing is to test the stability of the structure for a given period of time in explicit solvent in order to validate

the modifications proposed.

This hairpin is homologous to HP5F peptide, both peptides were derived from the GB1p sequence by Fesinmeyer et al. [191]. The differences between these two hairpins are illustrated in figure 37. Both hairpins have comparable stability at room temperature (82 ± 4% for HP5F and 86 ± 3% for GB1m3). Structurally, the only difference between these two hairpins is at the 3rd and 5th positions, where tryptophan and tyrosine are interchanged between the sequences (see figure 37). Both of these hairpins have been used for the validation of revised implicit Generalized Born parameters and protein folding studies [56, 75, 91, 193, 194, 291], where GB1m3 has more simulation reports than HP5F with explicit solvent models. We chose GB1m3 since this hairpin has more previous explicit solvent simulation data available for assessment of the quality of simulations.



HP5F Sequence	KK**Y**T**W**NPATGK**F**T**V**QE
GB1m3 Sequence	KK**W**T**Y**NPATGK**F**T**V**QE

Figure 36 GB1m3 hairpin and comparison between GB1m3 and HP5F hairpin sequences

2.4.3 Definition of folded structures for hairpin simulations

2.4.3.1 CLN025 and HP5F REMD simulations

Native structures from simulations were defined as snapshots from the trajectory that had a backbone Root Mean Square Deviation (RMSD) of 2.0Å or less to the NMR structure (BMRB entry 20009). This metric was obtained by plotting the histogram distribution of the backbone RMSD values obtained from simulations with ff99SB force field.

As seen in figure 38, the RMSD distribution is the greatest for values close to 2.0Å. Then there are other distributions around 3-4 Å. In order to determine if the characteristics of the native hairpin were still present in these populations around 2.0Å, the trajectory was visually inspected to evaluate for the correct backbone fold and hydrogen bonding pattern of residues. We did not elaborate on other metrics such as side-chains RMSD; side-chain stacking that can define the native state more thoroughly. This is a limitation of the metrics used for this analysis because the description of native can be considerably broad and not specifically defined. Nevertheless, this metric can help to determine if the modifications are affecting the canonical backbone definitions for a hairpin which would indicate problems with the secondary structure propensities of the force fields being tested.

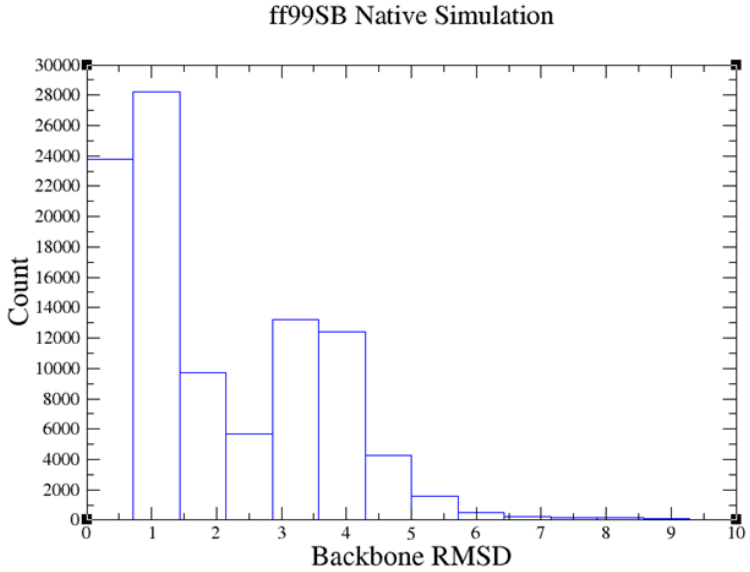


Figure 37 Histogram analyses for the definition of the native state cutoff. This information was obtained from replica exchange molecular dynamics post-processed temperature trajectory at 300K. The simulation was done with ff99SB for CLN025 hairpin. The backbone RMSD was calculated by measuring the distances for the N, C_α and C atoms of all amino-acids excluding the N and C terminus.

After this analysis was done, it was decided that this was an appropriate measure since the features of the hairpin were seen within this range according to the histogram analysis. The trajectories used for this analysis were the post-processed temperature trajectories at 300K. We compared the results obtained from simulations for each of the force fields being tested.

In the case of the HP5F hairpin the definition of native was borrowed from the work described by Nguyen et al. where they used a similar hairpin definition for the validation of their revised implicit solvent parameters [193]. They used a similar approach as described here to define their cutoff of 2.5Å.

2.4.3.2 Melting curves from backbone RMSD values

One of the reasons for using Replica Exchange Molecular Dynamics is that you have several simulations running at different temperatures; in the case of CLN025 and HP5F the temperatures used for the simulations were 280K, 300K, 315K, 325K, 340K, 360K, 380K and 400K. Since the method allows for exchanges between the different replicas, the replica trajectories were post-processed to include the trajectory information of each temperature into separate "temperature trajectories"; in other words separating each temperature trajectory information from the independent replica trajectories. After this, backbone RMSD values were measured for each of the temperature trajectories ranging from 280-400K. Then the structures that had backbone RMSDs $\leq$ 2.0Å were counted and divided by the total number of structures of each temperature trajectory and multiplied by 100 in order to determine the fraction folded. This was done to obtain populations of native structures as a function of temperature as in a melting curve. In order to have error bars for this information, this procedure was applied to native and extended simulations of both hairpins as in the case of the work of Nguyen et al. [193].

2.4.4 Chignolin and GB1m3 MD simulations

In the case of chignolin and GB1m3 simulations, only the backbone RMSD of the MD trajectories was measured as a way to determine how the simulation was sampling the different RMSD values for 100 ns. This was done to the native and extended structures of both hairpins in order to determine how the backbone of the hairpin was behaving in this given time and not intended to define folded states; instead this measurement was used to conclude the stability of the hairpin structure during this time.

2.4.5 Workflow for the validation of the modifications

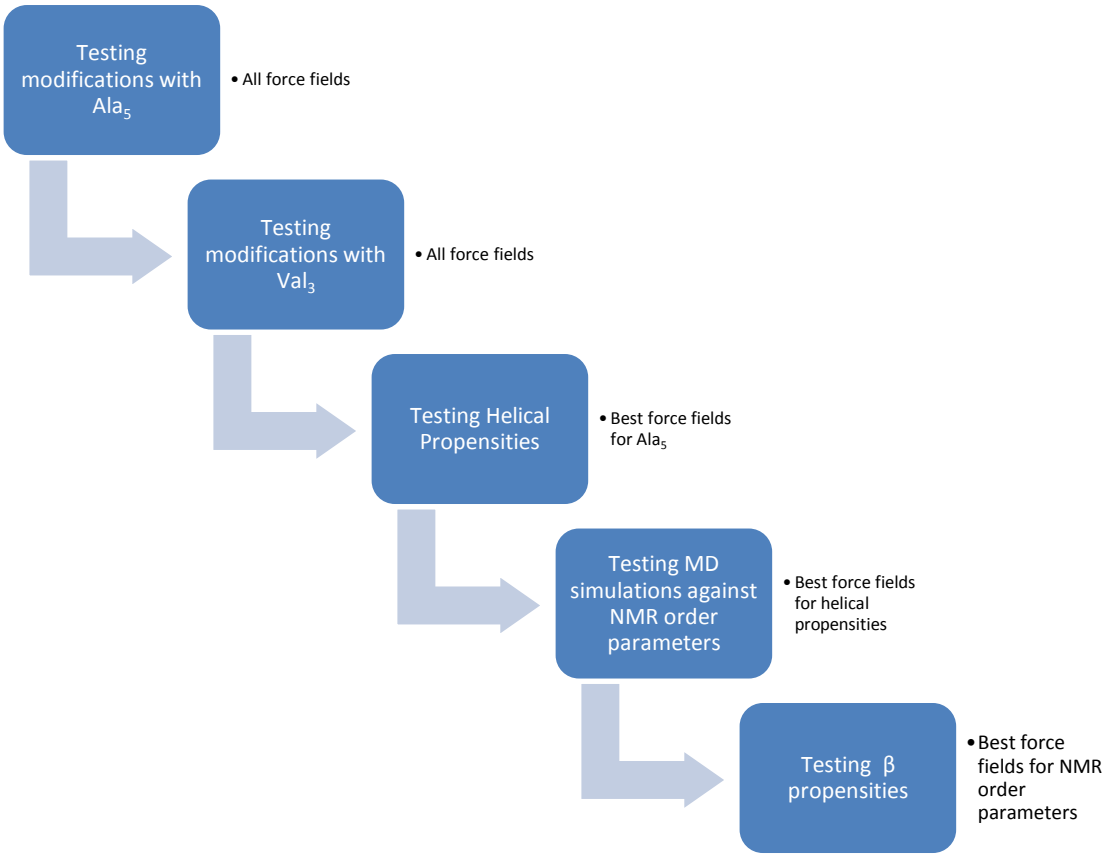


Figure 38 Workflow for the validation of force field parameters

2.5 Simulation Methodology

2.5.1 Molecular Dynamics and Monte Carlo Simulations

Molecular dynamic (MD) and Monte Carlo (MC) simulations have increased in popularity as computational power has increased. This is partially because other methods such as *ab initio* calculations require extensive quantum mechanical calculations of wave functions for each atom. As the size of molecules increases the amount of calculations increases as well, likewise requiring more computer power. Even with the increased speed and memory power in modern computers, this method still remains computationally expensive and time consuming. Furthermore, *ab initio* calculations limit the amount of possible conformations being studied.

On the other hand, the implementation of molecular mechanics methods to macromolecules like proteins and DNA has become routine. Proteins and DNA are of particular interest because of their role in cell biology and disease, making them preferred targets for drug development and novel therapeutics. Thus, MD and MC simulations have increasingly played a significant role in the interpretation of experiments, likewise serving as a tool to complement the results obtained. This can be appreciated in the growing application of MD/MC dynamics to structure refinement of NMR and X-ray crystallographic data as well as structure prediction.

At the moment MD/MC simulations in the nanosecond time scale are quite feasible allowing for direct assessment of the quality of simulations against hetero-nuclear NMR spin relaxation. In certain cases, other NMR observables such as chemical shifts, scalar J couplings, residual dipolar couplings (RDC) and order parameters can be detected at this time range as well; nevertheless wider ranges to the millisecond time scale can also be identified. Likewise, certain protein motions have longer time scales that require longer simulation time in order to be sampled.

As faster algorithms for computer simulations are being developed, μs/ms simulations have become more achievable and customary. Nevertheless, enhanced sampling MC methods are still necessary to study events such as protein folding and chemical processes at higher time scales. One of the methods described below is Replica Exchange Molecular Dynamics.

2.5.2 Replica Exchange Molecular Dynamics REMD

This method also known as multiple Markov chain method and parallel tempering[249] is an enhance sampling method in which multiple replicas of the same simulation run simultaneously and independently for a certain number of Monte Carlo or Molecular Dynamic steps [250]. The method allows for effective sampling of rough energy landscapes so trapped states in the simulation can escape local minima until the reach the global minimum. As shown in figure 40 multiple copies of simulations are setup to different temperature conditions. The simulations are allowed to proceed and every few simulation steps the copies of the simulation or "replicas" attempt to exchange following a stochastic criterion.

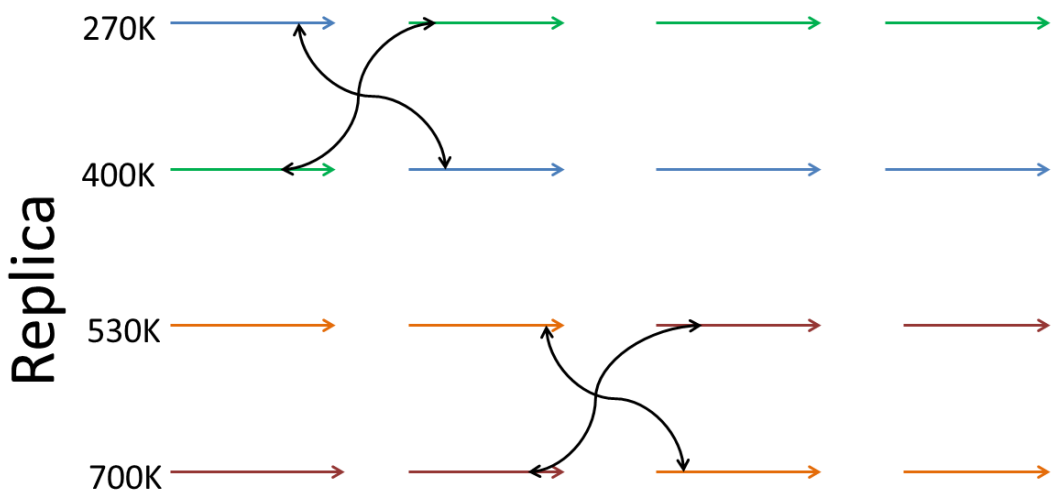


Figure 39 Replica exchange molecular dynamics scheme [292] . Multiple replicas proceed and attempt to exchange given the REMD metropolis criterion

In principle the simulations follow a random walk in energy and coordinate space as described by figure 41. This random walk allows for different states of the simulations to exchange as the simulation proceeds promoting wider sampling of the energy landscape in comparison to the sampling that could be obtained from canonical MD simulations. As seen in figure 41 the low and high energy conformations can interchange, speeding the path for the structure to reach its native state.

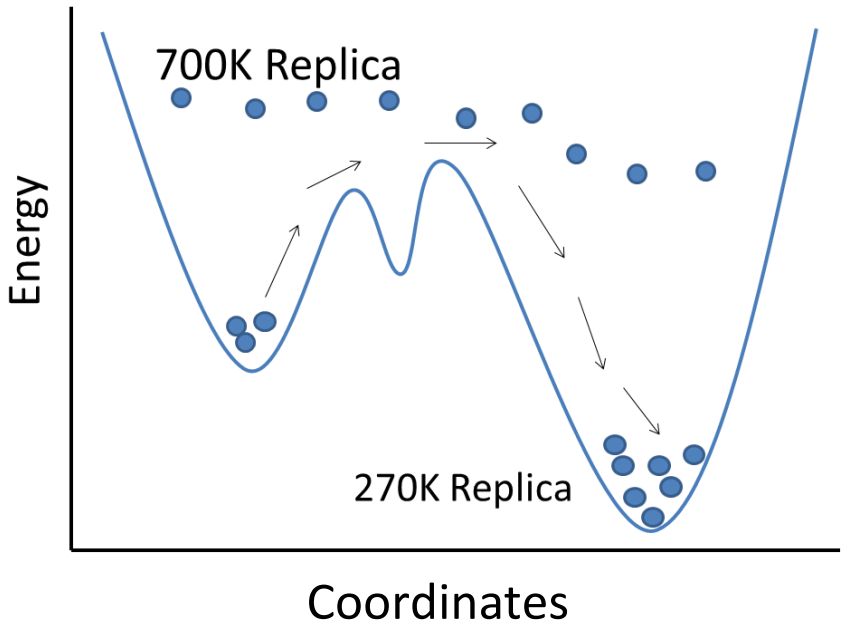


Figure 40 Replica Exchange Molecular Dynamics (REMD) diagram

The method requires that the simulation copies be closely distributed in temperature space. In other words, the energy distributions must overlap in order for the exchange probabilities not to be zero. If the temperature distribution is not closely and evenly spaced in between replicas the simulations will not attempt to exchange and the desired results will not be obtained. The results obtained would be comparable to multiple parallel canonical MD simulations. Due to this requirement the method has limitations; it is limited to the size of the system being simulated. The amount of required replicas for simulations is proportional to the size of the system. If the size of the system is considerably large it will require a large number of replicas as well. Since the temperature range must be closely and evenly distributed in order for the exchange probabilities not to be zero the range of temperatures must be large in order for the probabilities to overlap. This makes the energy space difficult to sample in order to obtain converged results because the exchange between distant replicas will not be as often as in the case of more compact energy landscapes. Besides this, computer power also limits the possibility of large number of replicas as large numbers of processors are required for this type of calculations.

The criterion for the exchange process to converge towards an equilibrium distribution

requires imposing the detailed balance condition on the transition probability *w (X → X′)* as described by equation 32 where the criterion for exchange is explained in detail:

$$W_{REM}\,(X)w(X \to X') = W_{REM}(X')w(X' \to X)$$

$$\frac{w(X \to X')}{w(X' \to X)} = \exp\{-\beta_m[K(p^{[j]'}) + E(q^{[j]})] - \beta_n[K(p^{[i]'}) + E(q^{[i]})] + \beta_m[K(p^{[i]} + E(q^{[i]})] + \beta_n[K(p^{[j]}) + E(q^{[j]}]\},$$

$$= \exp\left\{-\beta_m \frac{T_m}{T_n} K(p^{[j]}) - \beta_n \frac{T_n}{T_m} K(p^{[i]}) + \beta_m K(p^{[i]}) + \beta_n K(p^{[j]}) - \beta_m[E(q^{[j]}) - E(q^{[i]})] - \beta_n[E(q^{[i]} - E(q^{[j]})]\right\},$$

$$= \exp(-\Delta),$$

Where

$$\Delta \equiv [\beta_n - \beta_m](E(q^{[i]}) - E(q^{[j]}))$$

And *I*, *j*, *m* and *n* are related by the permutation functions before the exchange:

$$i = f(m),\ \ j = f(n)$$

And the Metropolis criterion:

$$w(X \to X') \equiv w(x_m{}^{[i]} | x_n{}^{[j]}) = \begin{cases} 1, & for\ \Delta \leq 0, \\ \exp(-\Delta), & for\ \Delta > 0, \end{cases}$$

Equation 33 Derivation of the energy criterion for exchanges in REMD simulations[250].

The method has been widely reported to simulate molecular processes such as domain dynamics[250], amyloid formation[293], protein folding [39, 48, 64, 76, 294] and protein ligand binding[295] . We used this method in order to produce converged results from simulations and to generate robust statistics. This method was used with implicit Generalized Born (GB) and explicit solvent models.

2.6 Solvation

Protein-solvent interactions play an important role in the events conducting to protein folding, stability, catalytic activity, association and ligand binding. They have also become an attractive target for biotechnology development of protein applications such as development of strategies for the purification and storage of proteins, as well as food processing and preservation. Therefore, solvation plays an essential role in molecular dynamic simulations because it is necessary to describe the solute - solvent interactions appropriately in order to obtain accurate results.

The solvent models used for simulations are implicit and explicit solvent models, in particular the revised Generalized Born Implicit solvent model[193] and TIP3P solvent model[135].

2.6.1 Generalized Born Implicit Solvent Models

Implicit solvation is a method of representing the solvent around the molecule of interest as a continuous medium instead of describing each explicit molecule of the solvent. The method is most often used in molecular dynamic simulations and in other applications of molecular mechanics. The method is used to estimate free energy of interactions in structural and chemical processes, such as protein folding [37, 282, 296] or conformational transitions of proteins[9, 281], DNA, RNA[297-299] and polysaccharides[300]. Other calculations include the free energy of association of biological macromolecules with ligands[169], or transport of therapeutics across cellular membranes[301].

In order to estimate the free energy of solvation the following equation is used:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{non-polar}$$

Equation 34 Solvation Free Energy

$\Delta G_{non\text{-}polar}$ is the free energy necessary to solvate a molecule from which all charges have been removed and ΔG_{el} is the free energy penalty for removing the charges from the molecule in a vacuum medium and then adding them back in the presence of solvent[283].

The implicit solvation model is applied for liquid substances, where the potential mean of force can be useful to estimate the averaged behavior of many highly dynamic solvent molecules[302]. There are two basic types of implicit solvent methods: models based on accessible surface areas and continuum electrostatic models, although various modifications and combinations of the different methods are possible[303]. The accessible surface area (ASA) method is based on experimental linear relations between Gibbs free energy of transfer and the surface area of the molecule being solvated[280].

For molecular dynamics the method is coupled directly with free energy of solvation estimation, unlike molecular mechanics or electrostatic methods that include only the enthalpy component of free energy. The continuum representation of solvent is also useful to make simulations faster and diminishes errors in statistical averaging that arise from incomplete sampling of solvent conformations. The energy landscapes obtained with implicit and explicit solvents are different; however the trends and tendencies in simulations can be determined between them. Even though the implicit solvent model is useful for simulations of biomolecules, this approximation have limitations and problems related to parameterization and treatment of ionization effects[303].

Generalized Born implicit solvent (GBIS) is an approximate method for calculating molecular electrostatics in solvent as described by the Poisson Boltzmann equation which models water as a dielectric continuum[280]. GBIS enables the simulation of atomic structures without including explicit solvent water. The elimination of explicit solvent makes simulations faster; however as the number of atoms in the molecule increases, the electrostatics calculation becomes more complex because the pairwise interactions increase as well. This makes the scaling of the simulations sometimes slower or comparable to the explicit solvent simulations.

2.6.1.1 The Generalized Born (GB) equation

The Generalized Born free energy of solvation is [304]:

$$\Delta G_{pol} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}}$$

Equation 35 Generalized Born Free Energy of Solvation

Where q_i and q_j are partial charges, ε_w is the solvent dielectric constant, and f_{GB} is the function that interpolates an "effective Born Radius" R_i . The effective Born radius is originally obtained by numerical integration, however this calculation has been simplified by obtaining the effective Born radius from pairwise approximations. The approximation is estimated from a sum over atom pairs [297, 298, 305-308].

In order to obtain the electrostatic potential model the Poisson-Boltzmann equation has to be taken into account:

$$\nabla[\varepsilon(\boldsymbol{r})\nabla\varphi(\mathbf{r})] = -4\pi\rho(\mathbf{r})$$

Equation 36 Poisson Boltzmann equation for a continuum

Where ρ is the charge distribution, and the dielectric constant ε. The dielectric constant can be calculated at the interior (ε_{in}) and exterior (ε_{ext}). In the gas phase the dielectric is equal to 1 while in solvent conditions the dielectric is ε_{solv}. The potentials for these conditions are φ_{solv} and φ_{vac} . The difference for the potentials is:

$$\varphi_{reac} = \varphi_{solv} - \varphi_{vac}$$

Equation 37 Difference between the potentials in the gas and solvent phase

The electrostatic component of the solvation free energy is:

$$\Delta G_{pol} = \frac{1}{2}\int \varphi_{reac}\,(\mathbf{r})\rho(\mathbf{r})dV$$

Equation 38 Electrostatic component of the solvation free energy taken from the reaction potentials

If the distribution is approximated by partial atomic charges q_i the electrostatic component then becomes:

$$\Delta G_{pol} = \frac{1}{2}\sum_i q_i\,\varphi_{reac}(\boldsymbol{r}_i)$$

Equation 39 Electrostatic energy approximated by partial atomic charges

In the case of an ion of radius *a* and charge *q* electrostatic equation becomes:

$$\Delta G_{Born} = -\frac{q^2}{2a}\left(1 - \frac{1}{\varepsilon_{solv}}\right)$$

Equation 40 Born energy equation

For a molecule with charges $q_i \ldots q_N$ surrounded by spheres of radius $a_i \ldots a_N$ then the free energy can be approximated by a series of pairwise of Born terms and Coulombic terms:

$$\Delta G_{pol} = \sum_{i}^{N} \frac{q_i^2}{2a_i}\left(\frac{1}{\varepsilon_w} - 1\right) + \frac{1}{2}\sum_{i}^{N}\sum_{j\neq 1}^{N} \frac{q_i q_j}{f_{GB}(r_{ij})}\left(\frac{1}{\varepsilon_w} - 1\right)$$

Equation 41 Polar free energy for pair-wise terms of Coulombic and Born terms

Where the Generalized Born Function is:

$$f_{GB}(r_{ij}) = \left[r_{ij}^2 + R_i R_j \, exp\left(\frac{r_{ij}^2}{4R_i R_j}\right)\right]^{1/2}$$

Equation 42 Generalized Born Function that introduces the concept of effective radii and intrinsic radii = r_{ij}

$$R_i^{-1} = a_i^{-1} - \frac{1}{4\pi}\int_{in, r>a_i} \frac{1}{r^4}\, dV$$

Equation 43 Effective Born radii Definition[9]

2.6.1.2 Explicit Solvent Models : TIP3P

Explicit solvent models are often used to represent water in simulations. There are several

models available depending on the level of representation that the user wants to achieve. Some models include "dummy" atoms to represent the lone pairs in the water molecule.

Some water models treat the water molecule as rigid and rely only on non-bonded interactions. These non-bonded interactions are similar to the non-bonded interactions in a force field such as the electrostatic interaction, modeled using Coulomb's law and the dispersion and repulsion forces using the Lennard-Jones potential [309]. The potential for water models such as TIP3P and TIP4P is described as:

$$E_{ab} = \sum_{i}^{on\ a} \sum_{j}^{on\ b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}}$$

Equation 44 Potential energy term for water models as per Jorgensen et al.[135]

Where k_C, the electrostatic constant, has a value of 332.1 Å·kcal/mol; q_i are the partial charges relative to the charge of the electron; r_{ij} is the distance between two atoms or charged sites; and *A* and *B* are the Lennard-Jones parameters. The Lennard-Jones term is usually applied to the oxygen atom from the water molecule. TIP3P belongs to a class of water models called three-site water models. These models have three interaction sites, corresponding to the three atoms of the water molecule. Each atom gets assigned a point charge, and the oxygen atom also gets the Lennard-Jones parameters. The TIP3P water model is very popular for molecular dynamics simulations because of its simplicity and computational efficiency.

Table 9 Summary of Parameters for the TIP3P Model

Parameters	TIP3P Water Model
r (OH), Å	0.9572
H-O-H, deg	104.52
A × 10^{-3}, kcal $Å^{12}$/mol	582.0
B, kcal $Å^6$/mol	595.0
q (O)	-0.834
q (H)	+0.417

Table 9 Summary of Parameters for the TIP3P water model [206]

All the calculations presented here in explicit solvent were done with the TIP3P water model.

2.7 Simulation details Ala_5

All runs for Ala_5 were simulated with Amber 11 and 12 packages. The peptide had free and protonated C-terminus without counter ions as described by Graf et al. [172]. The parameter source for the protonated C-terminus was borrowed from the work of Wickstrom et al. [151]. The number of water molecules used in the solvation of the truncated octahedron box was 891 for native and extended simulations using TIP3P as the water model [135]. The time step was setup to 2 fs. The topology and coordinate files for all systems were built with the tleap module in AMBER.

The native conformations were defined by building the Ala_5 structure and imposing helical dihedral angles in all amino-acids (φ = -60º ψ = -40º). The extended simulations were defined by building the same structure without imposing any dihedral angles in the amino-acids. Each one of these conformations was run for 50 ns. As previously described, there were thirty combinations of force fields (ff99SB + 29 variants) in which each possible modification to the φ' dihedral term was matched with a set of ψ/ψ' modifications. Therefore, the total amount of simulations added to 60, which were ran with Replica Exchange Molecular Dynamics (REMD)[249, 250] under the sander module in AMBER.

Each of these conformations was equilibrated at 300K through gradual heating during 50 ps while imposing harmonic restraints on the system. After equilibration, the system was minimized while progressively reducing positional restraints at constant pressure (1 atm) and temperature (300K) following on the published protocol of Wickstrom et al. [151]. The coordinate files from this last step were taken to perform REMD. The temperatures for the runs were 293K, 299K, 306.9K, 314.1K, 321.5K, 329K, 336.8K, 344.7K, 352.7K, 361K, 369.5K, 378.2K, 387K, 396.1K, 405.4K and 414.9K. Exchanges were attempted every ps with a target acceptance ratio of 20%. Production runs were 50 ns long, again following on the work published by Wickstrom et al. [151]. The simulations were run in the NVT ensemble with a

weak Berendsen thermostat [310] and particle mesh Ewald [311] for long range electrostatic interactions with a cutoff of 6Å. We also performed long MD simulations for 160 ns with a cutoff of 8.0Å under the same conditions as described above in order to determine the effect of the change in the cutoffs between simulations.

2.8 Simulation details alanine dipeptide

All runs for alanine dipeptide (N-acetyl alanyl-N-methyl-amide) Ace-Ala-NMe were simulated with AMBER 12 package. The initial topology and coordinate files were created with the tleap module of AMBER. The free energy surface was obtained through umbrella sampling simulations by constructing and minimizing independent conformations where the φ and ψ were rotated every 5º to cover the dihedral region of -180º to 180º for both angles, following on the procedure described by Bergonzo et al. [312]. These coordinates were solvated with 378 TIP3P[135] waters, generating 1296 umbrella sampling windows. During the minimization procedure the φ and ψ angles were restrained with a force constant of 500 kcal/mol.rad^2 through 1000 steps. Each minimized structure was used to generate an independent umbrella sampling window, while the force constant used for these simulations was 75 kcal/mol.rad^2 on φ and ψ dihedral angles. Each window was run for 10 picoseconds in time steps of 2 femtoseconds at 300K while saving the dihedral angle information every 2 femtoseconds. The data was post-processed using the Weighted Histogram Analysis Method (WHAM) by Grossfield et al. [313].

2.9 Simulation details Val_3

All runs were done with AMBER 11 package [125]. Val_3 was simulated with free N and protonated C terminus as described by Graf et al.[172]. Like in the case of Ala_5, native conformations were defined by building the Ala_5 structure and imposing helical dihedral angles in all amino-acids (φ = -60º ψ = -40º). The extended simulations were defined by building the same structure without imposing any dihedral angles in the amino-acids. The number of water molecules used in the solvation of the truncated octahedron box was 916 for both simulations. The time step was setup to 2 fs. The topology and coordinate files for all systems were built with

the tleap module in AMBER. For this case thirty force fields were tested (ff99SB + variants) in which each possible modification to the φ' dihedral term was matched with a set of ψ/ψ' modifications. Therefore, the total amount of simulations added to 60, which were ran with Replica Exchange Molecular Dynamics (REMD)[249, 250] under the sander module in AMBER.

Each of these conformations was equilibrated at 300K through gradual heating during 50 ps while imposing harmonic restraints on the system. After equilibration, the system was minimized while progressively reducing positional restraints at constant pressure (1 atm) and temperature (300K) following on the published protocol of Wickstrom et al. [151]. The coordinate files from this last step were taken to perform REMD. The temperatures for the runs were 293K, 299K, 306.9K, 314.1K, 321.5K, 329K, 336.8K, 344.7K, 352.7K, 361K, 369.5K, 378.2K, 387K, 396.1K, 405.4K and 414.9K. Exchanges were attempted every ps with a target acceptance ratio of 20%. Production runs were 50 ns long, again following on the work published by Wickstrom et al. [151]. The simulations were run in the NVT ensemble with a weak Berendsen thermostat [310] and particle mesh Ewald [311] for long range electrostatic interactions with a cutoff of 6Å.

We also performed REMD simulations with revised GB Neck implicit solvent model as reported by Nguyen et al. [193], using mbondi 3 radii as specified in tleap in order to determine how the calculation of scalar coupling values would agree between explicit and implicit solvent models. This was done as a test that would determine how the implicit solvent model compares when calculating experimental observables. For these simulations, the definition of native and extended conformations previously described for the explicit solvent simulations was re-applied in this case. Langevin dynamics were used for minimization and production runs [314]. Initially, hydrogen atoms were restrained for a thousand cycles while the rest of the structure was allowed to minimize, followed by 500 ps of gradual heating with decreasing restraints on hydrogen atoms. Then, the backbone atoms were restrained for a thousand cycles followed by gradual heating to 300K. The next three rounds of equilibration required gradually reducing restraints from 10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then to 0.1 kcal/mol.Å^2 to complete unrestrained dynamics. After this last step of equilibration, the coordinates were used for REMD production runs of 50 ns long as this was the time were the deviations between the results obtained from the

native and extended simulations were similar. The temperatures used for the replicas were 280K, 300K, 315K, 325K, 340K, 360K, 380K and 400K with a target exchange ratio of 20%.

2.10 Simulation details Hydrogen Bond Surrogate (HBS) peptide

All simulations described here were run with AMBER 11 package [125]. The topology and coordinate files for all systems were built with the tleap module in AMBER. We built native conformations by imposing helical conformations on all the residues while the semi-extended ones had only imposed helical conformations on the first five residues to resemble a pre-organized helical turn that propagates helix to the rest of the peptide as described by experiment [168]. This pre-organized α turn is achieved by an unnatural covalent bond between an N-terminal cap with a C=C double bond and a carbon that is bound to the amide nitrogen of Alanine5, the four atoms here named 1C=C´-5C-5N. The most accurate simulation of this system would require parameter fitting, including partial charges. Since the first five residues are pre-organized by the covalent modification just described, however, enforcement of covalent bond lengths and angles is expected to be a sufficient approximation for examination of helical propensity.

To imitate this feature, we harmonically restrained bonds and angles to analogous bonded parameters of AMBER force fields. The length of the bond between the cap and the $alanine_5$ (1C´-5C) was restrained to 1.522Å with a weight of 317 kcal/mol.$Å^2$. The 1C=1O-5H angle was restrained to 120º with a weight of 80 kcal/mol.rad^2 and the 1C´-5C-5N angle was restrained to 110.1º with a weight of 63 kcal/mol/rad^2. These systems were solvated with 1339 waters and initially minimized while restraining the hydrogen atoms and the water molecules for ten thousand cycles followed by gradual heating from 0 to 100K and then from 100K to 300K for 100 ps while imposing harmonic restraints in the water molecules and hydrogen atoms. Then, restrained molecular dynamics were performed gradually reducing the constraints in the water and hydrogen bonds for 350 ps. After this step, restraints were applied to the backbone atoms and slowly removed. The steps of equilibration were performed under constant pressure of 1 atm.

The coordinate files from this last step were taken to perform REMD simulations, uncertainties were calculated from independent native and semi-extended runs. These simulations were run for 50 ns until the uncertainties between the independent runs were similar.

The temperatures used for the replicas in the simulation were 272.2K, 277.5K, 283K, 288.6K, 294.2K, 300.0K, 305.9K, 311.9K, 318.0K, 324.3K, 330.6K, 337.1K, 343.8K, 350.5K, 357.4K and 364.4K. This testing was applied to the modifications of the φ' term known as mod1φ, mod2φ and mod5φ as well as the combined set of mod1φ with modifications to the ψ/ψ' term known as mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ because these force fields had the best results when testing the vicinal scalar couplings for Ala_5.

2.11 Simulation details for HBS peptide in implicit solvent

Simulations described here were run with the AMBER 12 package. The topology and coordinate files were built with the tleap module in AMBER. The time step was setup to 2 fs. Simulations were performed with parallel tempering also known as replica exchange molecular dynamics[249, 250] under the sander module in AMBER. Langevin dynamics and revised GB Neck implicit solvent model with mbondi 3 radii as specified in tleap were used for the minimization and production runs. For the minimization protocol, initially, hydrogen bonds were restrained for a thousand cycles while the rest of the structure was allowed to minimize, followed by 500 ps of gradual heating with decreasing restraints on hydrogen bonds. Then, the backbone atoms were restrained for a thousand cycles while gradually heating to 300K. The next three rounds of equilibration required gradually reducing restraints from 10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then to 0.1 kcal/mol.Å^2 to complete unrestrained dynamics. After this last step of equilibration, the coordinates were used for production runs of 50 ns long as this was the time were the deviations between the results obtained from the native and extended simulations were similar. The temperatures used for the REMD simulations were 280K, 300K, 315K, 325K, 340K, 360K, 380K and 400K with a target exchange ratio of 20%. We built native conformations by imposing helical conformations on all the residues while the semi-extended ones had only imposed helical conformations on the first five residues to resemble a pre-organized helical turn that propagates helix to the rest of the peptide as described by experiment [168]. To imitate this feature, we harmonically restrained bonds and angles to analogous bonded parameters of AMBER force fields. The length of the bond between the cap and the $alanine_5$ (1C´-5C) was restrained to 1.522Å with a weight of 317 kcal/mol.Å^2. The 1C=1O-5H angle was restrained to 120º with a weight of 80 kcal/mol.rad^2 and the 1C´-5C-5N angle was restrained to 110.1º with a weight of 63 kcal/mol/rad^2.

2.12 Simulation details K19 peptide

All simulations were run with AMBER 12 package [125]. The topology and coordinate files for all systems were built with the tleap module in AMBER. The time step was setup to 2 fs. Explicit water simulations were performed in a truncated octahedron box with TIP3P [135]waters. Canonical MD simulations were run using the PMEMD module in AMBER [315]. The simulations were run in the NVT ensemble with a weak Berendsen thermostat [310] and particle mess Ewald (PME)[311] for long range electrostatic interactions. Real space electrostatics had a cutoff of 8.0Å.

We ran native and semi-extended conformations for K19. The native conformations were generated by imposing helical conformations on all residues. The semi-extended conformation was generated from a snapshot of an ff99SB trajectory that had random coil conformation after 80 ns of simulation in explicit water. This snapshot was stripped from its waters and generated a PDB structure that was solvated to the same amount of waters that the native conformation had. In order to have both systems with the same number of atoms, the semi-extended conformation was generated first with the size of the octahedron box to be 8.0Å. Then the native conformation was built from the sequence while imposing helical conditions in all residues and the size of the box was 12.6Å. The systems were initially minimized while restraining the hydrogen atoms and the water molecules for ten thousand cycles followed by heating for 100 ps while keeping the restraints in the water molecules and hydrogen atoms. Then, restrained molecular dynamics were performed gradually reducing the constraints from 10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then the restraints were lowered from 1 kcal/mol.Å^2 to 0.1 kcal/mol.Å^2 until they were eventually turned off completely through 350 ps cycles. The steps of equilibration were performed under constant pressure (1 atm). Both sets of simulations were run for approximately 200 ns at 300K in order to compare to the results from ff99SB simulation. In other words, the behavior of the force fields is compared to the behavior of ff99SB. The results presented here correspond to the mod1φ, mod2φ and mod5φ variations. We also tested other force fields mainly mod1φ-mod1ψ, mod1φ-mod2ψ and mod1φ-mod3ψ because they had promising results based on the testing conducted with Ala_5. However, these simulations were only prolonged to the 100 ns mark. The lengths of the simulations described here are not intended to define folded states, as previously reported by

Clarke et al.[316] helix nucleation is in the millisecond time scale which will require 5000 times more simulation time in order to have enough statistics relevant to this event. Previously work by Song et al. [184] reported REMD simulations for 56 ns per replica (12 replicas), totaling 672 ns of simulation in GB implicit solvent model of Tsui et al. [299]. Although the work presented here has considerably less statistical data than previously reported work, the simulations are intended to benchmark the performance of the modifications against the performance of ff99SB. In other words, the modifications should have at least similar behavior than what is observed for ff99SB.

In the case of Hydrogen Bond Surrogate Peptide, the simulations were done using REMD simulations for 16 replicas for 40 ns each which is 640 ns total simulation time. Furthermore, the pre-organized turn of the helix overcomes the nucleation rate limiting step required for helix formation. Therefore, these results can resemble helix folding process better than the simulations of K19.

2.13 Simulation details Lysozyme

The topology and coordinate files for all systems were built with the tleap module in AMBER. The native conformations were built from NMR structure deposited in the Protein Data Bank known as 1IEE [317]. The time step was setup to 2 fs. Explicit water simulations were performed in a truncated octahedron box with the TIP3P water model. Real space electrostatics had a cutoff of 8.0Å. MD simulations were run using the PMEMD module in AMBER. The simulations were run in the NVE ensemble with a weak coupling algorithm thermostat [318] and particle mesh Ewald for long range electrostatic interactions. Lysozyme simulations were run in order to estimate the force field quality against NMR order parameters S^2 corresponding to the N-H amide bond. The method used for the calculation was the iRED method developed by Prompers et al. [230]. The simulations were initially minimized while restraining the hydrogen atoms and the water molecules for ten thousand cycles followed by gradual heating for 100 ps while keeping the restraints in the water molecules and hydrogen atoms. Then, restrained molecular dynamics were performed gradually reducing the constraints from 10 kcal/mol.Å^2 to 1

kcal/mol.Å^2, then the restraints were lowered from 1 kcal/mol.Å^2 to 0.1 kcal/mol.Å^2 until they were eventually turned off completely through 350 ps cycles. The steps of equilibration were performed under constant pressure (1 atm). We generated ~ 25 ns at 300K of MD dynamics for mod1φ, mod2φ, mod5φ, mod1φ-mod1ψ, mod1φ-mod2ψ and mod1φ-mod3ψ. For each of these simulations we ran two simulations with different random seed values in order to determine uncertainties following on the recommendations provided by Koller et al. [198] who suggested that is more appropriate to calculate NMR order parameters from several short MD simulations than from one long one.

2.14 Simulation details Ubiquitin

The topology and coordinate files for all systems were built with the tleap module in AMBER. The conformations were built from NMR structure deposited in the Protein Data Bank known as 1UBQ [319]. The time step was setup to 2 fs. Explicit water simulations were performed in a truncated octahedron box with TIP3P water model. Real space electrostatics had a cutoff of 8.0Å. MD simulations were run using the PMEMD module. The simulations were run in the NVE ensemble with a weak coupling algorithm thermostat [318] and particle mesh Ewald for long range electrostatic interactions. Order parameters S^2 corresponding to the N-H amide bond were calculated. The method used for this calculation was the iRED method developed by Prompers et al. [230]. The simulations were initially minimized while restraining the hydrogen bonds and the water molecules for ten thousand cycles followed by heating for 100 ps while keeping the restraints in the water molecules and hydrogen atoms. Then, restrained molecular dynamics were performed while gradually reducing the constraints from 10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then the restraints were lowered from 1 kcal/mol.Å^2 to 0.1 kcal/mol.Å^2 until they were eventually turned off completely through 350 ps cycles. The steps of equilibration were performed under constant pressure (1 atm). We generated ~25 ns at 300K of MD dynamics for mod1φ, mod2φ and mod5φ. For each of these force fields we ran two simulations with different random seed values in order to determine uncertainties following on the recommendations provided by Koller et al. [198]. In certain cases such as mod2φ and mod5φ we ran three simulations to minimize the deviations between the runs.

2.15 Simulation details CLN025

Simulations described here were run with the AMBER 12 package. The topology and coordinate files were built with the tleap module in AMBER. The time step was setup to 2 fs. Simulations were performed with parallel tempering also known as replica exchange molecular dynamics[249, 250] under the sander module in AMBER. Langevin dynamics and revised GB Neck implicit solvent model with mbondi 3 radii as specified in tleap were used for the minimization and production runs [193, 314]. For the minimization protocol, initially, hydrogen bonds were restrained for a thousand cycles while the rest of the structure was allowed to minimize, followed by 500 ps of gradual heating with decreasing restraints on hydrogen bonds. Then, the backbone atoms were restrained for a thousand cycles while gradually heating to 300K. The next three rounds of equilibration required gradually reducing restraints from 10 kcal/mol.$Å^2$ to 1 kcal/mol.$Å^2$, then to 0.1 kcal/mol.$Å^2$ to complete unrestrained dynamics. After this last step of equilibration, the coordinates were used for production runs of 50 ns long as this was the time were the deviations between the results obtained from the native and extended simulations were similar. The temperatures used for the REMD simulations were 280K, 300K, 315K, 325K, 340K, 360K, 380K and 400K with a target exchange ratio of 20%. The native simulated sequence was taken from the Biological Magnetic Resonance Data Bank BMRB entry 2009 [186]. The extended conformation was created from the amino-acid sequence using the tleap module in AMBER.

2.16 Simulation details HP5F

Simulations described here were run with the AMBER 12 package. The topology and coordinate files were built with the tleap module in AMBER. The time step was setup to 2 fs. Simulations were performed with parallel tempering also known as replica exchange molecular dynamics[249, 250] under the sander module in AMBER. Langevin dynamics and revised GB Neck implicit solvent model with mbondi 3 radii as specified in tleap were used for the minimization and production runs [193, 314]. For the minimization protocol, initially, hydrogen bonds were restrained for a thousand cycles while the rest of the structure was allowed to minimize, followed by 500 ps of gradual heating with decreasing restraints on hydrogen bonds.

Then, the backbone atoms were restrained for a thousand cycles while gradually heating to 300K. The next three rounds of equilibration required gradually reducing restraints from 10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then to 0.1 kcal/mol.Å^2 to complete unrestrained dynamics. After this last step of equilibration, the coordinates were used for production runs of 50 ns long as this was the time were the deviations between the results obtained from the native and extended simulations were similar. The temperatures used for the REMD simulations were 280K, 300K, 315K, 325K, 340K, 360K, 380K and 400K with a target exchange ratio of 20%. The native conformation was taken from an MD simulation with ff99SB that started from a fully extended conformation. The structure with the lowest potential energy that resembled the topology of a folded hairpin was chosen as "native", although this assumption does not have experimental information such as NMR or X-ray crystal structures that support it. REMD simulations were carried out with this conformation and we chose the structure with the highest backbone RMSD from the 400K temperature replica (~ 8.0Å) as the extended conformation.

2.17 Simulation details chignolin

Simulations were run with AMBER 12. The native conformations were built from NMR structure deposited in the Protein Data Bank known as 1UAO [185]. The semi-extended conformations were taken from snapshots of a REMD trajectory simulated with ff99SB for 75 ns where the temperature of the replica was 414K where the highest RMSD values were 6.0Å. These conformations were taken from these snapshots because after solvation both the native and semi-extended conformations would have around the same number of atoms. It was necessary to have the same number of atoms in order for the simulations to have similar trajectory times. The topology and coordinate files were built with the tleap module in AMBER. The time step was 2 fs. Explicit water simulations were performed in a truncated octahedron box with the TIP3P [135] water model. The native and semi-extended conformations were solvated with the same number of waters and had the same number of atoms (7413). The systems were initially minimized while restraining the hydrogen atoms and the water molecules for ten thousand cycles followed by heating for 100 ps while keeping the restraints in the water molecules and hydrogen atoms. Then restrained molecular dynamics were performed while gradually reducing the constraints from10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then the restraints were lowered from 1

kcal/mol.Å^2 to 0.1 kcal/mol.Å^2 until they were eventually turned off completely through 350 ps cycles. The steps of equilibration were performed under constant pressure (1 atm). Both sets of simulations were run for approximately 100 ns at 300K in order to compare to the results from ff99SB simulation. In other words, the behavior of the force fields is compared to the behavior of ff99SB. The results presented here correspond to the mod1φ since it was the best performing force field overall.

2.18 Simulation details GB1m3

Simulations were run with AMBER 12. The native conformation was taken from an MD simulation with ff99SB that started from a fully extended conformation. The structure with the lowest potential energy that resembled the topology of a folded hairpin was chosen as "native", although this assumption does not have experimental information such as NMR or X-ray crystal structures that support it. The semi-extended conformations were taken from snapshots of a REMD trajectory simulated with ff99SB for 75 ns where the temperature of the replica was 414K where the highest RMSD values were 12.0Å. These conformations were taken from these snapshots because after solvation both the native and semi-extended conformations would have around the same number of atoms. It was necessary to have the same number of atoms in order for the simulations to have similar trajectory times. The topology and coordinate files were built with the tleap module in AMBER. The time step was 2 fs. Explicit water simulations were performed in a truncated octahedron box with the TIP3P [135] water model. The native and semi-extended conformations were solvated with the same number of waters and had the same number of atoms (7413). The systems were initially minimized while restraining the hydrogen atoms and the water molecules for ten thousand cycles followed by heating for 100 ps while keeping the restraints in the water molecules and hydrogen atoms. Then restrained molecular dynamics were performed while gradually reducing the constraints from10 kcal/mol.Å^2 to 1 kcal/mol.Å^2, then the restraints were lowered from 1 kcal/mol.Å^2 to 0.1 kcal/mol.Å^2 until they were eventually turned off completely through 350 ps cycles. The steps of equilibration were

performed under constant pressure (1 atm). Both sets of simulations were run for approximately 120 ns at 300K in order to compare to the results from ff99SB simulation. In other words, the behavior of the force fields is compared to the behavior of ff99SB. The results presented here correspond to the mod1φ since it was the best performing force field overall.

3. Results

3.1 Ala_5

3.1.1 Dihedral distribution of the second residue in Ala_5

We began by plotting the dihedral populations for the second residue of Ala_5, in order to determine how the modifications compared against the results obtained from ff99SB as seen in figure 42. The error bars from the graphs were obtained from the native and extended simulation runs. For clarity only ff99SB, mod1φ, mod2φ and mod5φ were plotted.

Following on the rationale that motivated the modifications, the sampling for the φ = -60º region increased, while the sampling of the -150º $\leq \varphi \leq$ -120º region decreased. In the case of mod5φ the populations around this region increased in comparison to ff99SB which was expected based on dihedral energy function for this modification (see figure 19). In order to further investigate the full dihedral map we plotted the Ramachandran plot with the dihedral populations of φ and ψ (see figure 42-43).

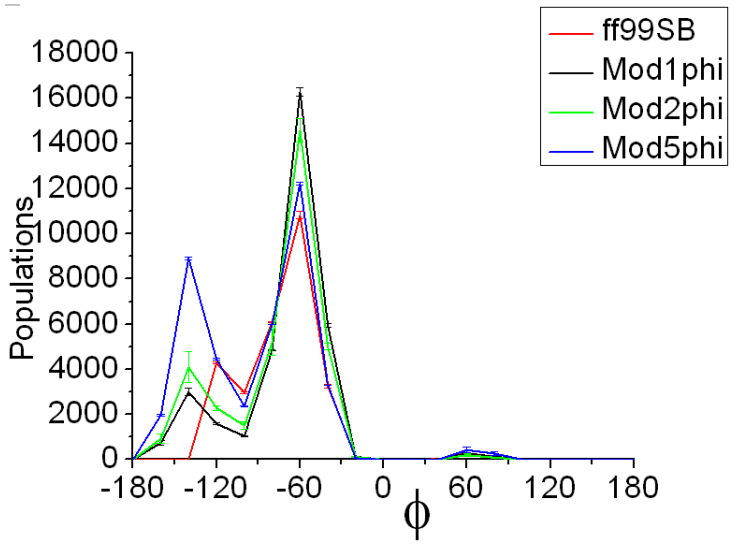


Figure 41 Dihedral populations for φ angle from the modifications to the φ' backbone dihedral term. Error bars taken from native and extended simulations.

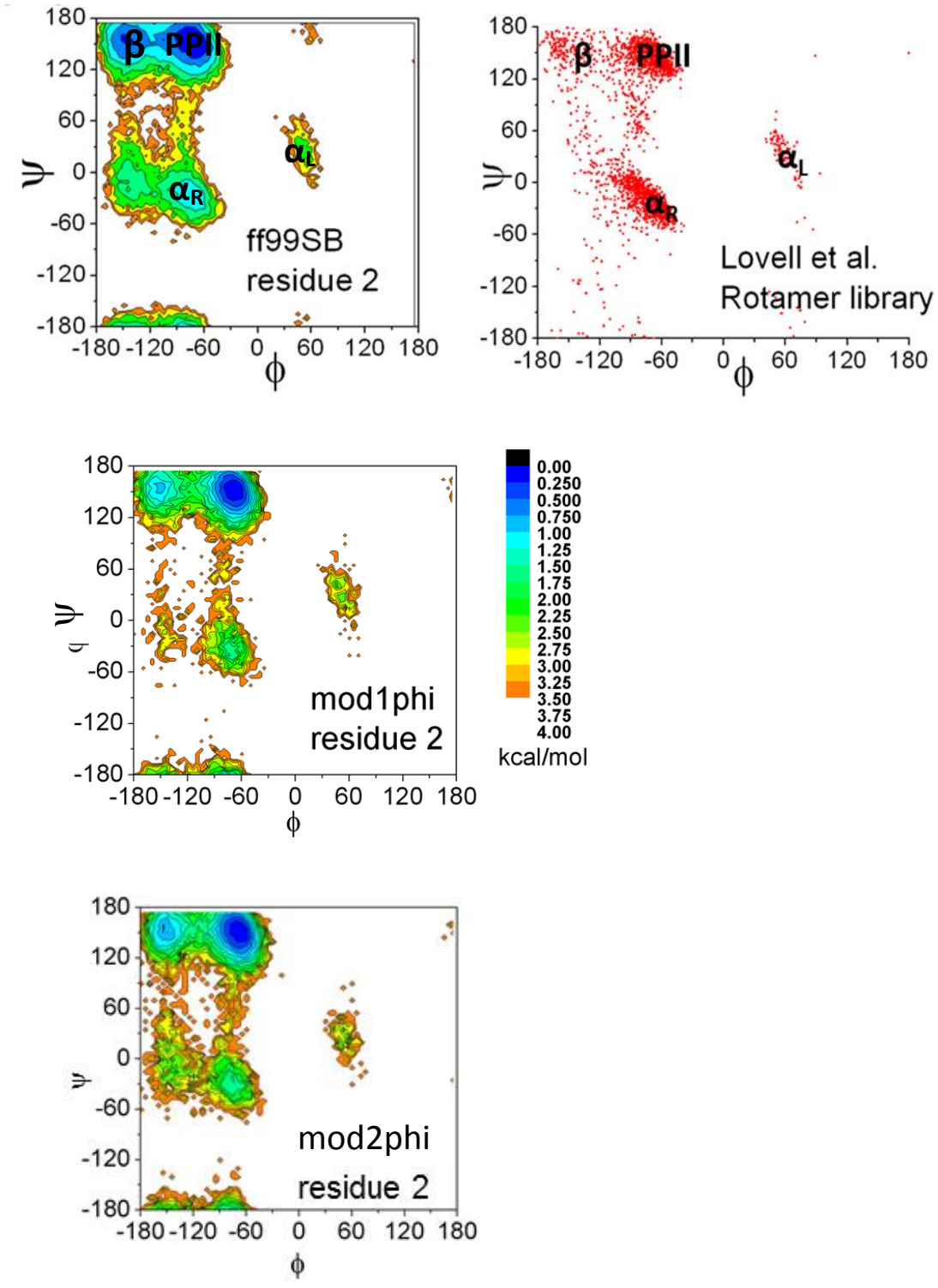


Figure 42 Comparison for backbone dihedral angle populations between ff99SB, mod1φ and mod2φ

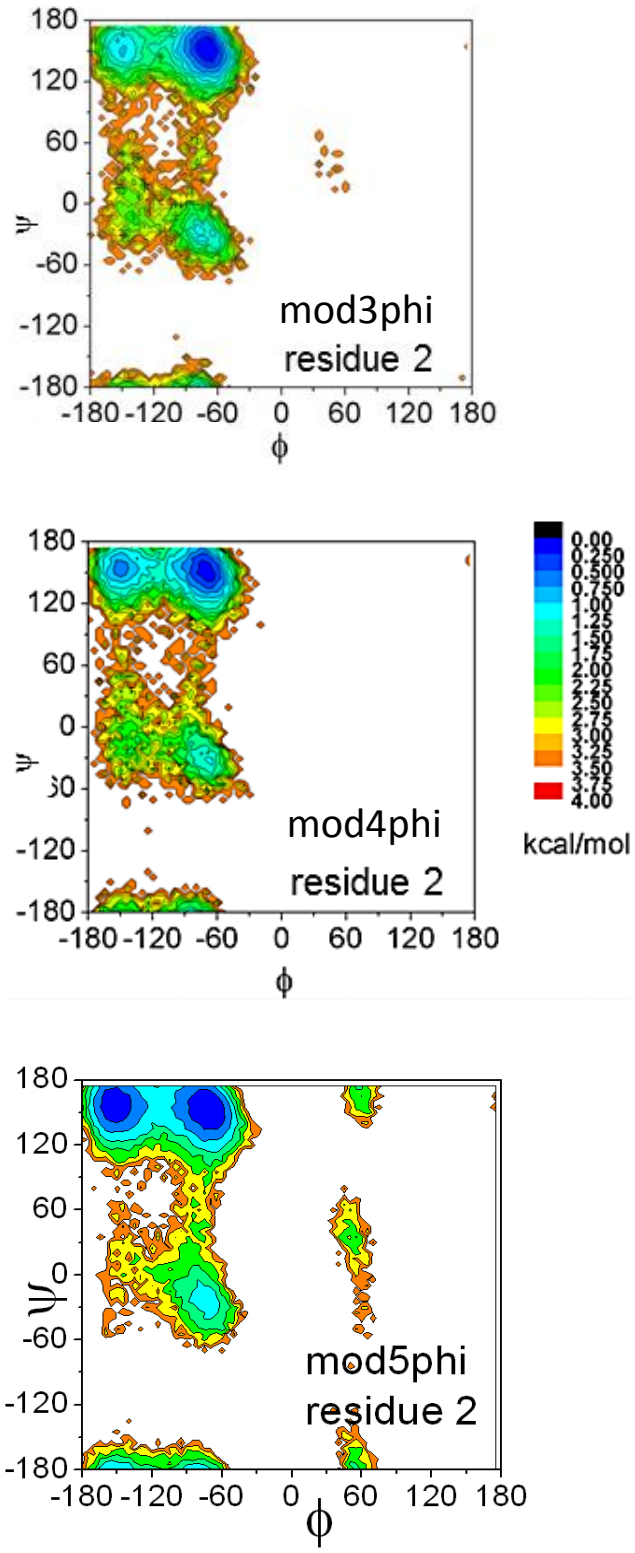


Figure 43 Comparison of backbone dihedral angle populations between mod3*φ*, mod4*φ* and mod5*φ*. Ramachandran plots for the rest of the modifications can be found in the appendix section figure 1

The Ramachandran maps show that sampling of the -150º ≤ φ ≤ -120º region decreased in comparison to ff99SB with the exception of mod5φ. Although, these are rough estimates for two reasons: First, there is still some dihedral sampling in the region next to the α-helical region. As mentioned previously, it is difficult to increase the sampling of the α-helical region without affecting the β-ppII equilibrium by only modifying the dihedral potentials for the φ angle (φ' term). As seen from the graphs even if the potentials for the ψ angle are modified along with the φ' term it is difficult to reduce the sampling in this region. To achieve this it would require that the sampling of the β region would be unfavorable in comparison to ppII which would destabilize β sheets in simulations of proteins.

Recent modifications to the force field have introduced "Gaussian" corrections that couple the φ and ψ terms of AMBER such as the work of Li et al. [155] [90] in order to correct for this problem without affecting the β - ppII equilibrium. This correction followed up on the CHARMM CMAP correction by Mackerell et al. in 2004 [148]. Although this type of corrections appear to solve some issues, there have been some concerns of over-stabilization of helices even at high temperatures [91]. It appears that these types of corrections are designed to fit particular regions in dihedral space without accounting for the effects in other parts of the force field function. Furthermore, for several years many groups have concentrated in changing backbone dihedral parameters of AMBER force fields, where the remainder of the force field terms, in particular the non-bonded terms (van der Waals and electrostatics) have been left untouched for almost 20 years. Recently, re-parameterization of the non-bonded terms have been attempted [157, 320] as well as the development of charge models [321] that required updated Lennard-Jones parameters that would fit the new charge model. It is known that the backbone dihedrals and non-bonded terms are influenced by each other [157] and only fitting the backbone dihedrals does not compensate for the deficiencies in the non-bonded terms of the force field.

Second, the rotamer library that is used for comparison is based on x-ray crystals and NMR structures deposited in the Protein Databank (PDB); effects such as crystal packing artifacts, experimental conditions and the dihedral propensities of alanine in proteins [322] are different than the peptide system that is described in the simulations presented here. This is because dihedral propensities of amino-acids are influenced by their neighboring residues and environment [322]. Nevertheless, for our purposes this comparison serves as a qualitative test; if

the dihedral sampling would resemble the major regions in the Ramachandran map the modifications would be appropriate for simulations of proteins.

The Ramachandran maps for the rest of the modifications are included in figure 45. The maps correspond to the simulations of the native conformations; the graphs for the simulations of the extended conformations are in the Appendix section. This graph was included here in order to describe the dihedral populations obtained and how they compare to the energy differences between ff99SB and the modifications (figure 46).

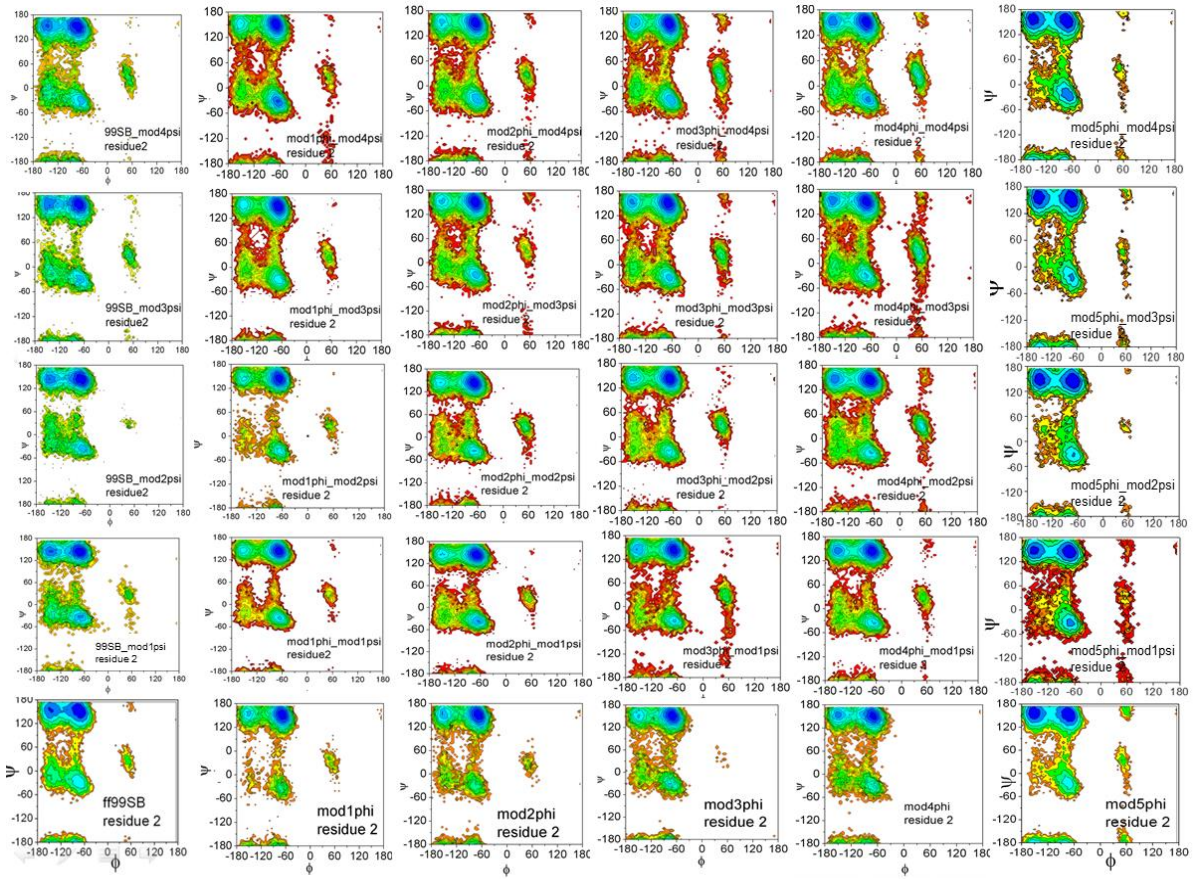


Figure 44 Backbone dihedral populations for all modifications and ff99SB

The backbone energy differences were calculated by subtracting the total dihedral energy for the backbone angles between ff99SB and modifications. These maps show how the set of modifications termed mod1-5φ were designed to make the region around φ = -60º favorable for sampling. This stabilizing effect was also observed for the φ = +60° (left handed α-helix) to a less extent, although some of the modifications (mod5φ) appear to make the region around φ =

+60º considerably more favorable than the rest. This effect is enhanced when the modifications to the ψ/ψ' were introduced where the modifications mod5φ-mod3ψ and mod5φ-mod4ψ show that the φ = +60° region is slightly more stable than the φ = -60° (figure 46); although this effect cannot be clearly established from the dihedral populations shown in figure 45. The modifications known as mod3φ-mod1ψ, mod3φ-mod3ψ, mod3φ-mod4ψ, mod4φ-mod2ψ, mod4φ-mod3ψ appear to have more sampling in the φ = +60°. To further investigate how the dihedral sampling of the simulations is being influenced by the modifications, we decided to measure the populations of the secondary structure basins. In order to define these basins, we used the definition previously used by Wickstrom et al [151] explained in the next section.

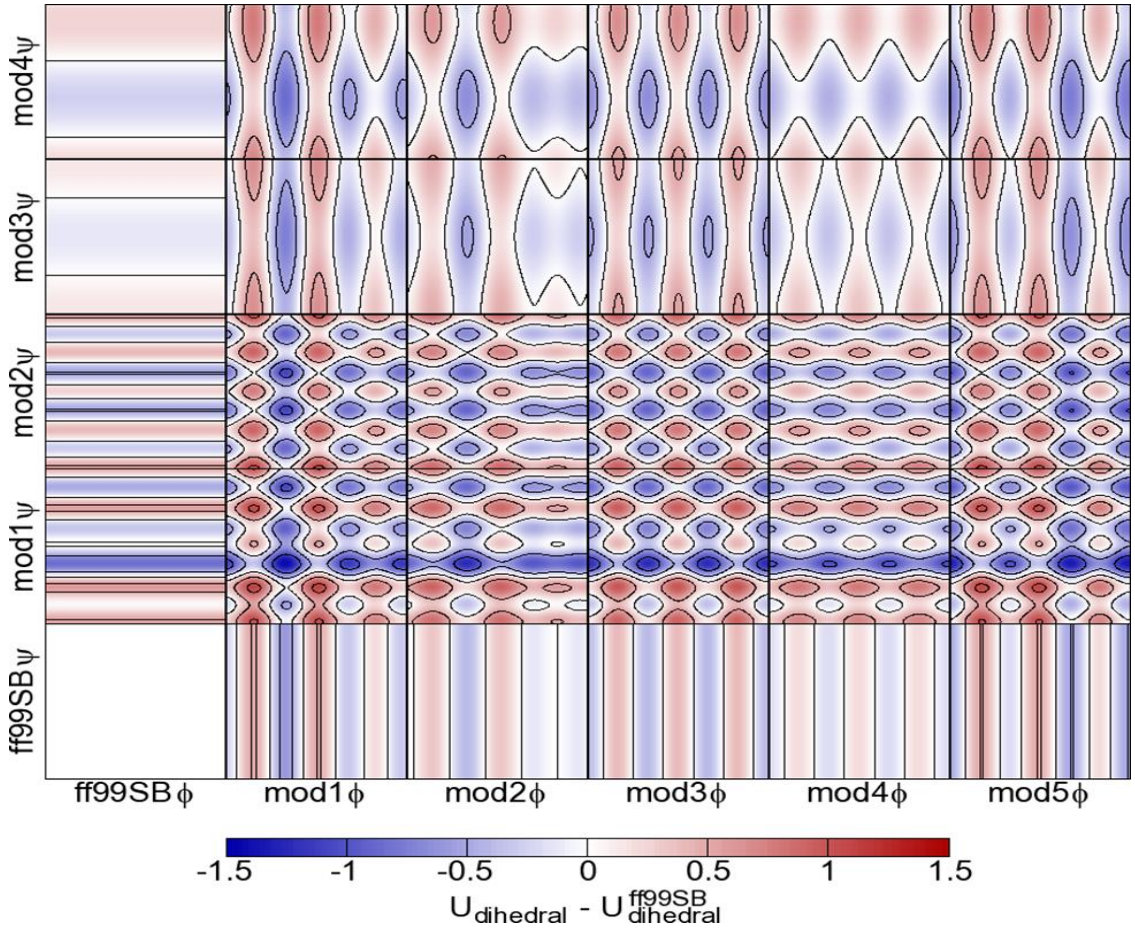


Figure 45 Backbone energy differences between ff99SB and the modifications. Each of the squares correspond to a dihedral map were the horizontal axis is the φ angle (-180° to 180°) and the vertical axis is the ψ angle (-180° to 180°). Graph generated by James Maier.

3.1.2 Secondary structure basin definitions

In order to determine how the secondary structure propensities were being affected by the modifications, we calculated the basin populations from simulations according to the definition given by Wickstrom et al. The definitions of the four principle regions were as follows: right handed helix (α_R), (φ,ψ) ~ (-160° to -20°, -120° to +50°); extended β-strand conformation, (-180° to -110°, 50° to 240°; or 160° to 180°, 110° to 180°); and poly-proline II, (-90° to -20°, 50° to 240°). The number of structures in individual regions were summed and divided by the total number of structures. Then, they were multiplied by 100 to get the percentages in each basin. Error bars were taken from native and extended runs (table 10 -13).

3.1.2.1 Basin populations with deviations for second and third residues in Ala_5

Table10 α_L basin percentages

Exp = 0	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	16.47± 4.98	23.02±3.91	20.50±3.18	19.93±1.97	18.96 3.55
Mod1φ	**5.65±5.59**	14.12±3.49	12.01±2.67	13.54±1.54	24.98±0.72
Mod2φ	10.36±1.36	19.26±1.56	14.10±1.22	17.30±2.27	19.52±0.55
Mod3φ	15.29±6.32	13.13±3.21	11.97±0.63	15.48±4.73	17.48±3.23
Mod4φ	7.05±5.83	14.07±3.07	13.82±1.49	15.58±0.75	18.92±0.12
Mod5φ	8.03±0.72	15.17±0.3	12.58±0.68	12.32±0.33	16.71±0.2

Table 10 α_R basins for residue 2 of Ala_5

From the basin populations of the α-helical region we can see that the least populated basin is mod1φ. According to the experimental results taken from vicinal scalar coupling measurements (Graf et al. [172]), where the authors reported fitting from these values to each of the major secondary structure regions in the Ramachandran plots (α_R, β, ppII and α_L); mod1φ has the best agreement with experiments. Although this comparison cannot be done directly because the basins measured here come from simulations of Ala_5 where there are more conformations than what is obtained from NMR measurements which is an ensemble average of conformations obtained at the given time. Furthermore, the vicinal scalar couplings are converted to φ/ψ angles using Karplus functions that have uncertainties as previously discussed. Nevertheless the trends

from the results of the modifications when compared to ff99SB show if the sampling of the given regions improved or got worse. The results for mod1φ were highlighted because this force field had the biggest population for the ppII basin as expected due to the rationale behind the modifications.

Table 11 β basin percentages

Exp =14	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	25.01±3.19	24.19±0.14	23.30±0.24	23.95±0.20	24.81±2.48
Mod1φ	**16.37±3.84**	12.76±0.10	12.49±1.27	12.82±1.75	11.59±1.14
Mod2φ	21.12±4.73	14.48±0.46	14.39±0.41	15.00±0.87	15.19±0.97
Mod3φ	15.41±1.60	21.14±1.27	18.11±0.61	17.42±1.52	17.43±0.29
Mod4φ	28.26±3.96	23.55±1.70	22.26±0.97	21.60±0.95	20.79±0.65
Mod5φ	36.85±0.81	32.46±0.44	33.50±0.37	34.16±0.11	32.32±0.50

Table 11 β basins for residues 2 of Ala_5

The β basins for mod1φ, mod2φ, mod3φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ, mod1φ-mod4ψ, mod2φ-mod1ψ, mod2φ-mod2ψ, mod2φ-mod3ψ, mod2φ-mod4ψ, mod3φ-mod1ψ, mod3φ-mod2ψ, mod3φ-mod3ψ and mod3φ-mod4ψ have comparable results.

Table 12 α_L basin percentages

Exp = 0	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	0.87±0.43	1.13±0.52	0.36±0.65	2.37±0.18	1.72±1.6
Mod1φ	**1.04± 0.50**	0.80±0.80	1.27±1.27	1.45±0.95	0.69±0.69
Mod2φ	0.77±0.18	0.88±0.88	2.50±2.34	0.83±0.64	1.36±0.85
Mod3φ	0.07±0.01	2.49±2.49	3.15±3.06	6.59±4.13	3.86±2.12
Mod4φ	0.00±0.00	1.35±1.99	4.42±0.06	4.48±0.22	4.34±0.03
Mod5φ	0.41±0.44	0.47± 0.14	0.28±0.59	0.12±0.52	0.17±0.48

Table 12 α_L basins for residue 2 of Ala_5

Similar to the results of the β basins, the results for the α_L basins are comparable between modifications which indicate that the modifications are performing similarly in the β and α_L regions. The only regions that can help to distinguish between the modifications are the ppII and α_R regions where is clear that mod1φ had one of the best samplings for the ppII region. It is known that poly-alanine peptides have ppII secondary structure as reported experimentally [256] although as they elongate they adopt helical secondary structure [242], which indicates good agreement with experiments.

Table 13 ppII basin percentages

Exp = 86	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	51.53±2.01	48.83±3.38	51.78±4.14	47.37±3.04	48.59±2.55
Mod1φ	**71.78±2.20**	68.11±3.07	69.69±2.90	66.21±4.09	55.40±0.5
Mod2φ	61.88±4.39	60.90±2.8	63.72±4.73	59.73±1.75	56.67±0.36
Mod3φ	63.98±3.82	58.70±4.94	61.50±3.28	52.3±3.09	53.64±6.2
Mod4φ	59.14±2.01	56.52±0.61	53.39±2.86	50.59±1.05	48.09±1.19
Mod5φ	46.89±0.95	48.42±0.67	48.70±0.43	44.92±0.70	41.88±0.40

Table 13 ppII basins for residue 2 of Ala_5. The force field that had the most ppII percentage was mod1φ (highlighted)

We also investigated the secondary structure basins for the central residue (residue 3) in Ala_5. Although the results were different than what was observed for residue 2, the populations for the ppII region of mod1φ, mod2φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ, mod1φ-mod4ψ, mod2φ-mod1ψ, and mod2φ-mod2ψ were comparable. In the case of the helical content, the values were similar between all the modifications.

Table 14 α_R basin percentages

Exp=0	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	15.63± 1.66	17.30±3.67	20.50±3.18	19.93±1.97	18.96 3.55
Mod1φ	16.02±1.85	17.62±3.49	12.01±2.67	13.54±1.54	24.98±0.72
Mod2φ	14.52±0.15	17.70±1.56	14.10±1.22	17.30±2.27	19.52±0.55
Mod3φ	20.55±5.87	9.92±3.21	11.97±0.63	15.48±4.73	17.48±3.23
Mod4φ	7.49±10.56	10.99±3.07	13.82±1.49	15.58±0.75	18.92±0.12
Mod5φ	23.75±1.27	13.95±0.33	12.58±0.68	12.32±0.33	16.71±0.2

Table 14 α_R basins for central residue of Ala_5

Table 15 β basin percentages

Exp=16	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	25.23±1.55	24.19±0.14	23.30±0.24	23.95±0.2	24.81±2.48
Mod1φ	11.66±1.16	12.76±0.10	12.49±1.27	12.82±1.75	11.59±1.14
Mod2φ	14.80±2.87	14.48±0.46	14.39±0.41	15.00±0.87	15.19±0.97
Mod3φ	12.78±0.38	21.14±1.27	18.11±0.61	17.42±1.52	17.43±0.29
Mod4φ	19.59±0.22	23.55±1.7	22.26±0.97	21.60±0.95	20.79±0.65
Mod5φ	31.09±1.50	32.46±0.44	33.50±0.37	34.16±0.11	32.32±0.50

Table 15 β basins for central residue of Ala_5

Similar to the results observed from residue 2, the results for the β basins for mod1φ, mod2φ, mod3φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ, mod1φ-mod4ψ, mod2φ-mod1ψ, mod2φ-mod2ψ, mod2φ-mod3ψ, mod2φ-mod4ψ, mod3φ-mod1ψ, mod3φ-mod2ψ, mod3φ-mod3ψ and mod3φ-mod4ψ were close to each other. The α_L basins also had comparable results to the results from residue 2.

Table 16 α_L basin percentages

Exp=0	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	7.25±2.64	1.13±0.52	0.36±0.65	2.37±0.18	1.72±1.6
Mod1φ	2.28± 1.56	0.80±0.80	1.27±1.27	1.45±0.95	0.69±0.69
Mod2φ	4.16±0.90	0.88±0.88	2.50±2.34	0.83±0.64	1.36±0.85
Mod3φ	1.68±1.12	2.49±2.49	3.15±3.06	6.59±4.13	3.86±2.12
Mod4φ	1.68±3.57	1.35±1.99	4.42±0.06	4.48±0.22	4.34±0.03
Mod5φ	2.91±0.39	0.47± 0.14	0.28±0.59	0.12±0.52	0.17±0.48

Table 16 α_L basins for central residue of Ala_5

The results for the ppII basins from the central residue of Ala_5 were similar to the residue 2 results. The mod1φ through mod3φ results were comparable, as well as mod1φ-mod1ψ, mod1φ-mod2ψ, and mod1φ-mod3ψ results. This trend was also observed for the results of residue 2.

Table 17 ppII basin percentages

Exp =84	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	44.84±3.93	48.83±3.38	51.78±4.14	47.37±3.04	48.59±2.55
Mod1φ	63.84±2.69	68.11±3.07	69.69±2.90	66.21±4.09	55.40±0.5
Mod2φ	61.55±2.06	60.90±2.80	63.72±4.73	59.73±1.75	56.67±0.36
Mod3φ	60.10±7.43	58.70±4.94	61.50±3.28	52.3±3.09	53.64±6.2
Mod4φ	65.71±13.81	56.52±0.61	53.39±2.86	50.59±1.05	48.09±1.19
Mod5φ	35.24±0.28	48.42±0.67	48.70±0.43	44.92±0.70	41.88±0.40

Table 17 ppII basins for central residue of Ala_5

3.1.3 Potential Mean Force (PMF) energy maps for alanine dipeptide

In order to further investigate the regions of the Ramachandran map we built Potential Mean Force (PMF) energy maps for alanine dipeptide (N-acetyl alanyl-N-methyl-amide) using umbrella sampling calculations. This was done in order compare how the modifications were refining the secondary structure regions for some of the force fields that had the best results for the ppII basin percentages calculations. Here we are plotting mod1φ and mod2φ and comparing them to the results of ff99SB (figures 47-49).

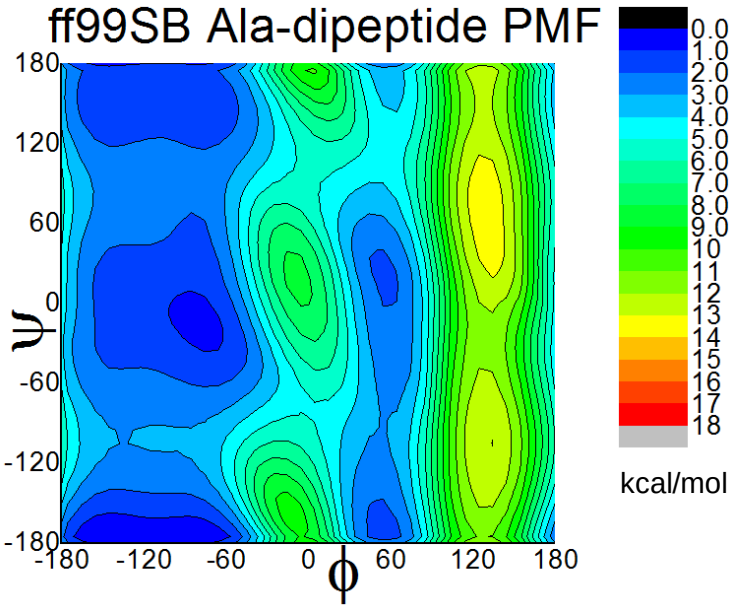


Figure 46 Potential Mean Force (PMF) dihedral map for Alanine dipeptide (N-acetyl alanyl-N-methyl-amide) calculated using umbrella sampling calculations.

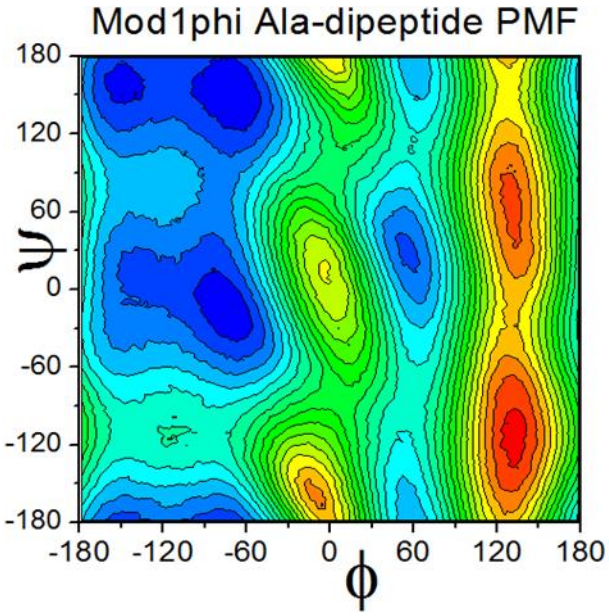


Figure 47 Potential Mean Force (PMF) energy maps for alanine dipeptide (N-acetyl alanyl-N-methyl-amide) calculated using umbrella sampling calculations.

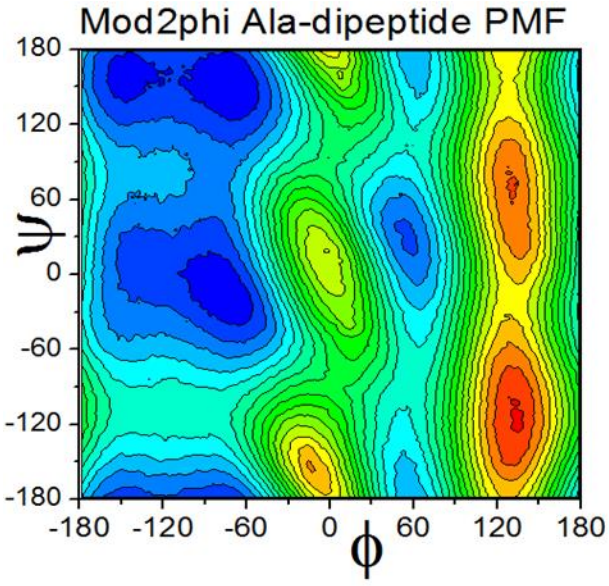


Figure 48 Potential Mean Force (PMF) energy maps for alanine dipeptide (N-acetyl alanyl-N-methyl-amide) calculated using umbrella sampling calculations

The energy maps show how the modifications define the β/ppII regions more than ff99SB. The helical region is also more defined than ff99SB. The modifications increased the energy barrier between these two regions by approximately 1 kcal/mol. The α_L region is also more defined and restricted to the $0 \leq \psi \leq 60$ region than ff99SB. These estimates are useful to

determine how the energy barriers between the basins were affected with the introduction of the modifications; nevertheless, they cannot be quantitatively compared to experimental results directly. In order to have a quantitative way of describing how the dihedral populations were affected by these modifications we used the vicinal scalar coupling analysis against the experimental values.

3.1.4 Scalar coupling values for Ala_5

Initially we tested the agreement with the ^3J (H_N, H_α) scalar coupling that probes for the φ angle. Figure 45 shows how the overlap between the dihedral populations for the modifications of the φ' angle and the ^3J (H_N, H_α) scalar coupling function. We can see that the dihedral populations for the φ angle increase in the φ = -60º region. This region has coupling values from 2 to 5 Hz as indicated in the figure 45. The average value obtained for simulations with ff99SB was 6.99 Hz which is higher than the experimental value of 5.59 Hz described by Graf et al. [172].

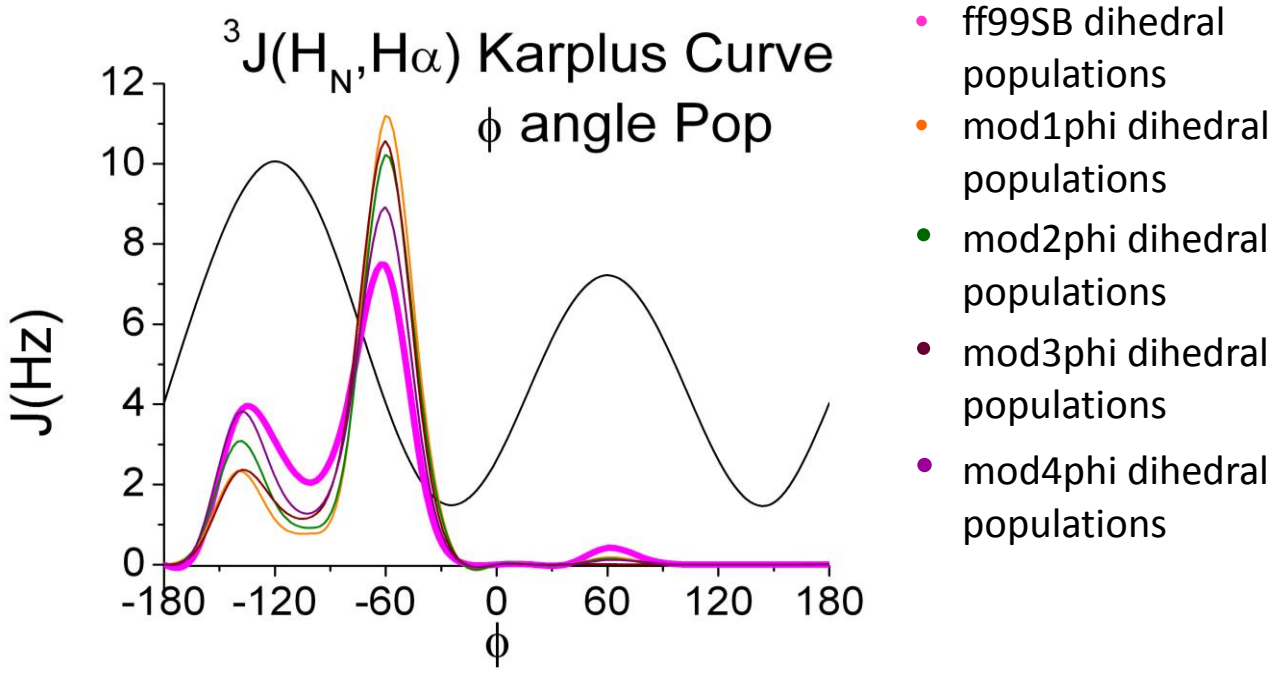


Figure 49 Comparison between dihedral populations for some of the modifications of φ'. The Karplus curve is included here in order to see how the sampling in the φ = -60 region increases the scalar coupling values in the range 2-5 Hz.

The ^3J (H_N, H_α) scalar coupling was chosen for this testing because it has a wide range that defines the angles between the β and the ppII/α-helix regions. Table 18 shows the calculated

vicinal scalar coupling values for all the simulations of the modifications proposed here. The results indicated that the simulation with mod1φ-mod1ψ had the least disagreement with experiments. Simulations with mod1φ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ had similar results to mod1φ-mod1ψ.

Table 18

$^3J(H_N,H_\alpha)$ values for residue 2 of Ala_5

Experiment 5.59	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	6.99±0.04	6.85±0.00	6.85±0.06	7.09±0.1	6.96±0.00
Mod1φ	**5.93±0.00**	**5.70±0.03**	**5.78±0.05**	**5.87±0.06**	**5.85±0.06**
Mod2φ	6.25±0.1	6.05±0.02	6.06±0.01	6.28±0.04	6.24±0.02
Mod3φ	6.24±0.00	6.14±0.01	6.13±0.03	6.06±0.20	6.15±0.05
Mod4φ	6.61±0.02	6.48±0.04	6.51±0.02	6.61±0.03	6.58±0.01
Mod5φ	6.98±0.01	6.71±0.00	6.81±0.02	6.93±0.01	6.87±0.00

Table 18 ^3J (H_N, H_α) values calculated from simulations using the modifications proposed. The best values were mod1φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ. Uncertainties were taken from native and extended simulations.

3.1.4.1 Caveats of using ^3J (H_N, H_α) scalar couplings to judge secondary structure propensities of force fields

Some of the disadvantages of using scalar couplings and Karplus functions to benchmark force fields have been discussed here previously; however it is important to re-visit these concepts in order to clarify the significance of the results shown here. Karplus functions have been used for evaluating the quality of force fields, however, Karplus functions attempt to represent spin/spin interactions of nuclei in a two dimensional manner. As described by Karplus et al. there are a significant amount of uncertainties and caveats that need to be taken into

account [174], in our case we are interested in the uncertainties of the ^{3}J (H_N, H_α) scalar couplings because they are used to benchmark the modifications described here. For this purpose we used the work of Salvador et al.[323] where they calculated *J*-coupling constants using density functional theory [323] for acetyl-$(Ala)_3$-NH_2 capped tri-alanine peptide. As shown in figure 51, a scalar coupling value of 9 Hz can cover a wide range of dihedral angles in the dihedral map and small changes of 0.5 Hz between the values can be associated with different secondary structure regions. This graph shows that ^{3}J (H_N, H_α) scalar couplings are not only dependent on the φ dihedral angle, but on the ψ angle as well. Therefore the uncertainties of the values calculated here can be as significant as the differences in the scalar coupling values between the modifications.

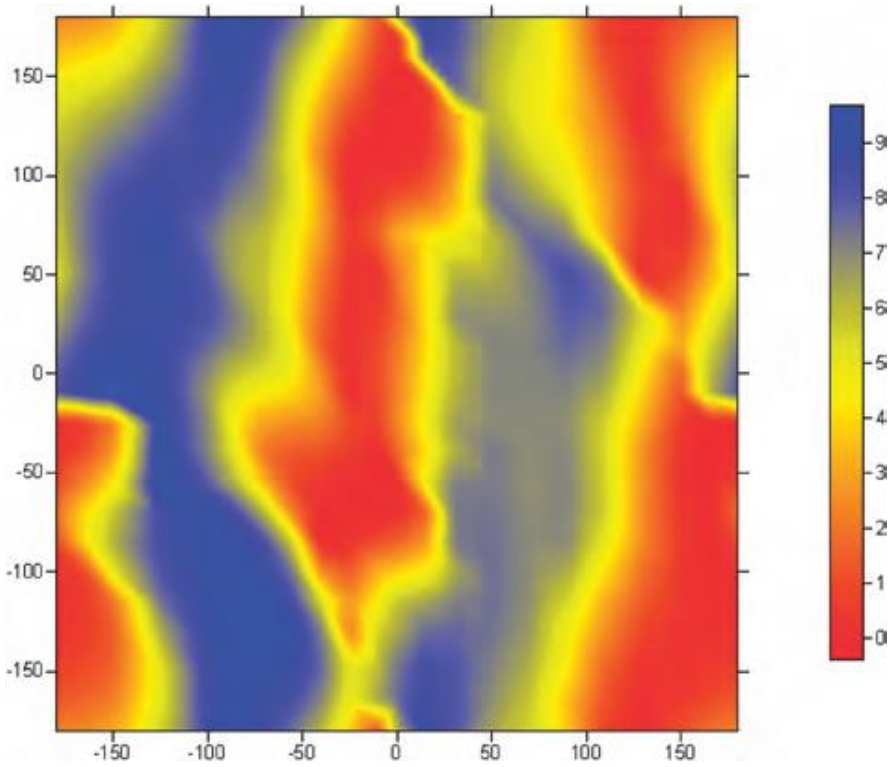


Figure 50 Dihedral map that shows the relation between ^{3}J (H_N , H_α) coupling values and the φ/ψ dihedral angles. Graph taken from the work of Salvador, Tsan and Dannenberg [323]

3.1.4.2 χ^2 values calculated from J-coupling constants values for residues 2-4 of Ala_5

In order to use more *J*-coupling constants besides ^3J (H_N , H_α), we calculated χ^2 values using equation 30. The *J*-coupling constants used for the calculation were $^3J(H_N,H_\alpha)$, $^3J(H_N,C_\beta)$, $^3J(H_N,C')$, $^3J(C',H_\alpha)$ and $^3J(C',C')$ that probe for the φ angle and $^1J(N,C_\alpha)$ and $^2J(N,C_\alpha)$ that probe for ψ. The uncertainties were taken from native and extended runs.

Table 19 χ^2 values for residues 2-4 of Ala_5

	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	1.80±0.1	1.80±0.1	1.90±0.05	1.80±0.1	2.00±0.1
Mod1φ	**0.84±0.02**	**1.00±0.01**	**0.91±0.05**	**1.00±0.07**	1.40±0.07
Mod2φ	**0.89±0.02**	1.10±0.1	1.10±0.00	1.20±0.03	1.50±0.01
Mod3φ	1.23±0.05	1.20±0.01	1.30±0.07	1.40±0.01	1.70±0.1
Mod4φ	1.30±0.00	1.50±0.04	1.70±0.07	1.70±0.02	1.80±0.02
Mod5φ	1.60±0.04	1.70±0.07	2.5±0.29	1.70±0.06	2.00±0.08

Table 19 χ^2 values for residues 2-4 of Ala_5. Uncertainties were taken from native and extended simulations

From these results we can conclude that the mod1φ and mod2φ had the lowest χ^2 values for all the modifications. Although, as discussed previously, these results have significant uncertainties since Karplus functions were used for the derivation of the *J*-coupling values shown here. We also used other Karplus function methods known as DFT1 and DFT2 where the parameters are given in detail in tables 5-8. From these calculations we could see that the lowest χ^2 values were obtained with the Original Karplus parameters, followed by the DFT2 parameters. The DFT1 parameters had the highest χ^2 values overall. As previously reported by Best et al [152] the results are dependent on the Karplus parameters used for the calculations.

Table 20 χ^2 values calculated with DFT1, DFT2 and Original parameters

Force Field	DFT1	DFT2	Original
99SB	1.44±0.02	1.62±0.02	1.81±0.00
99SB_mod1ψ	1.64±0.38	1.75±0.11	1.80±0.10
99SB_mod2ψ	1.67±0.04	1.83±0.08	1.90±0.05
99SB_mod3ψ	1.60±0.05	1.83±0.13	1.80±0.10
99SB-mod4ψ	1.80±0.06	2.03±0.12	2.00±0.10
Mod1φ_ff99SB	2.64±0.05	**1.17±0.02**	**0.84±0.02**
Mod2φ_ff99SB	1.68±0.14	**1.12±0.02**	**0.89±0.02**
Mod3φ_ff99SB	2.27±0.01	**1.08±0.01**	1.23±0.05
Mod4φ_ff99SB	1.48±0.01	1.21±0.03	1.38±0.05
Mod5φ_ff99SB	2.04±0.03	1.92±0.02	1.60±0.04
Mod1φ_mod1ψ	3.09±0.02	1.45±0.01	**1.00±0.01**
Mod1φ_mod2ψ	2.92±0.07	1.31±0.01	**0.91±0.05**
Mod1φ_mod3ψ	2.69±0.13	1.30±0.07	**1.00±0.07**
Mod1φ_mod4ψ	3.03±0.11	1.66±0.07	1.40±0.07
Mod2φ_mod1ψ	2.44±0.22	1.24±0.12	1.10±0.10
Mod2φ_mod2ψ	2.28±0.14	1.20±0.00	1.10±0.00
Mod2φ_mod3ψ	2.32±0.03	1.25±0.02	1.20±0.03
Mod2φ_mod4ψ	2.24±0.14	1.44±0.01	1.50±0.01

Mod3φ_mod1ψ	2.08±0.03	1.18±0.00	1.20±0.01
Mod3φ_mod2ψ	1.62±0.04	1.13±0.01	1.30±0.00
Mod3φ_mod3ψ	1.93±0.14	1.22±0.06	1.40±0.01
Mod3φ_mod4ψ	1.92±0.02	1.29±0.19	1.70±0.10
Mod4φ_mod1ψ	1.79±0.02	1.39±0.05	1.50±0.04
Mod4φ_mod2ψ	1.60±0.03	1.49±0.06	1.70±0.07
Mod4φ_mod3ψ	1.51±0.02	1.45±0.03	1.70±0.02
Mod4φ_mod4ψ	1.84±0.02	1.65±0.01	1.80±0.02
Mod5φ_mod1ψ	2.35±0.01	1.82±0.04	1.70±0.07
Mod5φ_mod2ψ	2.19±0.07	1.80±0.00	2.50±0.29
Mod5φ_mod3ψ	2.37±0.00	2.01±0.02	1.70±0.06
Mod5φ_mod4ψ	2.58±0.01	2.16±0.02	2.00±0.08

Table 20 χ^2 values calculated with DFT1, DFT2 and Original parameters. Deviations are calculated from native and extended simulations. The highlighted values indicate the modifications that had the best agreement with experimental scalar coupling values.

As seen on Table 20 the χ^2 results vary among the different Karplus coefficients used. The reason for this discrepancy is due to the differences among the Karplus curves. In certain cases these curves can differ by as much as 2 Hz in the amplitude of the function as seen in figures 52-55. This effect was previously reported[235] and is related to the dihedral conformations of the peptides used for the parameterization of the DFT1 and DFT2 coefficients. Interestingly the regions were some of the curves differ the most as in the case of the $^3J(H_N,C_\beta)$ coupling value is the φ = -60º region where the modifications are increasing the sampling.

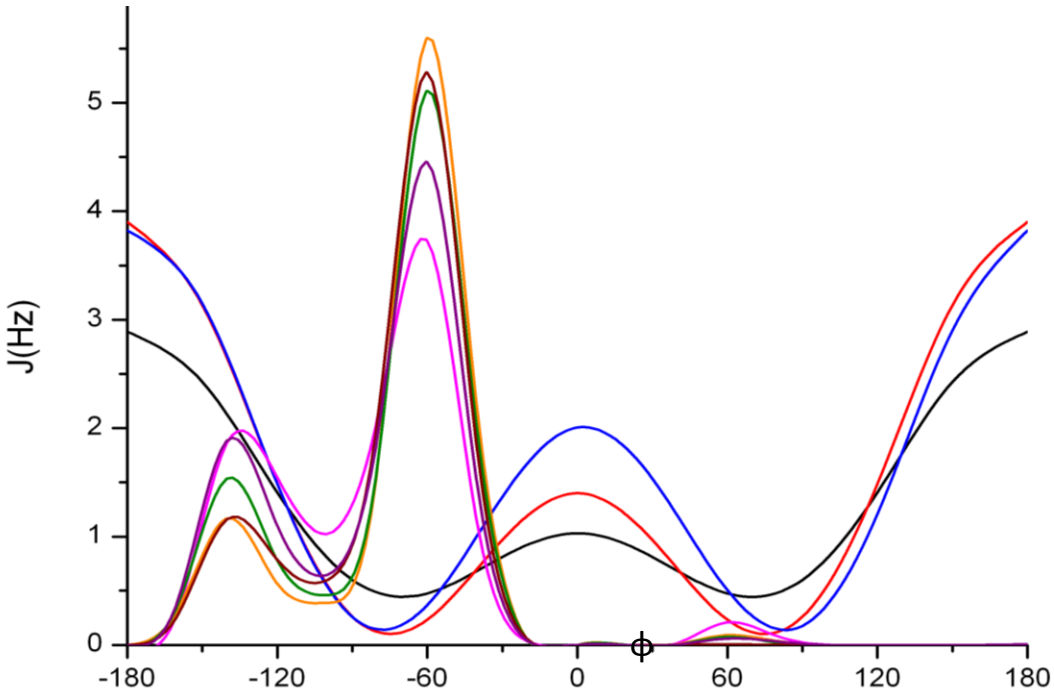


Figure 51 Karplus functions for DFT1 (blue), DFT2 (red) and Original (black) for ^3J (C, C') vicinal scalar coupling. The other curves (magenta, dark purple, green, brown and orange) correspond to the dihedral populations of φ for ff99SB, mod4φ, mod3φ, mod2$\varphi$and mod1φ respectively. The dihedral populations were included here as a reference to indicate how the dihedral sampling is being influenced by the modifications. The actual values for the populations of ff99SB, mod1φ and mod2φ obtained from simulations with error bars are included in figure 42.

Figure 52 illustrates how the Karplus functions differ among them, especially the DFT1/DFT2 curves compared to the original parameters. In certain regions like around φ = 0° the curves can differ from each other by almost 1 Hz. For our purposes, this region does not have significant populations in our simulations. For the regions of interest such as φ = -60° and -150 ≤ φ ≤ -120° the curves do not differ significantly. Likewise we investigated the results for the ^3J (H_N , H_α) Karplus curve shown in figure 53.

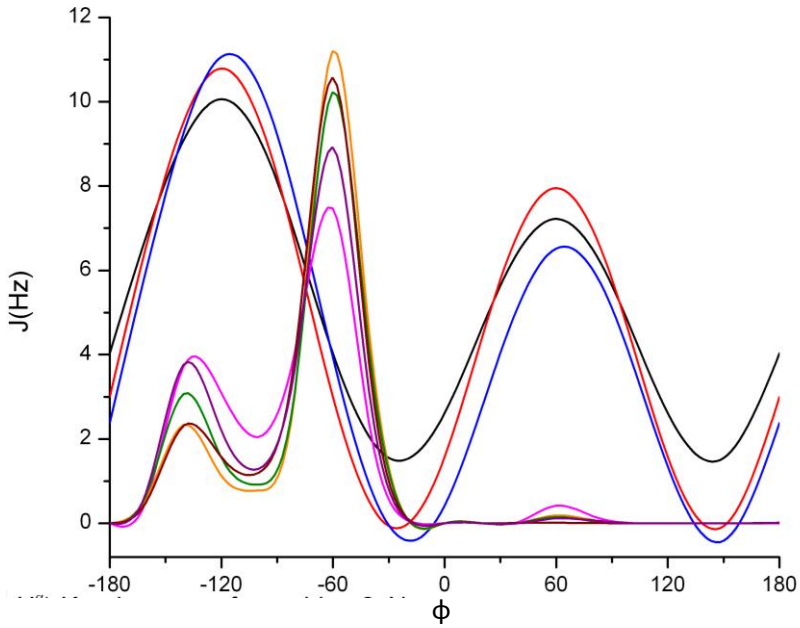


Figure 52 Karplus curves for DFT1 (blue), DFT2 (red) and Original (black) of ^3J (H_N, H_α) vicinal scalar coupling. The other curves (magenta, dark purple, green, brown and orange) correspond to the dihedral populations of φ for ff99SB, mod4φ, mod3φ, mod2$\varphi$and mod1φ respectively. The Karplus curves were smoothed between the data points, therefore they appear to have values below the zero mark which is not the case for DFT1 and DFT2 parameters. The dihedral populations were included here as a reference to indicate how the dihedral sampling is being influenced by the modifications. The actual values for the populations of ff99SB, mod1φ and mod2φ obtained from simulations with error bars are included in figure 42.

The Karplus curves for the ^3J (H_N, H_α) vicinal scalar coupling also appear to differ significantly in certain regions in particular in the φ = -120°, -30° and 150° where the curves are at maxima and minima. The results obtained from these dihedral regions in particular the region where φ = -30°, can be misleading because the curves differ by almost 2 Hz. Nevertheless the reason why the experimental values for the residues of Ala_5 do not agree well with the results from simulations with ff99SB is due to the sampling of the $-150 \leq \varphi \leq -120°$ where the scalar values are 8 – 11Hz. Having significant sampling in this region would make the average of the scalar coupling high in comparison to experimental values (5.5 – 5.7 Hz). In the region where φ = -30° the scalar coupling values range is 0 -2 Hz where increased sampling in this region would lower the average of the scalar coupling values.

Similarly to this case, the ^3J (H_α, C′) shown in figure 54 the curves differed at the maxima and minima. This scalar coupling is sensitive in the α_L and α_R regions where the ^3J (H_N,

H_α) is not as sensitive. The differences for this curve were not as significant as the rest of the scalar coupling curves.

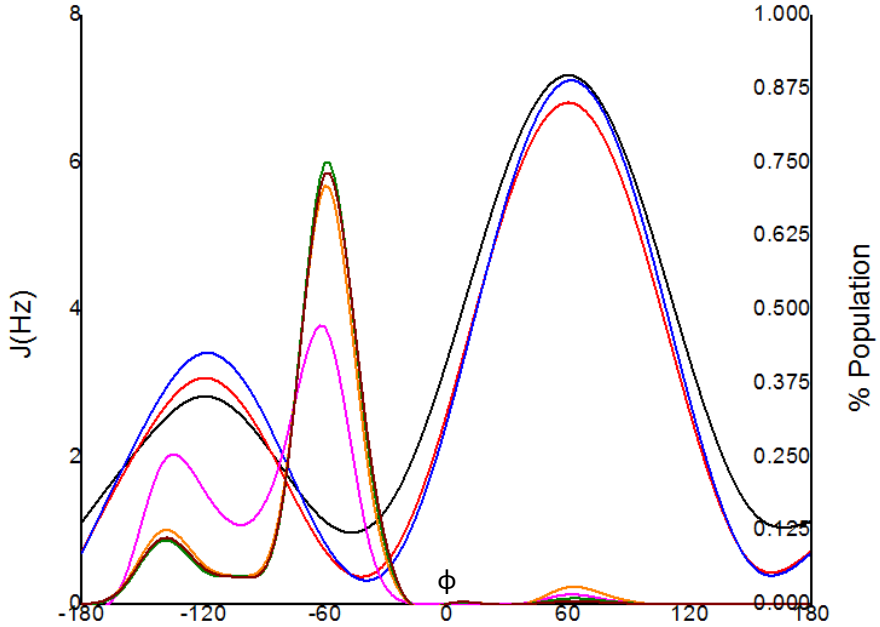


Figure 53 Karplus curves for DFT1 (blue), DFT2 (red) and original (black) for the $^3J(H_\alpha, C')$ vicinal scalar coupling. The other curves (magenta, dark purple, green, brown and orange) correspond to the dihedral populations of φ for ff99SB, mod4φ, mod3φ, mod2$\varphi$and mod1φ respectively. Contrary to the results shown for the rest of the curves the dihedral populations shown here correspond to the simulations of the extended conformations. The dihedral populations shown for the other Karplus curves were the results from the native conformation simulations. The dihedral populations were included here as a reference to indicate how the dihedral sampling is being influenced by the modifications. The actual values for the populations of ff99SB, mod1φ and mod2φ obtained from simulations with error bars are included in figure 42.

In the case of the $^3J(H_N,C_\beta)$ the differences were significantly noticeable in particular in the φ = -60° region. This feature was previously discussed in the simulation methods section. Nevertheless, is necessary to re-visit this concept because this scalar coupling is the major contributor to the differences in the χ^2 values between the DFT1, DFT2 and Original parameters. As seen in figure 55, in the φ = -60° region the difference between the curves can be as much as 3 Hz. These differences contribute significantly to the deviation calculation which affects the χ^2 comparison between the parameters because the modifications are designed to increase the sampling of the φ = -60° region.

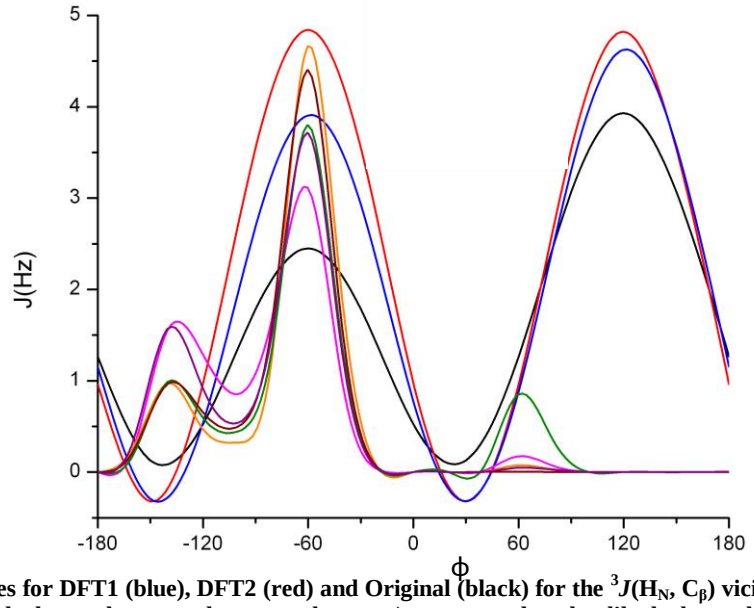


Figure 54 Karplus curves for DFT1 (blue), DFT2 (red) and Original (black) for the $^3J(H_N, C_\beta)$ vicinal scalar coupling. The other curves (magenta, dark purple, green, brown and orange) correspond to the dihedral populations of phi for ff99SB, mod4phi, mod3phi, mod2

Case et al. [235] reported that most of the quantum-calculated Karplus functions (DFT1 and DFT2) agree well with empirically derived Karplus functions with the exception of $^3J(H_N,C_\beta)$ that couples φ at -60°. The deviations between the empirical and quantum Karplus functions can be as high as 3 Hz as shown in figure 55. The authors argue that these discrepancies are due to motional averaging effects and to the size/identity of the side-chain used for the quantum calculation. Calculations with serine and higher basis set gave better agreement between the empirical, DFT1 and DFT2 parameters.

The lowest χ^2 values were obtained with the DFT2 and original parameters. Overall the best results were obtained with mod1φ, mod2φ, mod3φ mod1φ-mod1ψ, mod1φ-mod2ψ and mod1φ-mod3ψ. We also ran molecular dynamic simulations of Ala_5 because of the long range electrostatics cutoff used in the REMD simulations. The REMD simulations were ran with a 6.0Å cutoff, since the protocol published by Wickstrom et al. [151] used an 8.0Å cutoff for their simulations. We investigated the impact of the cutoff differences in the calculation of the χ^2 values. The results were comparable to what was observed from the REMD simulation (figures 56-60).

The graphs were plotted by sets of ψ/ψ' modifications, the error bars were obtained from native and extended MD simulations.

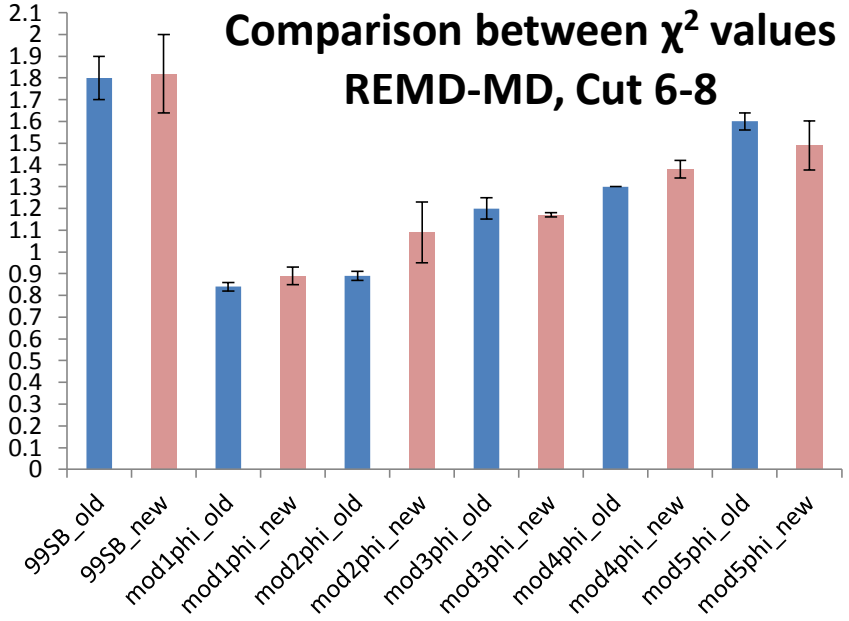


Figure 55 Comparison between χ^2 values for REMD and MD simulations of the modifications 99SB and mod1φ through mod5φ. The old data is the REMD data where the new data is the MD data.

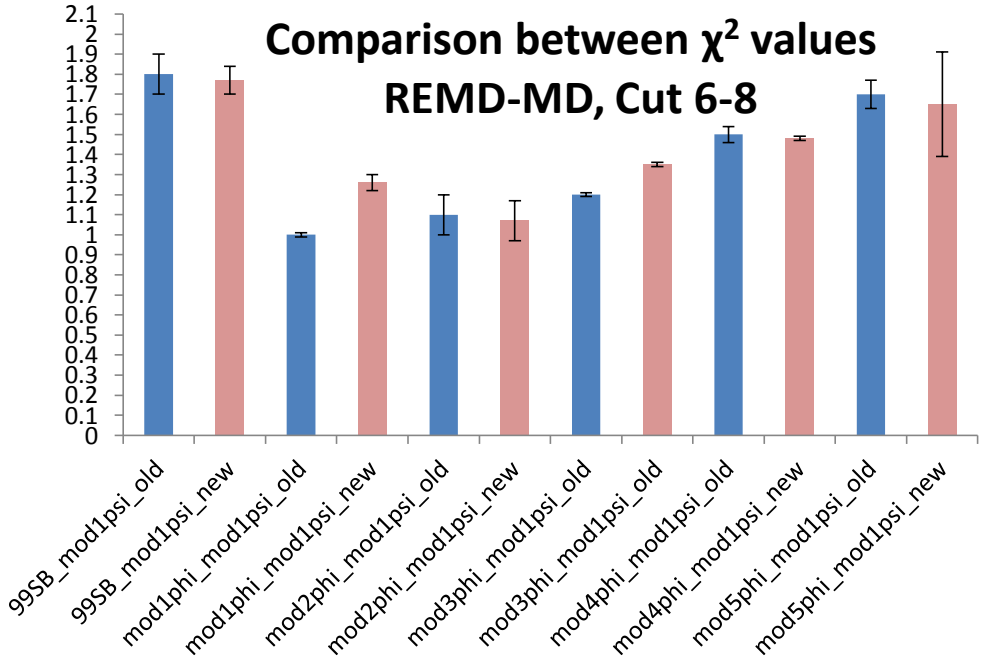


Figure 56 Comparison between χ^2 values for REMD and MD simulations of the modifications 99SB_mod1ψ through mod5φ_mod1ψ. The old data is the REMD data where the new data is the MD data.

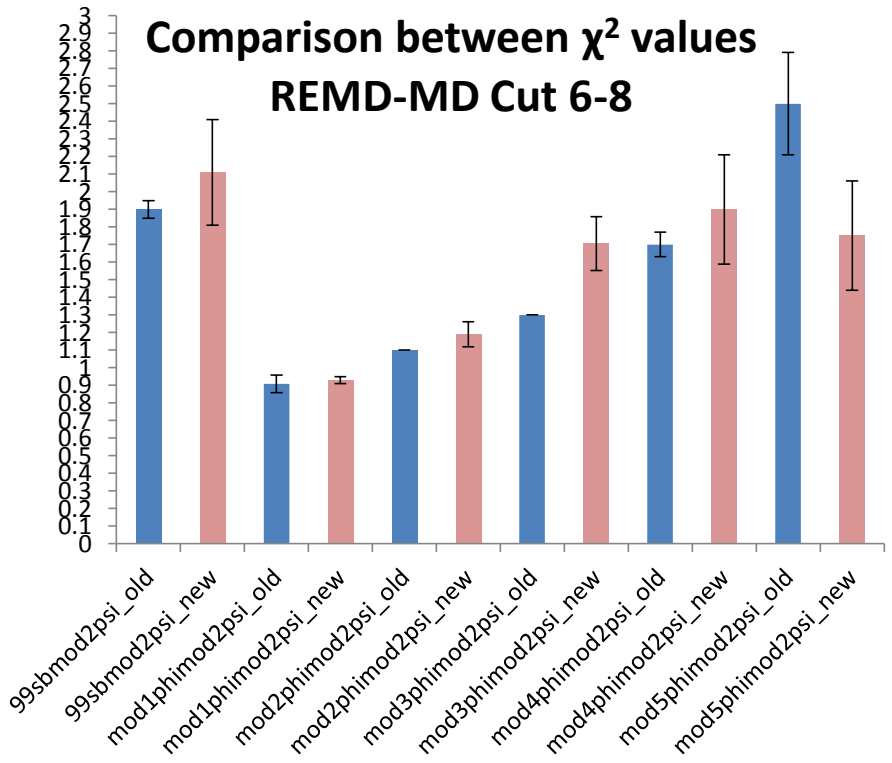


Figure 57 Comparison between χ^2 values for REMD and MD simulations of the modifications 99SB_mod2ψ through mod5φ_mod2ψ. The old data is the REMD data where the new data is the MD data.

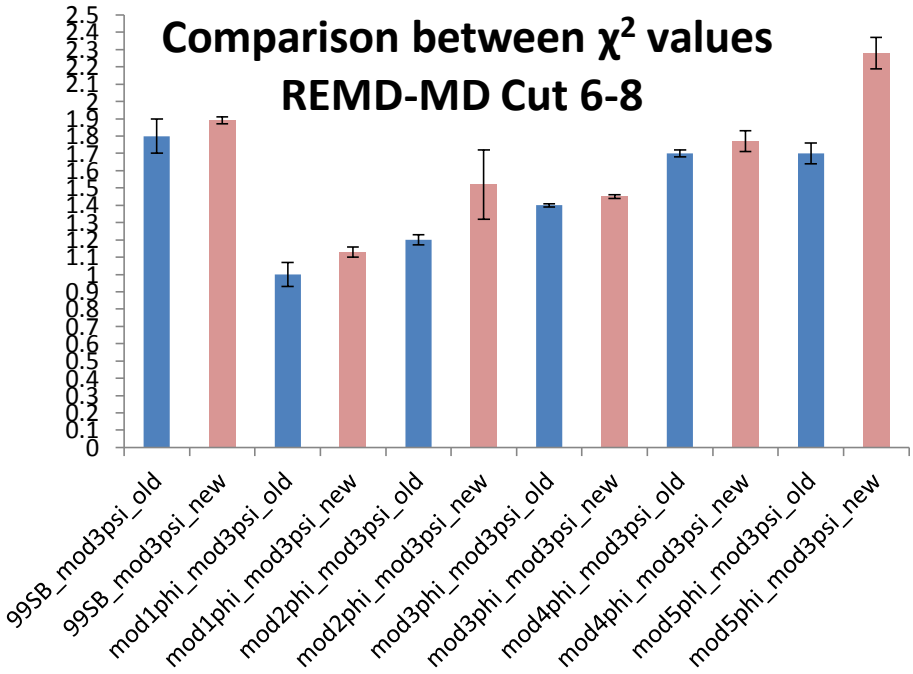


Figure 58 Comparison between χ^2 values for REMD and MD simulations of the modifications 99SB_mod3ψ through mod5φ_mod3ψ. The old data is the REMD data where the new data is the MD data.

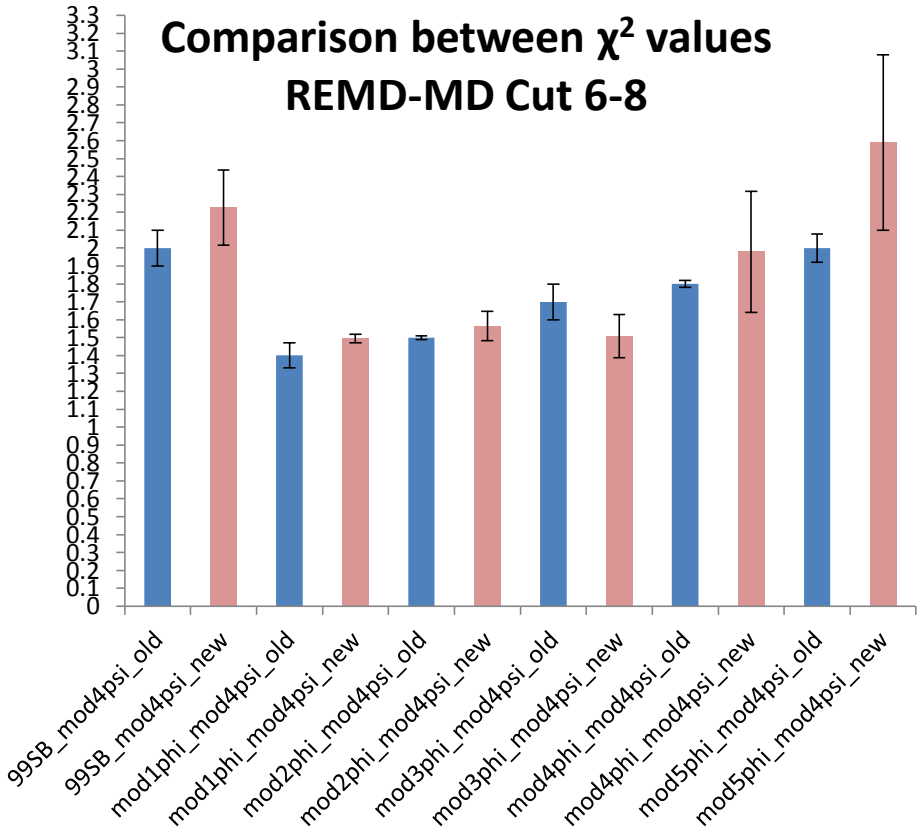


Figure 59 Comparison between χ^2 values for REMD and MD simulations of the modifications 99SB_mod4ψ through mod5φ_mod4ψ. The old data is the REMD data where the new data is the MD data.

Overall the MD data has good agreement with the REMD data for the χ^2 values, the highest discrepancies from the REMD data are found in mod5φ-mod2ψ, mod5φ-mod3ψ and mod5φ-mod4ψ. In the case of mod5φ-mod4ψ we extended the simulations and re-calculated the χ^2 values. The new value obtained was 3.36± 0.33 which is higher than the value obtained from the REMD simulation. We think that this is due to the sampling between the α_R and α_L regions that is taking long time to converge, because when we investigated which scalar coupling was causing the largest deviations we discovered that it was the ^3J (H_α, C') scalar coupling that is sensitive to this transition (see figure 54). As seen in figure 61-63, the simulation is sampling the α_L region often.

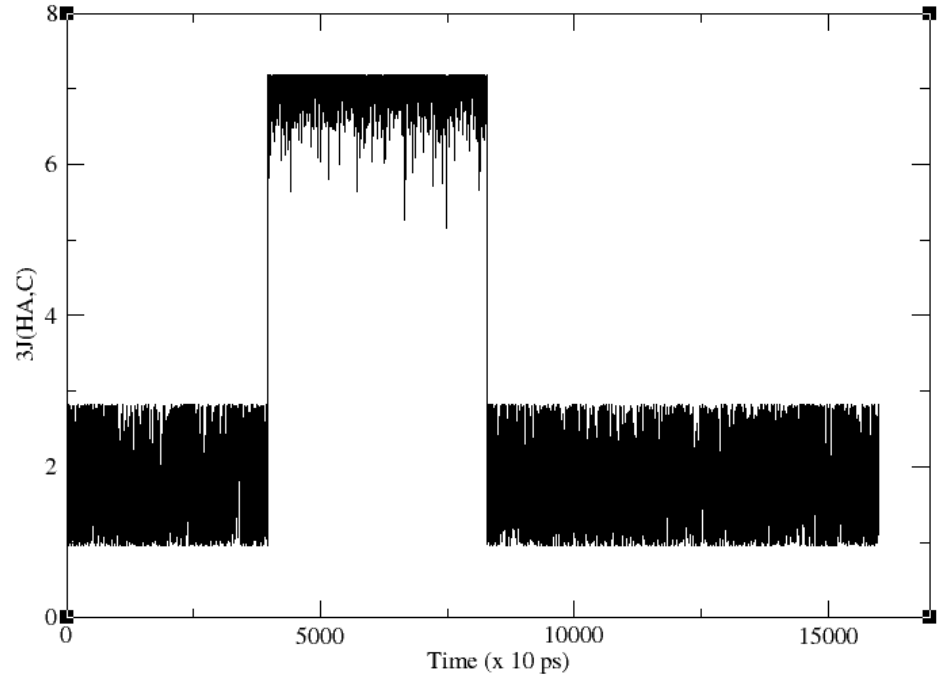


Figure 60 ^3J (H_α, C') scalar coupling values as a function of time for mod5φ-mod4ψ for the central residue of Ala_5. The data shown here is from the extended simulation.

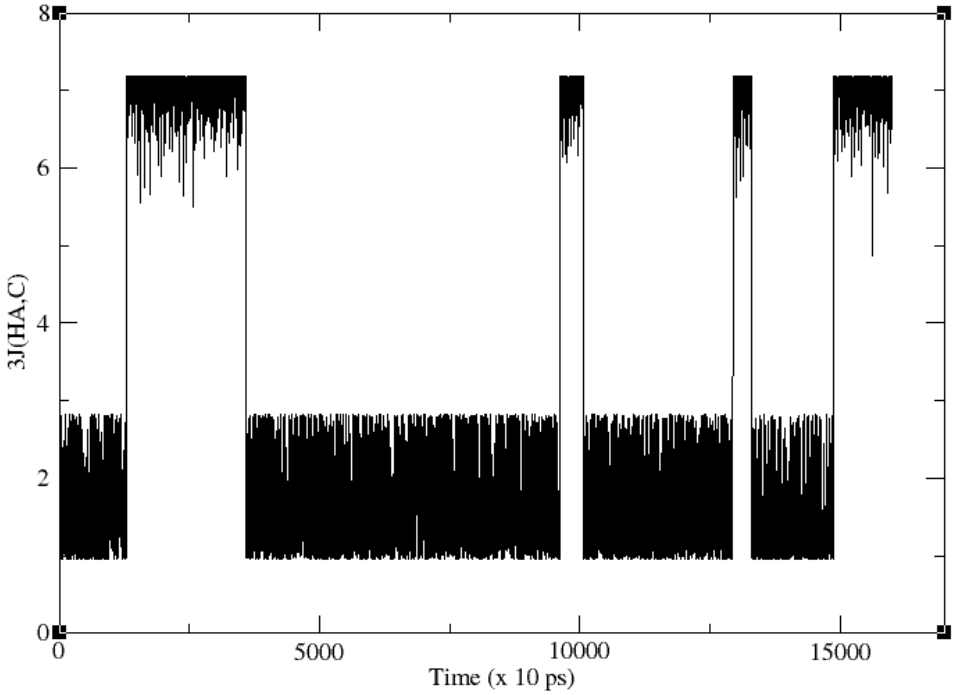


Figure 61 ^3J (H_α, C') scalar coupling values as a function of time for mod5φ-mod4ψ for the central residue of Ala_5. The data shown here is from the native simulation.

We also investigated the Ramachandran dihedral map and we corroborated that the α_L region is being considerably populated for the mod5φ-mod4ψ simulations.

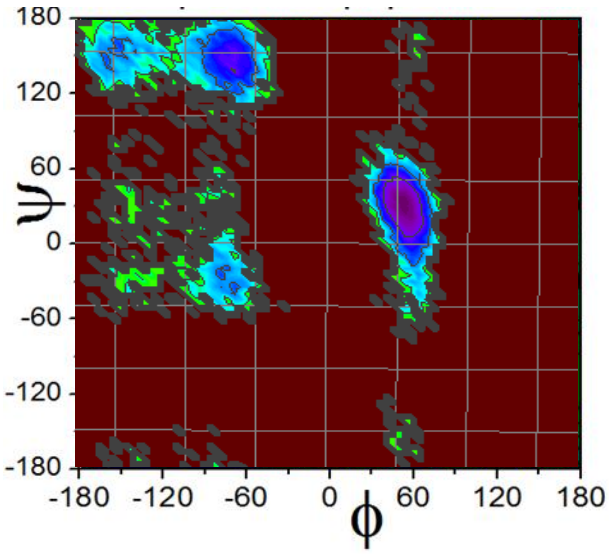


Figure 62 Ramachandran plot for the dihedral distributions of mod5φ-mod4ψ for the central residue of Ala_5. The data shown here is from the extended MD simulation.

As seen in figure 46, the mod5φ-modψ family of modifications appears to stabilize the α_R helical region in comparison to the other modifications. This effect was not so evident in REMD simulations because of the enhanced sampling of the method; however this is more evident in the MD simulations.

3.2 Dihedral distributions for central residue in Val_3

Similar to the results from the Ala_5 simulations, the results for the simulations show that AMBER ff99SB force field sample a broad distribution in the region -150° < φ < -60°, which gets more localized with mod1φ force field. The ppII region and the α_R region are more defined than with ff99SB (see figure 64). The sampling of the β region is considerably reduced; this is contradictory to what the experiments indicate for this peptide's secondary structure. The experiments indicate that this peptide has β extended conformation [195, 196]. The modifications were intended to reproduce the secondary structure propensities for all amino-acid residues; however the modifications did not show this effect on Val_3.

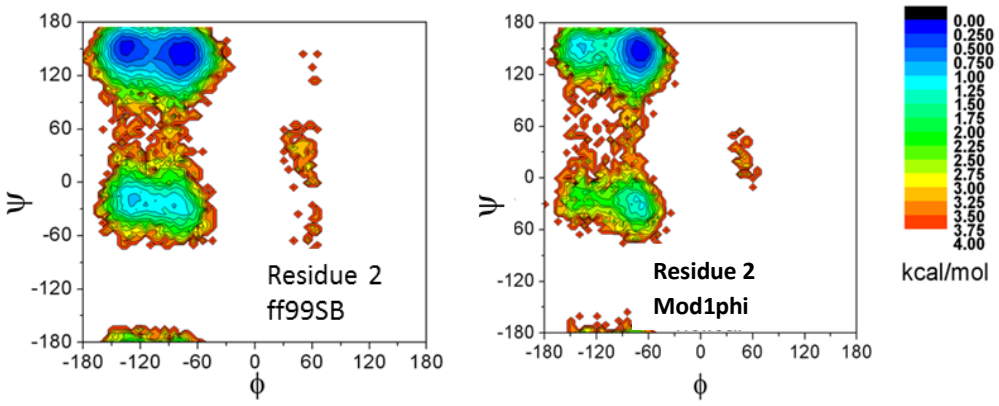


Figure 63 Ramachandran maps for the dihedral distributions of the central residue of Val_3. The left panel corresponds to the native simulation with ff99SB and the right panel corresponds to the native simulation with mod1φ

We further investigated this effect by plotting the dihedral propensities of alanine and valine from the dihedral library of Lovell et al. [149] to determine if both residues have similar dihedral propensities. As seen in figure 65 this is not the case, in fact alanine has a peculiar dihedral distribution in comparison to other amino-acids (figure 66).

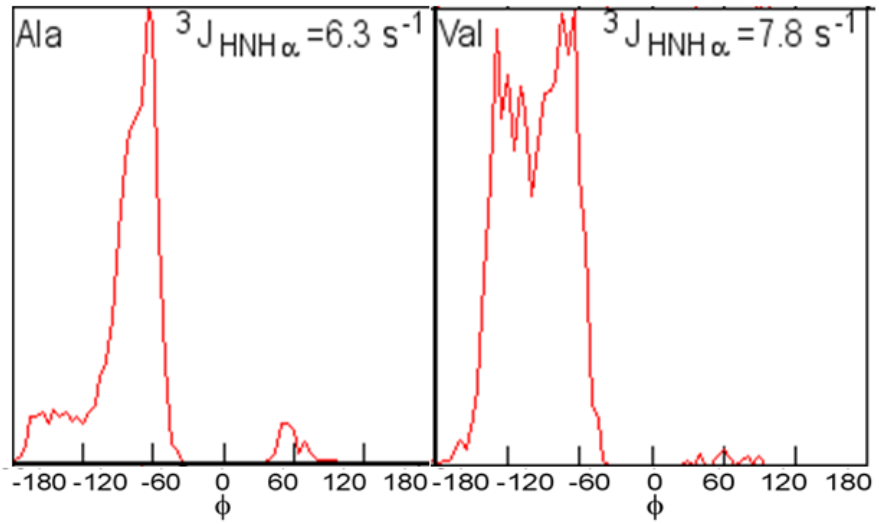


Figure 64 Dihedral propensities of alanine and valine from the dihedral library of Lovell et al. [149] for proteins from the Protein Databank (PDB). Most of the dihedral populations for alanine are located around the φ = -60° whereas the dihedral populations for valine have two major peaks around the φ = -150° and φ = -60° regions. The average value for ^3J (H_N , H_α) scalar coupling were also calculated from the dihedral angles using the original parameters for the Karplus function. Graph generated by James Maier.

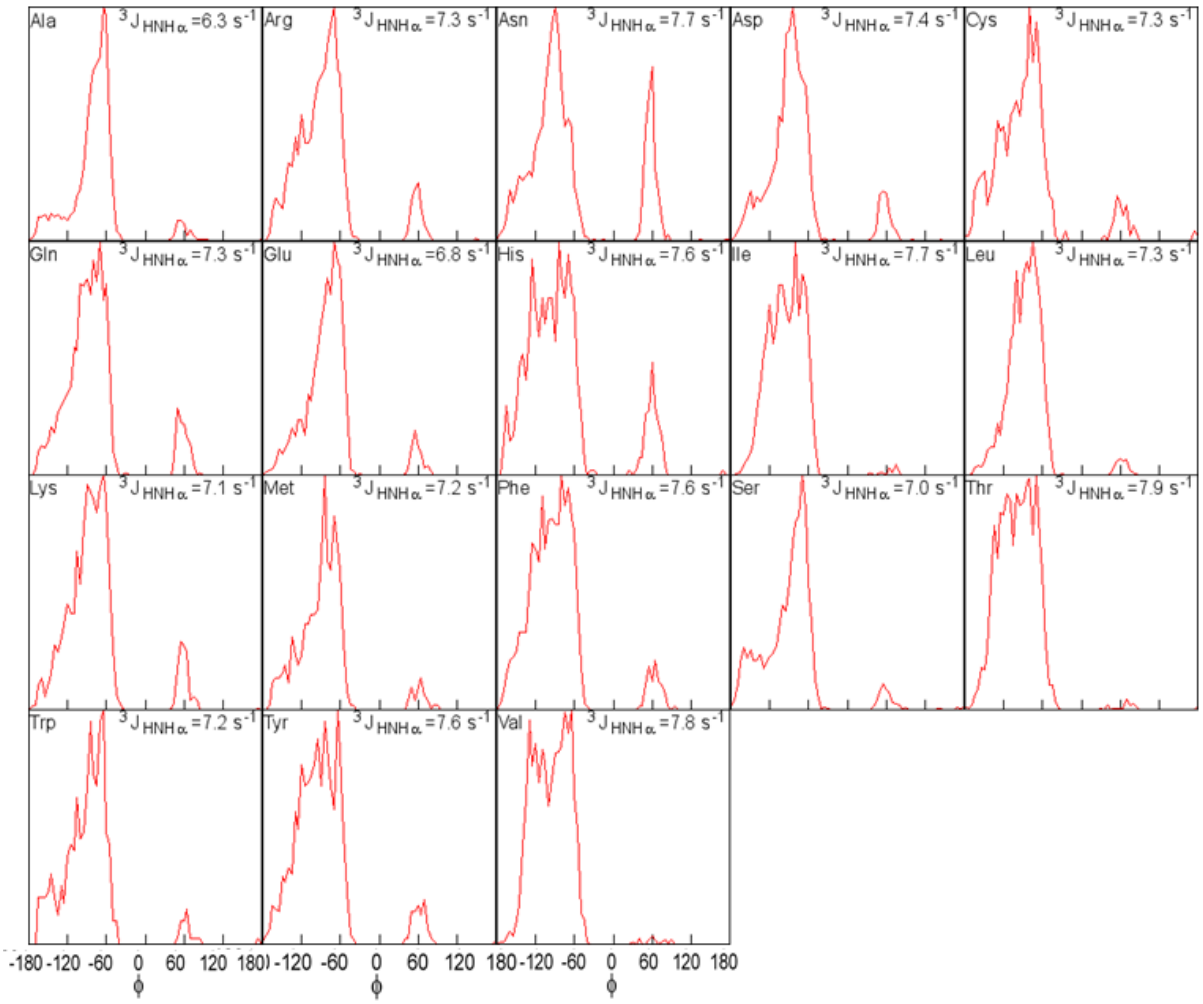


Figure 65 Dihedral propensities for all amino-acids with the exception of glycine and proline. The dihedral populations were taken from the dihedral library of Lovell et al. [149] for proteins from the Protein Databank (PDB). The dihedral distributions are different from alanine for most of the amino-acids with the exception of glutamic acid and serine. Higher branched amino-acids had the highest differences. Graph generated by James Maier.

The differences between the dihedral populations of alanine and the rest of the amino-acids except glycine and proline are due to the effects of side-chains on the backbone. As seen from figure 66 the higher branched amino-acids have the highest dihedral deviations for φ angle. Amino-acid residues such as aspartic acid, glutamic acid and serine resemble the propensities of alanine, whereas amino-acids such as asparagine, glutamine, tyrosine, tryptophan, histidine, phenylalanine, methionine, cysteine, threonine, isoleucine, leucine, lysine, arginine and valine are different from alanine. From the dihedral angle populations we also calculated the average ^3J (H_N, H_α) scalar coupling value for each of the amino-acid residues. As shown in figure 66 most of the values are higher than 7 Hz, only alanine, glutamic acid and serine have values lower or close to 7 Hz.

3.3 Comparison between implicit solvent and explicit solvent results

One of the advantages of using implicit solvent models is that the solvent molecules surrounding the solute are not explicitly accounted for. Instead, they are treated as a continuum which makes the calculations simpler when compared to explicit solvent models. Implicit solvent models are useful to speed up simulations at a fraction of computational cost in contrast to explicitly including solvent molecules [324, 325]. Initially we used the revised GB Neck model [193] for the simulations of Val_3. We used this solvent model because we wanted to investigate if this model could qualitatively and quantitatively reproduce the dihedral propensities and vicinal scalar coupling data obtained with TIP3P explicit solvent model for this small peptide. If the solvent model was able to reproduce these results, then this model solvent could be used for the calculations for the other model peptides required for testing. We began by plotting the dihedral populations from simulations as indicated in table 21. The information shown here corresponds to the data for mod1φ which was the force field that had the closest agreement with experiment for Ala_5.

Table 21 Val_3 secondary structure basin comparison between TIP3P and revised GB neck model [193] for ff99SB and mod1φ

Method	α_R	β	ppII	α_L
ff99SB (TIP3P)	16.80±0.60	29.50±2.00	50.40±2.60	0.19±0.00
ff99SB (rev. GB Neck)	6.91±1.31	40.17±2.22	48.95±3.73	0.01±0.09
Mod1φ (TIP3P)	12.70±4.10	13.90±0.80	70.00±2.90	0.20±0.20
Mod1φ (rev.GB Neck)	5.15±2.62	22.42±0.13	68.59±2.41	0.01±0.02
Exp	19±0.70	52±0.50	29±0.60	0±0.00

Table 21 Secondary structure comparison between TIP3P and revised GB neck model for ff99SB and mod1φ. The results indicate that the α_L populations of mod1φ are comparable for both solvent models while the results for ff99SB show that the revised GB neck model had less population than TIP3P. The β populations were less for TIP3P than revised GB neck model while the results for ppII and α_L populations were comparable between both solvent models. Error bars were taken from native and extended conformations.

The results for the dihedral populations between the solvent models indicate that the revised GB neck model only reproduces the dihedral propensities for the ppII basin for ff99SB and mod1φ. The β populations were higher for the implicit solvent model than the explicit one, while the α_R populations were not consistent between force fields; in the case of ff99SB the populations were less for revised GB neck model than TIP3P while for mod1φ the results were comparable. To further investigate this, we also compared the results for the vicinal scalar couplings following the same methodology described for Ala_5. The results shown here were calculated with the original Karplus parameters reported by Hu et al. [175]. As in the case of Ala_5 we began by calculating the ^3J (H_N, H_α) scalar couplings which are shown in table 22 for TIP3P explicit solvent model and table 23 for revised GB Neck model [193].

Table 22 $^3J(H_N, H_\alpha)$ Value Comparison Val_3

Exp 7.94	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	**8.04±0.02**	**7.94±0.00**	**8.00±0.01**	**7.96±0.02**	**8.02±0.00**
Mod1φ	6.70±0.01	6.55±0.01	6.58±0.02	6.67±0.00	6.66±0.00
Mod2φ	7.09±0.05	7.03±0.07	7.04±0.02	7.15±0.03	7.13±0.03
Mod3φ	7.20±0.02	7.02±0.02	7.07±0.01	7.11±0.06	7.07±0.03
Mod4φ	7.61±0.00	7.46±0.03	7.56±0.00	7.55±0.01	7.58±0.00
Mod5φ	7.38±0.04	7.28±0.03	7.28±0.00	7.32±0.01	7.41±0.01

Table 22 ^3J (H_N, H_α) coupling values for all 30 force fields tested. The force field that had the best agreement with experiments was mod3*ψ*, mod*ψ* family of force fields also had good agreement with experiments (in bold). Error bars were taken from the native and extended simulations.

Table 23 ^3J (H_N, H_α) values

Exp 7.94	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	**7.69±0.02**	**7.63±0.03**	**7.68±0.04**	**7.80±0.00**	**7.83±0.03**
Mod1φ	6.36±0.00	6.34±0.08	6.32±0.02	6.48±0.01	6.53±0.03
Mod2φ	6.87±0.00	6.84±0.01	6.74±0.06	6.88±0.08	6.96±0.01
Mod3φ	6.90±0.00	6.85±0.02	6.84±0.03	6.90±0.01	6.93±0.01
Mod4φ	7.32±0.03	7.32±0.03	7.25±0.02	7.39±0.04	7.50±0.04
Mod5φ	7.08±0.05	7.11±0.08	6.99±0.07	7.09±0.05	7.06±0.06

Table 23 ^3J (H_N, H_α) coupling values for all 30 force fields tested. Similarly to the TIP3P results, the force fields that had the closest agreement with experiments were mod3ψ and mod4ψ; the 99SB_modψ family of force fields also had good agreement with experiments (in bold). Error bars were taken from the native and extended simulations.

The TIP3P results indicate that mod1ψ force field in this case has the best agreement with experiments. Other modifications to the ψ parameters performed comparably well such as mod2ψ, mod3ψ and mod4ψ. Similarly, in the case of the revised GB Neck results, the trend was observed, the 99SB_modψ family for modifications had the closest agreement with experiments, although the actual scalar coupling values were lower than in the TIP3P case. This trend was observed for the rest of the scalar coupling values when comparing the χ^2 values between TIP3P and revised GB Neck model, although in the case of the implicit solvent model, the results are more comparable between the 99SB_modψ and the mod1φ families of force fields (see table 24 and 25). This is due to the discrepancies between the calculated scalar coupling values for the solvent models. Nevertheless given the uncertainties in the Karplus functions used for the calculations, the implicit solvent model might be able to reproduce the trends observed for explicit solvent calculations but it does not reproduce the actual values exactly. Therefore, it is recommended to use explicit solvent models like TIP3P for simulations that require this type of calculation.

Table 24 χ^2 values for Val_3 for TIP3P explicit solvent model

	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	1.72±0.03	1.46±0.03	1.62±0.01	1.74±0.13	1.64±0.00
Mod1φ	1.83±0.06	1.83±0.04	1.86±0.05	1.79±0.02	1.74±0.05
Mod2φ	**1.49±0.02**	**1.39±0.01**	**1.37±0.04**	**1.38±0.01**	**1.43±0.01**
Mod3φ	1.86±0.01	1.66±0.08	1.70±0.15	1.66±0.00	2.03±0.15
Mod4φ	1.65±0.02	1.49±0.00	1.54±0.09	1.49±0.02	1.70±0.09
Mod5φ	2.09±0.24	2.25±0.24	1.81±0.07	1.91±0.05	1.78±0.14

Table 24 χ^2 values for all 30 force fields tested. The force fields that had the best agreement with experiments were mod2φ and mod2φmodψ family of force fields.

Table 25 χ^2 values for Val_3 for revised implicit GB Neck model

	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	2.07±0.13	1.93±0.04	1.73±0.02	2.04±0.01	1.93±0.04
Mod1φ	2.84±0.24	2.74±0.09	2.79±0.19	2.60±0.02	2.24±0.06
Mod2φ	**2.09±0.00**	**2.27±0.18**	**2.16±0.21**	**1.97±0.08**	**2.14±0.18**
Mod3φ	2.23±0.24	2.46±0.03	2.14±0.23	2.18±0.00	2.41±0.06
Mod4φ	2.08±0.03	2.03±0.01	1.95±0.15	2.08±0.15	2.05±0.10
Mod5φ	2.40±0.09	1.93±0.03	2.10±0.01	2.14±0.06	2.01±0.06

Table 25 χ^2 values for all 30 force fields tested. The force fields that had the best agreement with experiments were mod2φ and mod2φmodψ family of force fields.

For the case of the χ^2 values the best performing force fields were the family of mod2φ-modψ force fields. Although the force fields that had the best χ^2 values for Ala_5 were not the best in this case, the results were close to each other within uncertainty of experimental values.

3.4 Differences in secondary structure propensities for alanine and valine

As previously discussed this small peptide is believed to have β strand partial character [195, 196]. The modifications proposed here were designed to increase the agreement with poly-alanine peptides. The changes applied to the dihedral energies of ff99SB force field were intended to increase the populations in the φ = -60º region which would lower the sampling of the $-150 \leq \varphi \leq -120°$ region. Therefore the scalar coupling values were lowered in order to have better agreement with the experimental values reported for Ala_5. When we compared the dihedral propensities for valine against alanine, we could see that they were different, in particular in the regions described above (see figure 65). It appears that in the case of alanine most of the dihedral populations are concentrated in the φ = -60º region, whereas in the case of valine the populations are more equally distributed between these two regions (φ = -60º and $-150 \leq \varphi \leq -120°$ regions). This effect appeared to be due to the side-chain group on the backbone propensities. In order to investigate this, we plotted the backbone dihedral populations from the Lovell et al. [149] amino-acid library color coded by χ_1 dihedral angles (see figure 67). The first graph from the left panel corresponds to the backbone conformations for valine where all the χ_1 angles (trans, gauche- and gauche+) were included. The second graph from the left panel corresponds to the backbone conformations were the gauche- χ_1 angles where included. Lovell et al. [149] defined gauche- to angles where χ_1 = -60º. In this graph we included the backbone conformations for φ/ψ where $-30º \geq \chi_1 \geq -90º$.

For the right panel the upper graph corresponds to the backbone conformations of φ/ψ where the trans χ_1 angles are included. The definition of Lovell et al. for trans is χ_1= 180º, therefore the backbone conformations for φ/ψ where $-150º \geq \chi_1 \geq -180º$ and $180º \geq \chi_1 \geq 150º$ were plotted in the graph. The lower graph corresponds to the backbone conformations of φ/ψ where χ_1 = gauche+. Lovell et al. defined the gauche+ conformation as χ_1 = +60º, therefore the conformations for φ/ψ where $90º \geq \chi_1 \geq 30º$ were plotted.

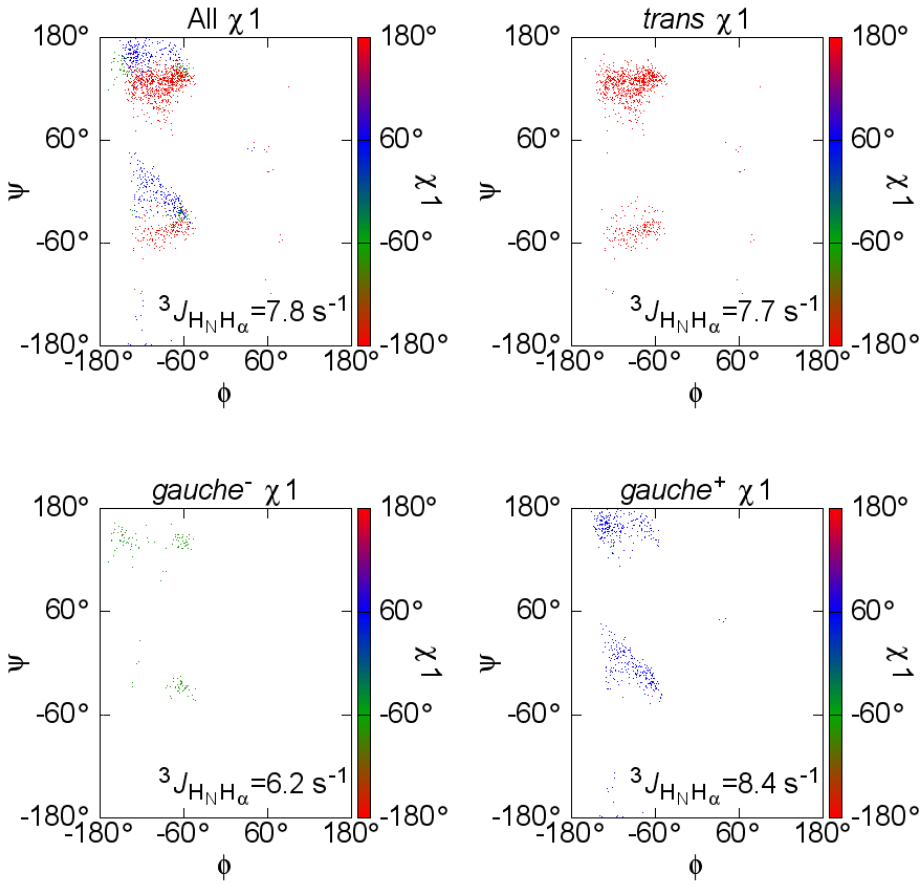


Figure 66 Ramachandran maps for valine taken from the Lovell et al. [149] dihedral maps. The upper left graph corresponds to *φ/ψ* conformations where all the χ_1 conformations were included (trans, gauche+ and gauche-). The lower left graph corresponds to the *φ/ψ* conformations where -30º ≥ χ_1 ≥ -90º. The upper right graph corresponds to *φ/ψ* conformations where -150º ≥ χ_1 ≥ -180º and 180º ≥ χ_1 ≥ 150º. The lower right graph corresponds to *φ/ψ* conformations where 90º ≥ χ_1 ≥ 30º. Graph generated by James Maier.

From the dihedral populations we also calculated the ^3J (H_N, H_α) values. From the graph we can see that values in the region of -30º ≥ χ_1 ≥ -90º make the ^3J (H_N, H_α) value around 6.2 Hz whereas values in the region 90º ≥ χ_1 ≥ 30º make the ^3J (H_N, H_α) value around 8.4 Hz. Similarly, values in the regions -150º ≥ χ_1 ≥ -180º and 180º ≥ χ_1 ≥ 150º give a scalar value around 7.7 Hz. This information corroborates the notion that the backbone conformations are dependent upon side-chain preferences, especially for higher branched amino-acids as illustrated in figure 66. This suggests that it is necessary to investigate the side-chain dihedral propensities of ff99SB to determine if this force field is biasing a specific configuration such as the gauche- χ_1 conformation where the angles are restricted to the -30º ≥ χ_1 ≥ -90 º regions. This effect was not

further investigated, because the work presented here was aimed at evaluating the effects of modifications to the backbone parameters. The side-chain parameters efforts will be covered in a separate report (Maier et al. [326]).

3.5 Hydrogen Bond Surrogate (HBS) peptide helical propensities

After investigating the agreement between experiments and simulations for Ala_5 and Val_3, we decided to investigate how the modifications that had the best agreement for Ala_5 were performing against ff99SB for helical propensities, which were mod1φ, mod2φ, mod1φmod1ψ, mod1φmod2ψ, mod1φmod3ψ and mod1φmod4ψ. We tested mod5φ since the agreement between experimental and simulated scalar couplings of ^3J (H_N, H_α) was in between the values of ff99SB and mod1φ. The metric for helical propensities shown here was Dictionary of Secondary Structure for Proteins (DSSP) analysis as described in section 2.1.10.3 from simulation methods. Figure 68 shows how the modifications increase the helical propensities in comparison to ff99SB.

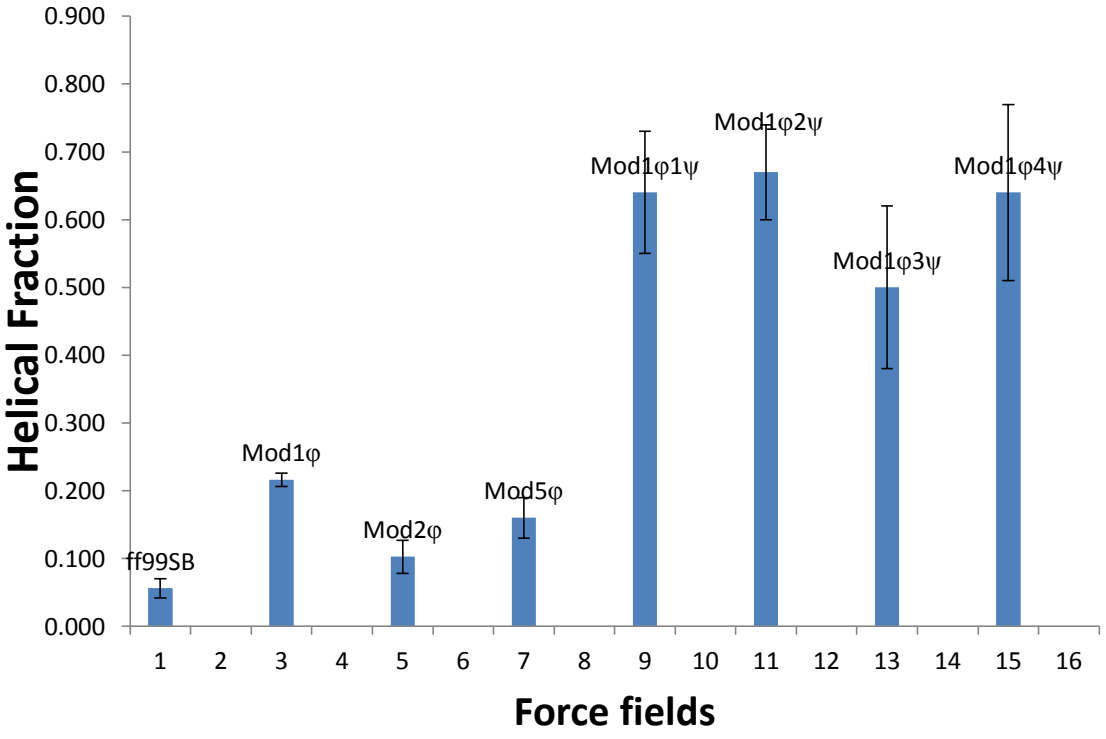


Figure 67 Comparison of helical propensities for the best performing force fields. Error bars were taken from native and semi-extended simulations as described in section 2.7.

The performance of some force fields such as mod1φ was comparable to experimental results that indicate this peptide is 46% folded at room temperature. However, in the case of this peptide simulations and experiments cannot be compared directly because this folded value was taken from CD spectra as the ratio $[\theta]_{222}/[\theta]_{max}$, where $[\theta]_{max}$ = -23,400. The concentration of each peptide was determined by monitoring the absorbance of the tyrosine residue under denaturing conditions. Nevertheless this can serve as an indicator of the boundaries of the fraction folded region, where values close to this can be taken as a good resemblance to what is seen in experiment. Having said this, the force field that resembles this value the most is mod1φ followed by mod5φ and mod2φ. The other force fields such as mod1φmod1ψ, mod1φmod2ψ, mod1φmod3ψ and mod1φmod4ψ had greater helical propensity in comparison to experiment, while mod1φmod3ψ (~40%) was the closest one to experiments. However, the upper limit of the mod1φmod3ψ data had considerably higher helical propensities (~0.6). Mod1φ (~0.25), mod2φ (~0.12) and mod5φ (~0.20) were further evaluated against ^3J (H_N, H_α) scalar couplings. The coefficients used for the Karplus curve were taken from Pardi [177] et al. since those were the parameters that Wang et al. [168] used to calculate the φ dihedral angles from the ^3J (H_N, H_α) scalar couplings that they measured (figures 69-70). We also used the DFT1, DFT2 and Original parameters as previously done for the Ala_5, Val_3 calculations (see figure 71-73).

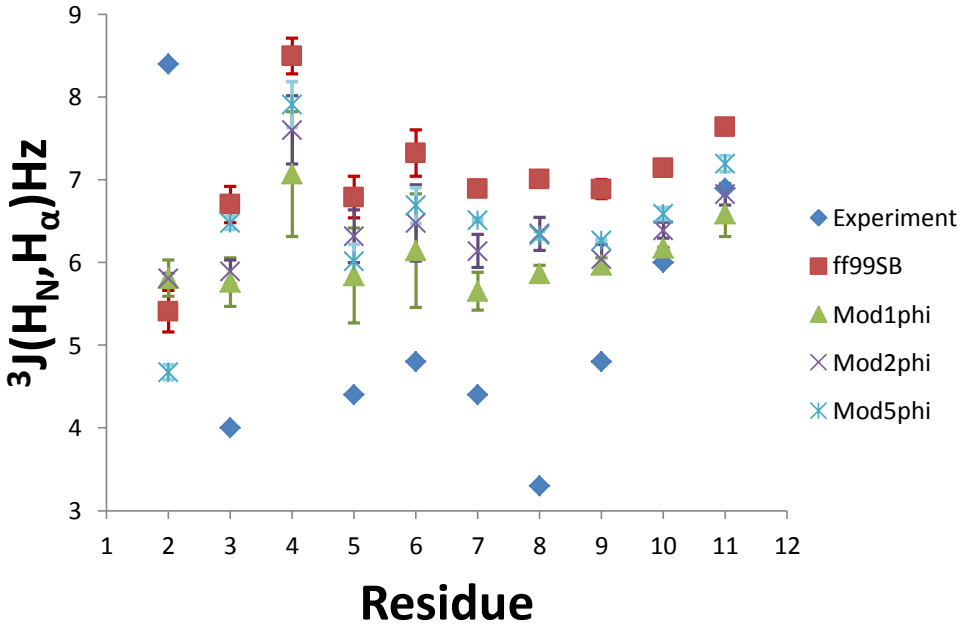


Figure 68 Comparison between ff99SB, modifications and experiment for $^3J(H_N,H_\alpha)$ scalar couplings of HBS peptide. Error bars were taken from native and semi-extended simulation results. The force fields shown are mod1φ, mod2φ and mod5φ. The experimental values were taken from Wang et al. [168]

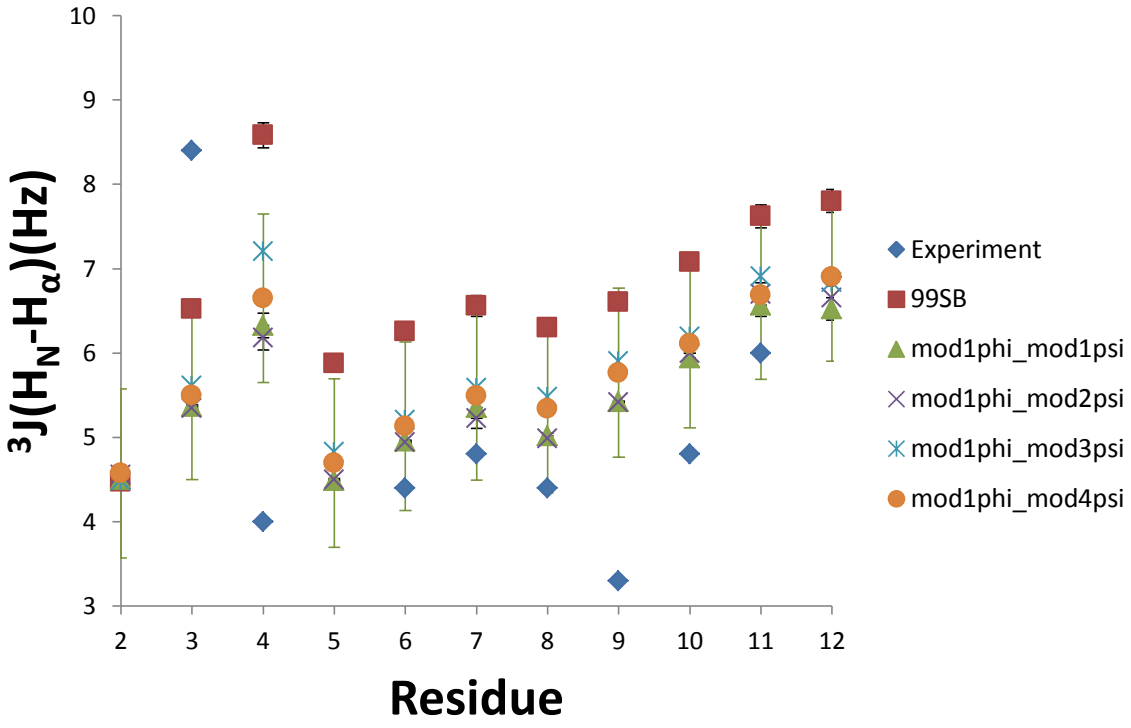


Figure 69 Comparison between ff99SB, modifications and experiment for the ^{3}J (H_N, H_α) scalar couplings of HBS peptide. The force fields shown here are mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ. The experimental values were taken from Wang et al. [168]. Error bars were taken from native and semi-extended simulations as described in section 2.7

Although the deviations shown here overlap between the different force fields, the trends show that the results from mod1φ, mod2φ and the mod1φ-modψ family of force fields have the closest agreement with experiment. Although there are some discrepancies due to the methods used for the simulation. For instance, from experiments the first turn of the helix was stabilized by a covalent bond between 1C=1C´--5C-5N covalent bond between residues 1 (cap) and 5 (alanine) at the N terminus. In the case of the simulations, we restrained the length of the bond between the cap and Alanine 5 (1O-5H) to 1.522Å with a weight of 317 kcal $mol^{-1}Å^{-2}$. The 1C=1O-5H angle was restrained to 120° with a weight of 80 kcal mol^{-1} rad^{-2} and the 1O-5H-5N angle was restrained to 110.1° with a weight of 63 kcal/mol/rad^2. Therefore, it is expected that the results of the first five residues will disagree from experiments because the simulation and experimental conditions are not the same. Furthermore, this effect of the first five residues will also influence other parts of the helix as in the case of Alanine9 which appear to have lower scalar values due to the geometry of the Alanine5 residue that is hydrogen bonding to. In the

case of the simulations this effect is not being reproduced because the geometry of the hydrogen bonding pattern of the helix is not the same given the differences between the restrained hydrogen bond portrayed in the simulations and the covalent C-C bond designed for the experiment. Nevertheless, the results show that the modifications increase the agreement with experiments for the other residues shown in figure 69 and 70. This trend was also seen when calculating the scalar couplings using the DFT1, DFT2 and original Karplus parameters previously used in the case of Ala_5 and Val_3 seen in figures 71-73 despite the differences in the parameters.

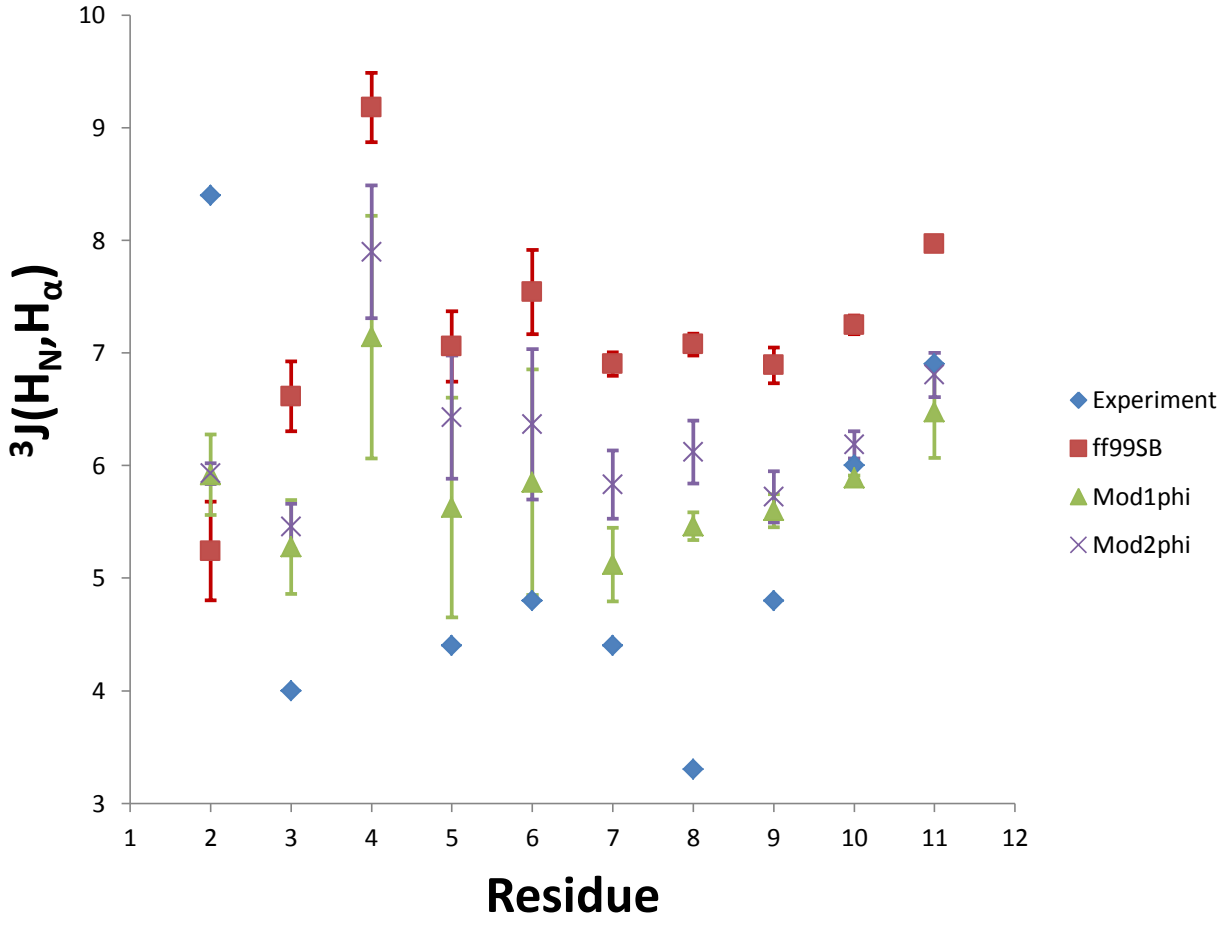


Figure 70 Comparison between scalar coupling values for DFT1 parameters on Hydrogen Bond Surrogate (HBS) peptide. Error bars taken from native and semi-extended simulations as described in section 2.7

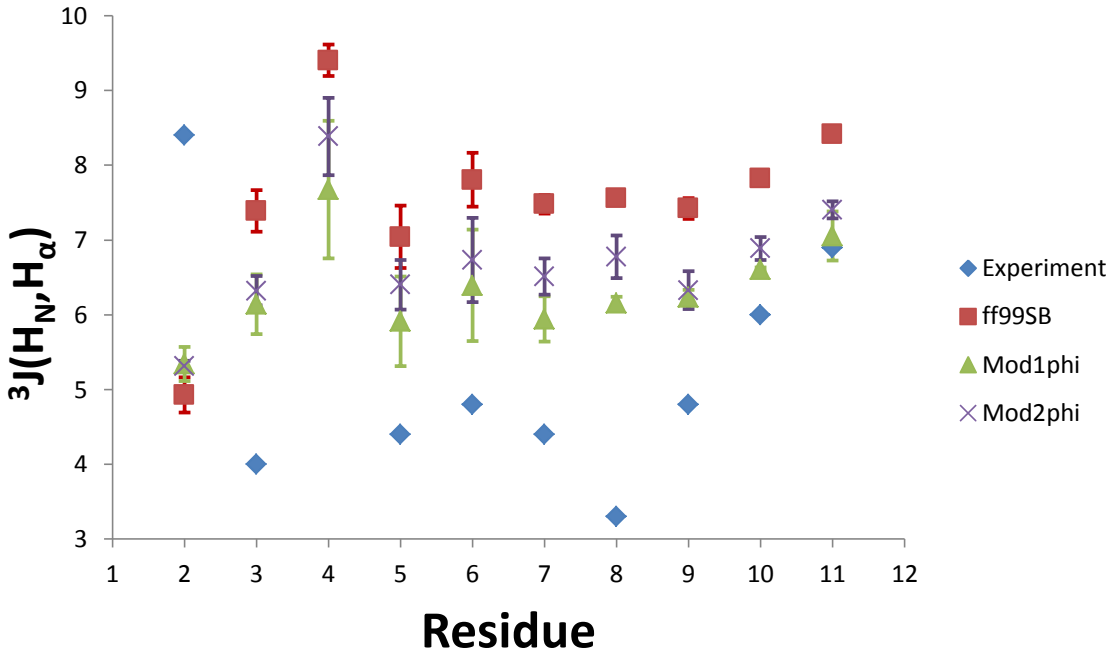


Figure 71 Comparison between scalar coupling values for DFT2 parameters of HBS peptide. Error bars taken from native and semi-extended simulations as described in section 2.7

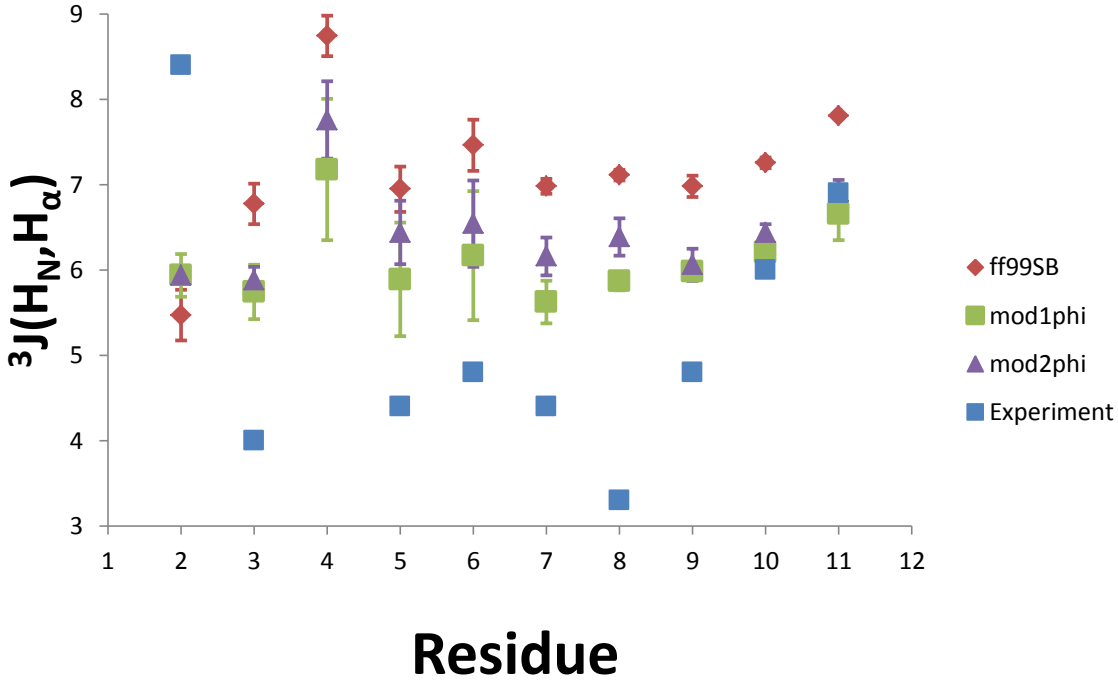


Figure 72 Comparison between scalar coupling values for original parameters of HBS peptide. Error bars taken from native and semi-extended simulations as described in section 2.7

3.6 Comparison between explicit TIP3P solvent simulations and revised GB Neck simulations

We wanted to investigate how the implicit solvent model was reproducing the trends observed with explicit TIP3P solvent. We used the revised GB Neck model system as described by Nguyen et al [193] and measured the helical propensities as described in section 2.1.10.3 from simulation methods. The results are shown in table 26. The trends show the same pattern for mod1φ (~0.69) mod2φ (~0.59), mod1φmod1ψ (~0.86) mod1φmod2ψ (~0.81), mod1φmod3ψ (~0.68) and mod1φmod4ψ (~0.76) as seen from the TIP3P simulations. Although the results are not exactly the same, as in the case of Val_3 we could see that the fraction helical content increases in a similar manner as in the results for TIP3P explicit solvent. Primarily, these simulations were performed to roughly estimate the results for the TIP3P explicit solvent, in other words, it served as initial testing in the peptide in order to see if the trends observed from Ala_5 and Val_3 would be observed in the case of this peptide. After this testing was done, the TIP3P explicit solvent simulations were performed.

Table 26 Helical fraction content for Hydrogen Bond Surrogate (HBS) peptide simulations with revised GB-Neck model

Exp=0.46	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	0.40±0.05	0.70±0.06	0.56±0.05	0.46±0.04	0.50±0.08
Mod1φ	0.69±0.07	0.86±0.1	0.81±0.07	0.68±0.07	0.76±0.08
Mod2φ	0.59±0.03				

Table 26 Fraction helical content for Hydrogen Bond Surrogate (HBS) peptide simulations with revised GB-Neck model. The highlighted values in bold correspond to the best performing force fields for Ala_5 and Val_3. Initially this testing was done to determine if the force fields that had the best performance for the other peptides would yield similar results in this peptide. The error bars were taken from native and semi-extended simulations as described in section 2.8

3.7 K19 peptide helical propensities

In order to further investigate the helical propensities of the force fields we study another helical peptide known as K19. This peptide was chosen in order to evaluate the helical propensities in a larger system for which experimental information is available (CD and NMR), without the need to impose distance and angle restraints as in the case of the Hydrogen Bond Surrogate Peptide (HBS). This peptide has been studied previously with implicit solvent simulations [167]. However in the work described here, the simulations were done with explicit TIP3P solvent model. As seen in figure 74 the helical fraction increased significantly with the modifications previously studied in comparison to ff99SB (0.05). Force fields like mod1φ (0.33), mod2φ (0.16), mod1φmod1ψ (0.80), mod1φmod2ψ (0.68), mod1φmod3ψ (0.76) and mod1φmod4ψ (0.75) had more helical fraction content than ff99SB.

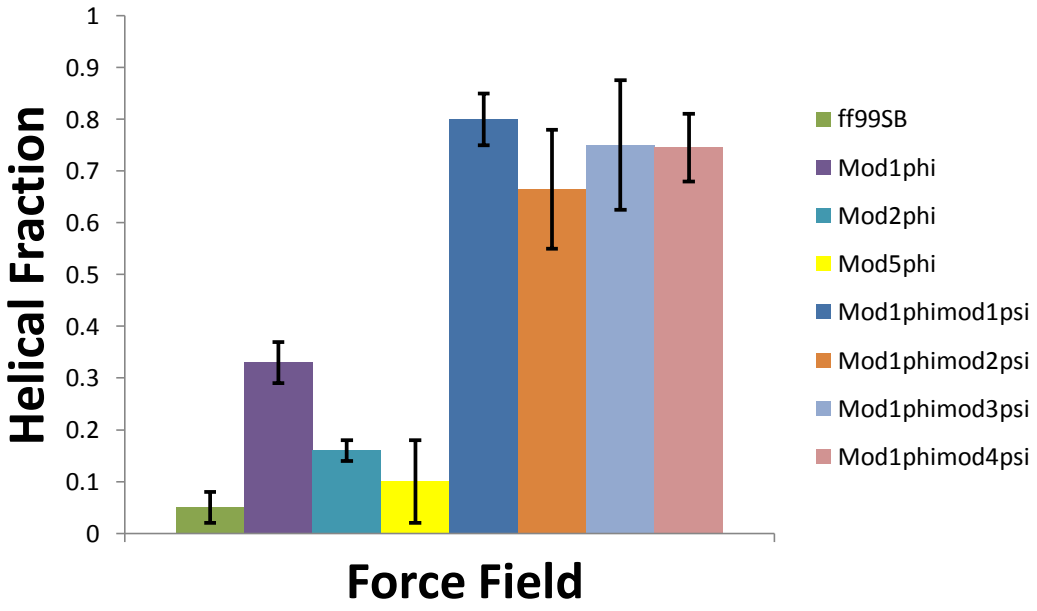


Figure 73 Helical fraction statistics for best force fields compared against ff99SB. Helical content for most of the force fields increased with the exception of mod5φ where the helical fraction was comparable to ff99SB.

Experimentally, the results from the CSD analysis indicate that this peptide is

approximately 48% folded at room temperature. Even though this is the temperature used for the simulations presented here, we cannot directly compare the experimental results to our simulation results because they are two different metrics. Currently, it is possible to use chemical shift libraries for prediction measurements from simulations; the assignments for the peaks are usually done with software packages that are based on empirical assignment from these libraries. Furthermore, these software packages are trained to predict chemical shifts which are taken from NMR ensemble averages; therefore the method can become less sensitive to conformational variations that contribute to the experimentally measured average. In other words, the software package can prefer the static experimental conformation and incorrectly predict chemical shifts from conformational dynamics from simulations [262].

However, as in the case of HBS peptide, they can provide an estimate of how folded is the peptide at this temperature. The results shown in figure 74 are taken from approximately 200 ns of MD simulation, where the actual plotted values were calculated by taken the average between the native and extended simulations at the time where the curves meet as shown in figure 75. The error bars were taken from the native and extended simulations. It is important to note that these simulations are not considered converged; rather they are used to evaluate the folding behavior of the peptide for this given period of time (200 ns). The simulation behavior, as indicated in figure 75 shows that the simulation with ff99SB does not seem to attempt to re-fold after the first unfolding event, whereas the simulation of mod1φ unfolds and attempts to re-fold. The simulations for the other force fields such as mod2φ and mod5φ unfold as well and attempt to re-fold, although the helical content is less than in the case of mod1φ. Similarly to the results seen for HBS peptide, the helical propensities of mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ were significantly higher than mod1φ, mod2φ and mod5φ. As previously described in section 2.1 these force fields had modifications to the φ' and ψ /ψ' parameters designed to improve the agreement with alanine vicinal scalar couplings and to increase the helical propensities of helical peptides. The helical effect can be seen here more than in the other tests previously described. Although according to the experimental information reported for Hydrogen Bond Surrogate and K19 these propensities seem higher. To further investigate the properties of these force fields (mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ

and mod1φ-mod4ψ) we evaluated their performance against NMR order parameters for Lysozyme in order as it is usually done when validating new force field parameters [90, 142, 266]

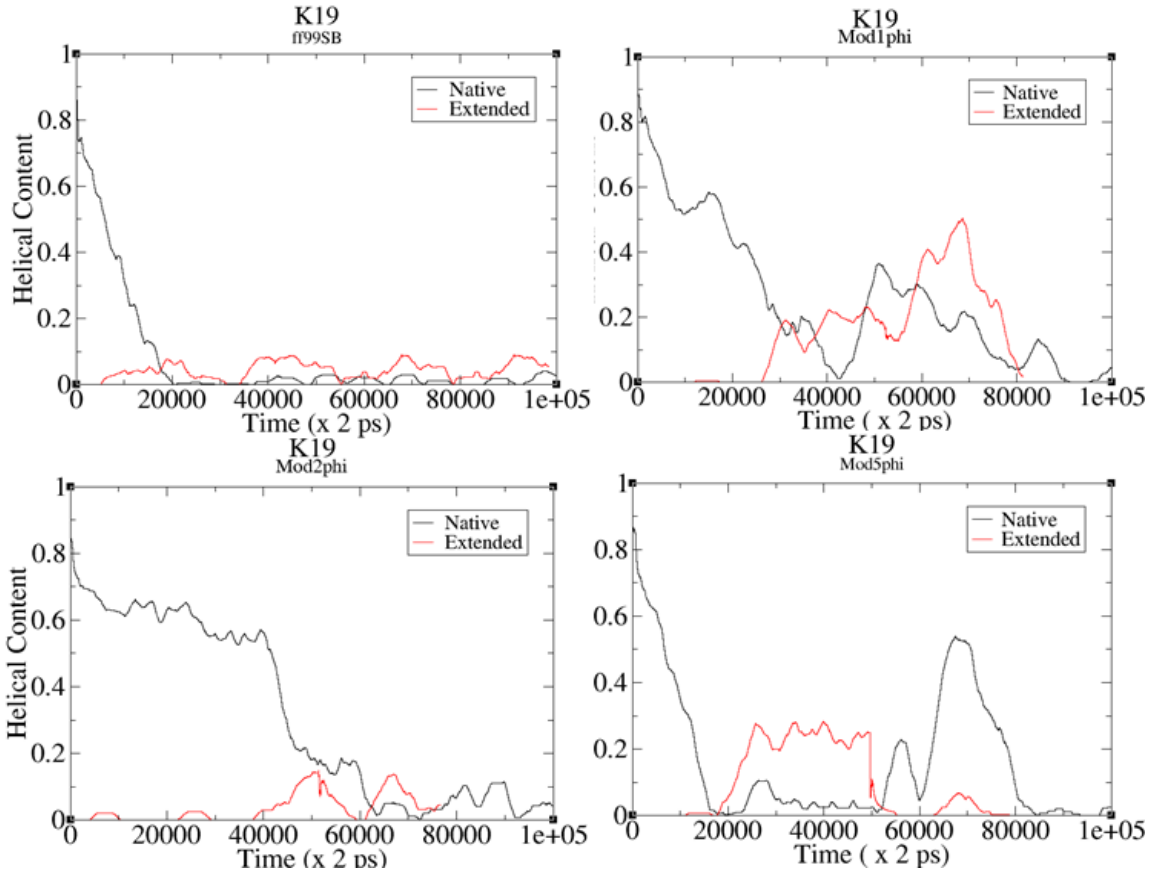
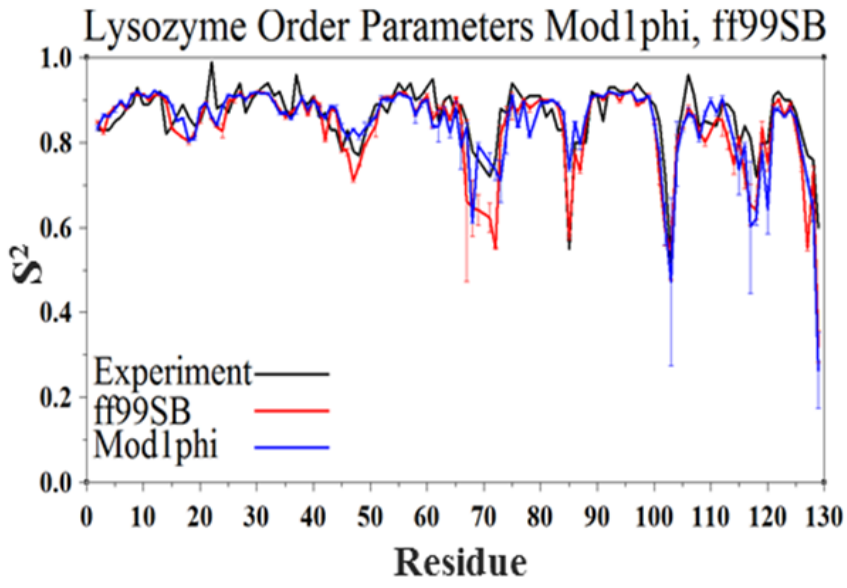
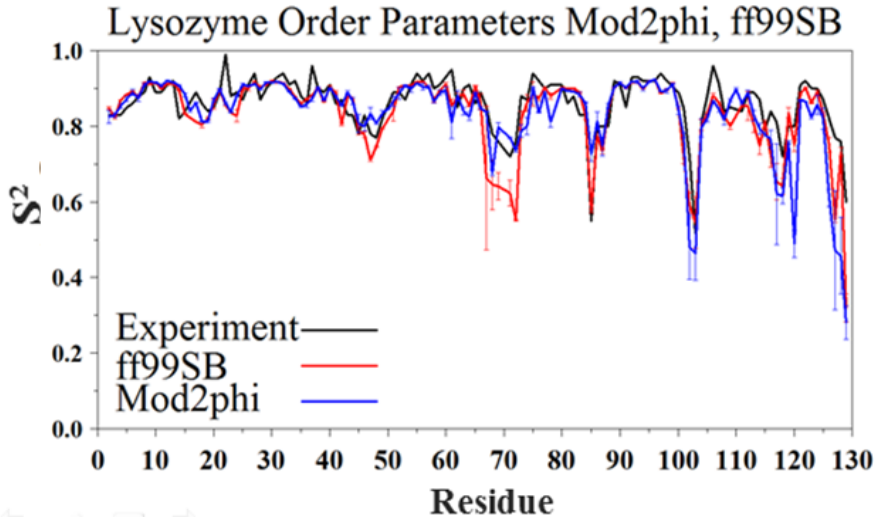


Figure 74 Helical propensity averages as a function of time for K19

3.8 Calculated NMR order parameters for Lysozyme

The modifications that had some of the best results from the testing performed were mod1φ, mod2φ, mod5φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ. We tested these force fields and compared the results to ff99SB as indicated in figures 76 and 77. Mod1φ had similar results to ff99SB, mod2φ and mod5φ when compared to experimental NMR order parameters [4]. Although in certain regions mod1φ had better agreement with experiments than ff99SB, this was seen particularly for residues 65-73. For certain residues such as Asn 103 mod1φ, mod2φ, mod5φ, mod1φ-mod1ψ and mod1φ-mod2ψ had discrepancies with experiment. This residue has been reported by Clore et al. to have fast and slow internal correlation times[4]. The authors mentioned that the method appears to reproduce the data within the experimental error, however for some residues in both proteins the internal re-orientational correlation function that is probed by NMR is not exponential with slow correlation times that are not in the extreme narrow limit as described by Lipari-Szabo [206]. This effect was particularly seen when comparing the ^{1}H-^{15}N NOE data from experiment to the fitted data using equation 6. The fitted values were significantly small or negative whereas the experimental values were positive. The issue is due to fast and slow correlation times for the relaxation motion of these bond vectors, when these differ by more than one order of magnitude the function does not decay exponentially. Instead, the function decays to a plateau intermediate before reaching the S^2 value. It is possible that the deviation between the runs for this residue's order parameter is related to this effect.


Lysozyme Order Parameters Mod1phi, ff99SB
1.0
0.8
0.6
0.4
0.2
0.0
S2
Experiment
ff99SB
Mod1phi
0 10 20 30 40 50 60 70 80 90 100 110 120 130
Residue


Lysozyme Order Parameters Mod2phi, ff99SB
1.0
0.8
0.6
0.4
0.2
0.0
S2
Experiment
ff99SB
Mod2phi
0 10 20 30 40 50 60 70 80 90 100 110 120 130
Residue

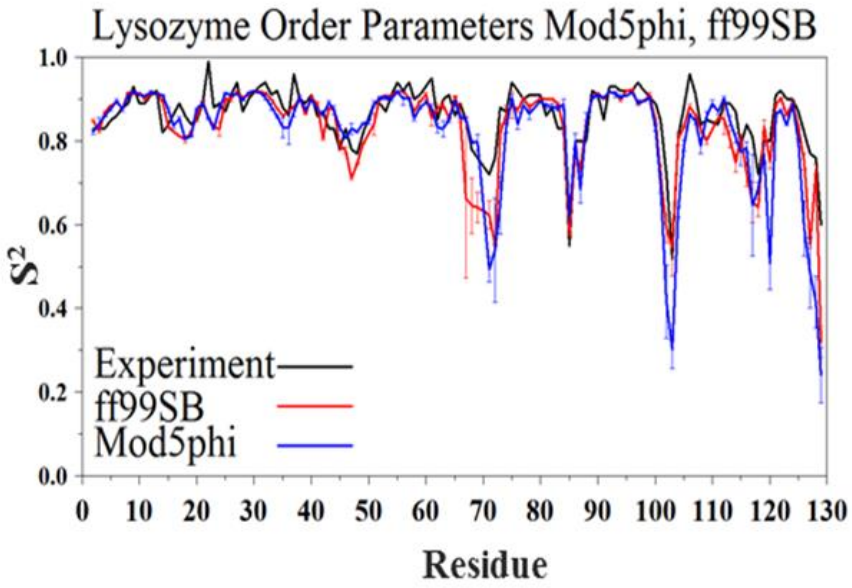


Figure 75 Lysozyme NMR order parameters for mod1φ, mod2φ and mod5φ. Error bars were taken from two independent runs with different random seeds.

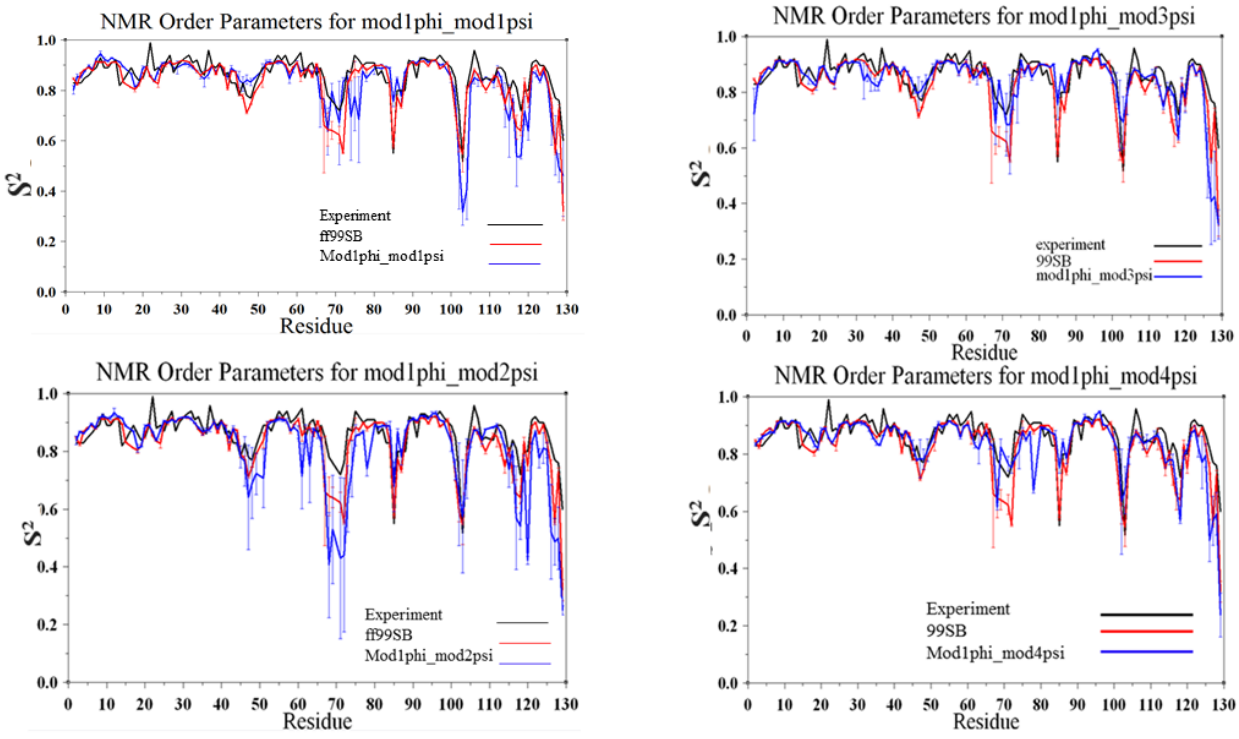


Figure 76 Lysozyme order parameters for mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ. Error bars are taken from two independent runs with different random seeds

The results for mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ showed increased flexibility in regions where ff99SB had good agreement with experiments, whereas mod1φ, mod2φ and mod5φ had better or comparable results to ff99SB. Since the results for the helical content of HBS and K19 peptide indicated that these force fields had elevated helical propensities that were atypical from experiments, we decided that these force fields would introduce undesired effects on the results from simulations that would go against the initial goal of having modifications that would have good agreement with transient helical peptides as well as reproducing backbone dynamics reported from NMR order parameters. Therefore, these force fields were not further evaluated against other test systems like ubiquitin and hairpins.

3.9 Calculated NMR order parameters for Ubiquitin

Mod1φ, mod2φ and mod5φ were evaluated against experimental values. The results indicated that there is no significant difference between them, with the exception of mod2φ which appears to be slightly more flexible for residues 52 and 53 of the sequence as seen in figures 78 and 79.

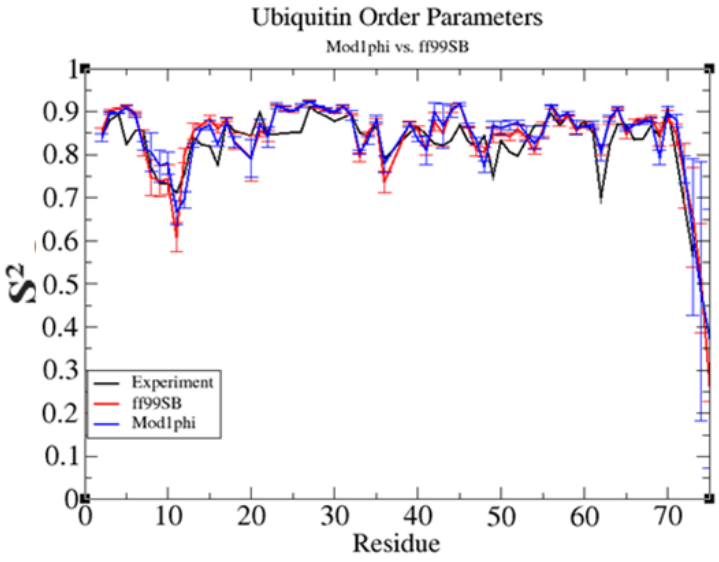


Figure 77 Ubiquitin NMR order parameters for mod1φ. Error bars were taken from two independent runs with different random seeds

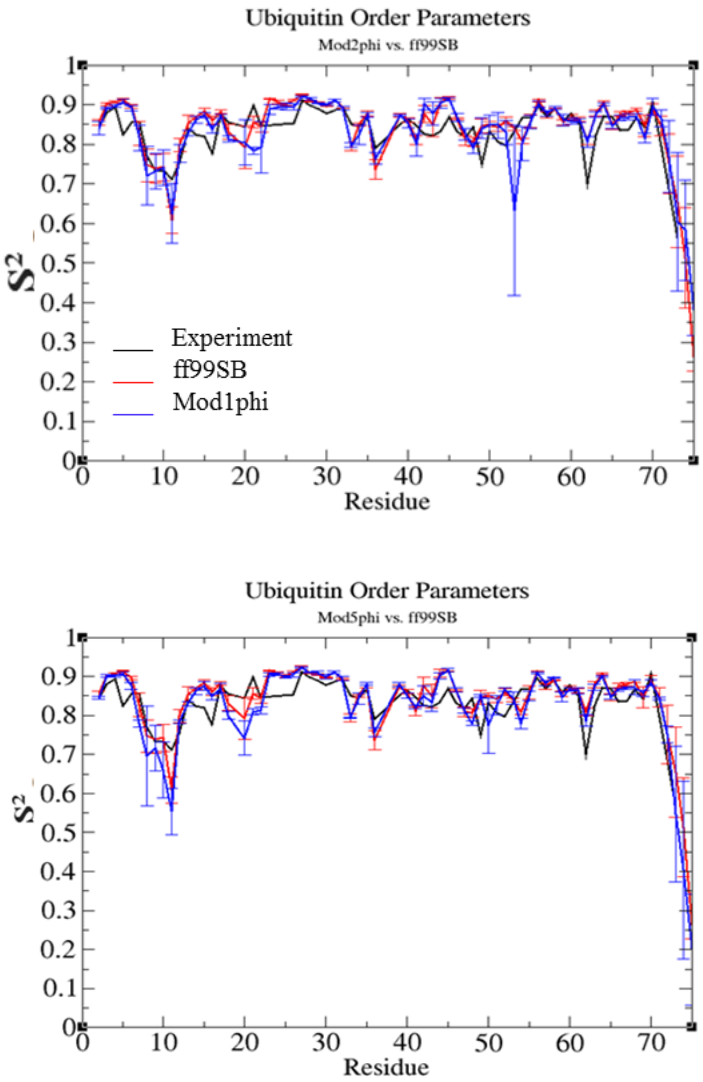


Figure 78 Ubiquitin order parameters for mod2φ and mod5φ. Error bars were taken from two independent runs with different random seeds.

NMR order parameters for ubiquitin have become a standard for the quality of force fields. Just like in the case of lysozyme, several groups have used these two systems for this purpose [90, 142, 154] . Nerenberg et al. [154] used this protein to validate their modification to the φ' potential and they observed root-mean-square errors of 0.044 between the experimental and simulated S^2 values for their modification and ff99SB. Hornak et al. [142] also used it to validate the parameters of ff99SB against ff94 and ff03. They reported that overall ff03 had the best agreement with experiments closely followed by ff99SB. Apparently the deviations are due to the magnitude of the fluctuations on the average conformations. Li et al [90] validated their modification to the backbone parameters using ubiquitin, lysozyme and interleukin-4 obtaining

results comparable to what is seen here. Overall the results shown here indicate that the modifications perform comparably when calculating NMR order parameters, although mod2φ showed increased flexibility for residues 52 and 53 (Aspartic acid and glycine) which are located near turn T4 region.

3.10 Evaluating β secondary structure propensities for modifications

3.10.1 CLN025

The results from simulations of small homologous and helical peptides as well as proteins indicated the best performing force fields; therefore we decided to test these force fields in β hairpins to determine how the modifications to the dihedral potentials were affecting the equilibrium between the β and ppII regions. Initially we tested mod1φ since it was the modification that had given the most promising results for all the tests performed here with the exception of Val_3 peptide. The reason, as previously stated is believed to be related to the influence of the side-chain dynamics on the backbone. As illustrated in figure 67 and described in section 3.4 the conformations of the χ_1 angle influence the dihedral propensities of φ/ψ angles. For the modifications described here we have only introduced changes to the φ', ψ/ψ' backbone terms leaving the side-chain parameters untouched. Nevertheless, figure 67 indicates that the side-chain parameters for χ_1 of valine prefer the gauche- conformation (where $-30^o \geq \chi_1 \geq -90^o$) therefore the backbone angles are sampling regions in φ dihedral space that are lowering the ^3J (H_N, H_α) vicinal scalar coupling and perhaps other scalar coupling values causing the χ^2 values to be larger for mod1φ. As shown in figure 80 we simulated CLN025, a 10 residue hairpin with faster folding properties [272] than its ancestor chignolin [185]. This peptide was chosen as a good model system because of its size making it suitable for implicit solvent simulations [193] using Replica Exchange Molecular Dynamics (REMD) [249, 250]. The reason why it was necessary to use REMD simulations was because we intended to study the folded populations for different temperatures. As seen in figure 80 we plotted the fraction folded populations for ff99SB (red) and mod1φ (black). The fraction folded populations were generated by calculating the backbone RMSD for N, C_α and C atoms and defining as folded the conformations that were less or equal to 2.0Å. We chose this cutoff based on the analysis described in section 2.3.5.3.1. Then we counted those structures and divided them by the total number of structures taken from the specific temperature replica trajectory in order to obtain the populations.

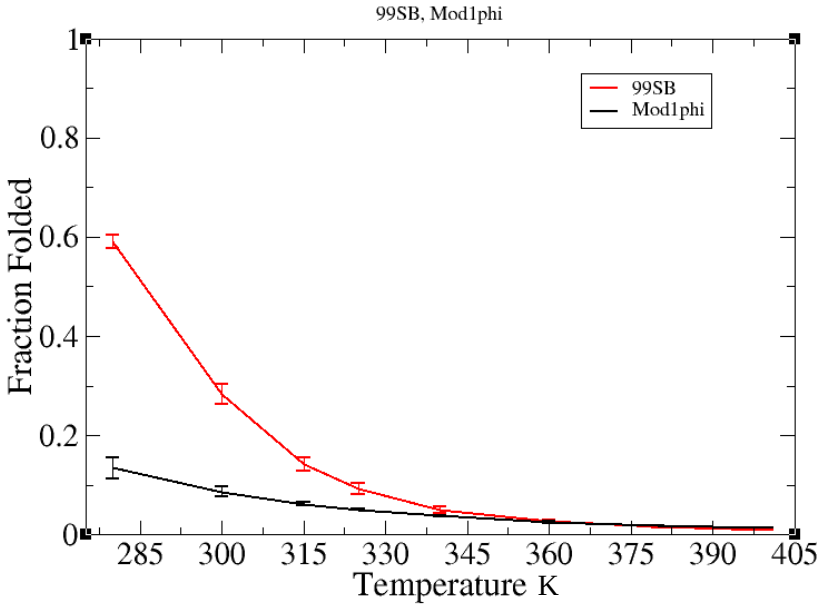


Figure 79 Melting curve for CLN025 from simulations with revised GB Neck model [193]. Error bars were taken from native and extended simulations for both force fields.

Mod1φ had low fraction folded (~ 0.15) when compared to ff99SB (~0.6). These results were concerning because it appeared as if this modification which had good agreement with the experimental information previously compared did not perform comparably to ff99SB. These results were taken from a simulation where the solvent accessible surface area (SASA) term was omitted from the calculations because the method has limitations. As previously reported by Chen et al. [327] the non-polar model was inclined to bias non-polar interactions that shifted conformations to non-native states. Others such as Levy et al. [328] indicated that solvent-accessible surface area models yield results based on the protein models used for their parameterization. For instance, solute-solvent energies of native protein conformations do not reproduce the solute-solvent energies of extended conformations as well as failing to reproduce this type of energies for mis-folded proteins. Wagoner et al. [329] also reported that these models cannot accurately discriminate between conformational fluctuations which could bias the calculation of thermodynamic properties from simulations. Chen et al. [330] also reported that the fallacies of the solvent accessible surface area model are due to the lack of proper estimation

of the length-scale dependence of hydrophobic solvation. In principle, neglecting this would over-stabilize the pairwise interactions and predict the wrong sign in the contributions.

Nguyen et al. [193] reported that the solvent accessible surface area method is a rough approximation of the non-polar term used for implicit solvent calculations. Partly, due to its over-estimation of the non-hydrogen term which biases other conformations than the native state. Furthermore, the performance of the implicit solvent model is not only dependent on the solvent accessible surface area term but on the radii set as well, since the radii defines the boundary between solute and solvent. To further investigate this, Nguyen et al. (personal communication) simulated CLN025 using the mod1φ modification with original Bondi radii [331] parameter for the oxygen atom of aspartic /glutamic acid and the solvent accessible surface area term. As seen in figure 81, including the solvent accessible surface area (SASA) term and changing the radii from mbondi 3 as described by Nguyen et al. [193] to the original Bondi et al. parameter set [331] for aspartic/glutamic acid increases the fraction folded for this hairpin in comparison to what is observed in figure 80 for mod1φ.

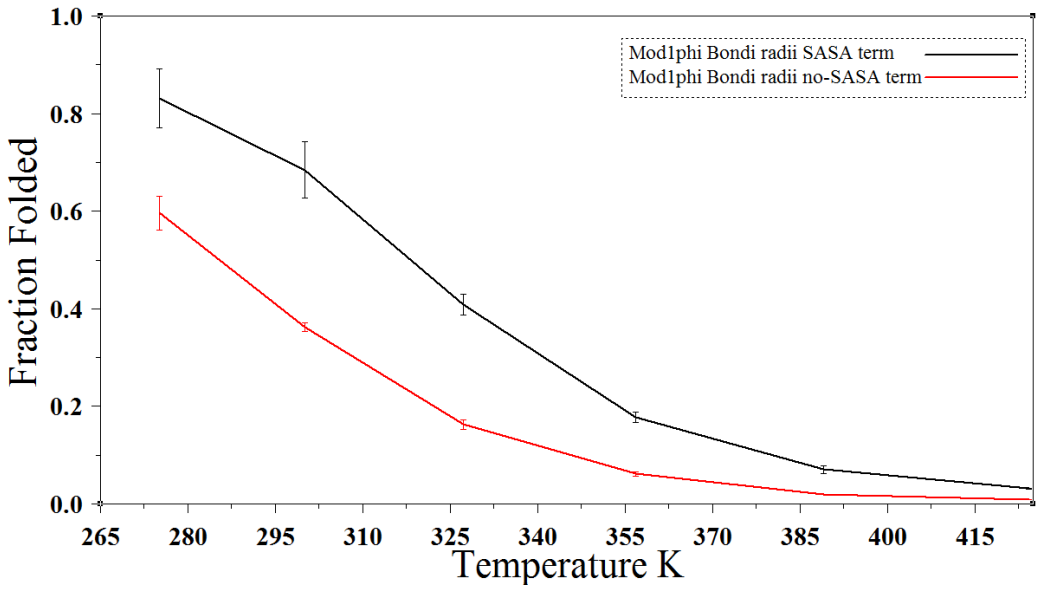


Figure 80 Melting curve for CLN025 from simulations of mod1φ with revised GB Neck model changing the oxygen radii from mBondi 3 [193] to Bondi [331] and including the solvent accessible surface area term (black) and excluding it from the calculation (red)

As seen in figure 81, the results of the simulation are not only dependent on the force field used but they are also dependent on the parameters used for the GB implicit solvent calculation. Even though as discussed previously the solvent accessible surface area term

(SASA) has many limitations for which it should not be included in the calculation, the radii terms used for the GB implicit solvent calculation also influences the folded populations obtained from simulations. The reason why this effect is observed in this case is due to the stabilization of salt bridges previously reported by Geney et al. [281]. In particular in the case of CLN025 the presence of glutamic acid (at the third position) and aspartic acid (at the fifth position) having increased oxygen radii stabilize hydrogen bonding to the neighboring threonine residues (the sixth and eight positions in the sequence). According to Nguyen et al. (personal communication) this stabilizing effect is approximately 1 kcal/mol. Although the results shown here do not justify modifying the existing parameters for the revised GB Neck model as described by Nguyen et al. [193]; on the contrary they highlight the limitations of the combination of methods used for the simulations such as force field and solvent model.

3.10.2 HP5F

Following up on the results observed from CLN025 peptide, we decided to use this 16 residue hairpin as a test system as Nguyen et al. [193] had done previously for the validation of their implicit solvent parameters. For HP5F we tested mod1φ, mod2φ and mod5φ in order to investigate how the combination of force field parameters with the revised GB Neck model [193] would affect the results. Similarly to the case of CLN025 we used REMD [249, 250] in order to calculate the melting profiles from backbone RMSD populations (figure 82) as Nguyen et al. [193] had done for the validation of their parameters.

Our simulations showed that the modified force fields are able to fold this hairpin, however the force field that had the most percent folded was mod5φ, followed by mod2φ and mod1φ had the least percent folded. In terms of convergence we also noticed that mod2φ had the longest simulation runs, requiring 300 ns to converge between native and extended simulations. Mod1φ was second with 240 ns and mod5φ was third requiring 180 ns simulation time. On the other hand, ff99SB had the least required simulation time, folding at 70 ns and having the most percent folded populations in comparison to the other force fields (see figure 82). An analog hairpin discussed in section 2.3.5.2 where tyrosine 3 was mutated to tryptophan and tryptophan 5

was mutated to tyrosine known as GB1m3[194] was simulated with similar results (see figure 83).

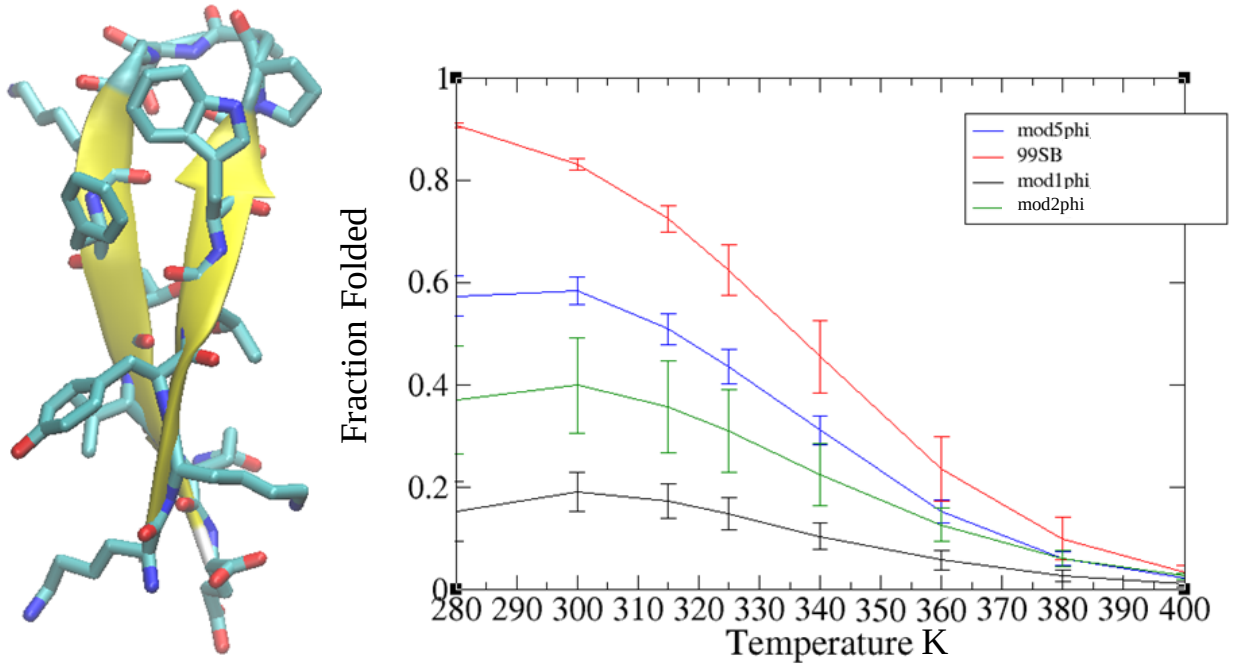


Figure 81 Melting curves for mod1φ, mod2φ, mod5φ modifications (right panel) from simulations of HP5F (left panel). The error bars were taken from two independent simulations (native and extended) using revised GB Neck model [193] as implicit solvent model. The melting curves were generated from backbone RMSD calculations for each post-processed temperature trajectory in the range of the simulation (280-400K). Experimentally, the melting curve was generated from NMR Chemical Shift Deviation (CSD) profiles. According to experiments the fraction folded for HP5F at 298K is 0.82 [191]

As previously discussed for CLN025, the results shown are dependent on the implicit solvent model parameters chosen for the calculation such as radii and solvent accessible surface area. Although Nguyen et al. [193] described these deficiencies, the propensities of the modifications for both hairpins (HP5F and GB1m3) show results that are expected. As seen in figures 82 and 83 the force field that has the least amount of fraction folded is mod1φ. Although the effect is more pronounced at room temperature for HP5F (0.2) than GB1m3 (0.4) and the folded populations were higher for GB1m3 than HP5F the trends were the same. Mod1φ the force field that had the greatest populations at the φ = -60º region and the lowest populations for -150º $\leq \varphi \leq$ -120º as indicated from the simulations of Ala_5, had the lowest folded populations for these hairpins. Similarly mod2φ, which had comparable results to mod1φ, had less folded populations than ff99SB, although slightly higher. Mod5φ which had similar populations as

ff99SB but greater sampling for the -150º ≤ φ ≤ -120º region (see figure 42) had the highest folded populations for the modifications simulated here.

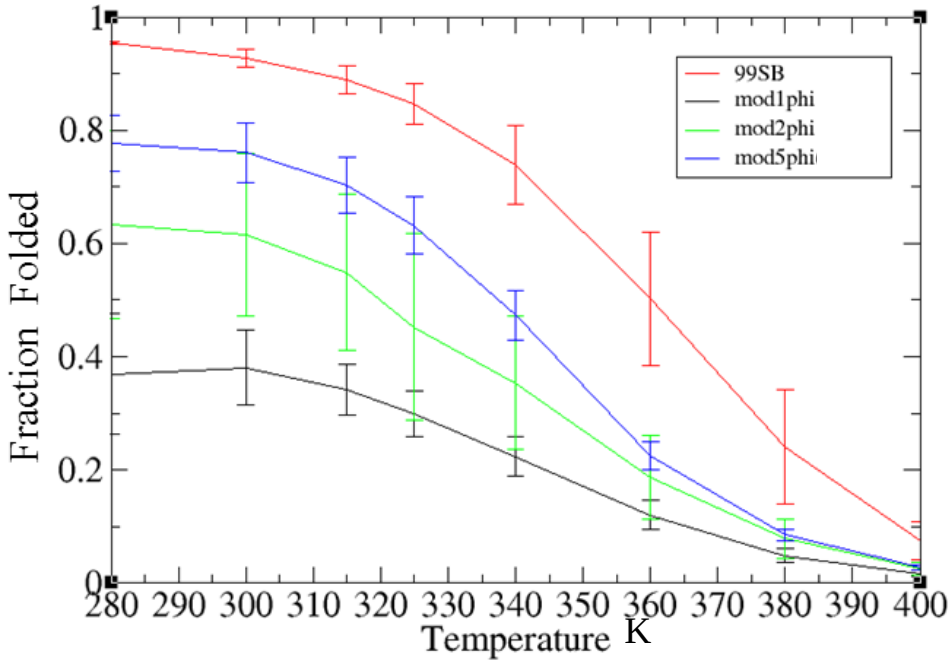


Figure 82 Melting curves for different modifications from simulations performed with GB1m3. The error bars were taken from two independent simulations (native and extended). The melting profiles were generated from backbone RMSD calculations for each post-processed temperature trajectory in the range of the simulation (280-400K). Experimentally, the melting curve was generated from NMR Chemical Shift Deviation (CSD) profiles. According to experiments this hairpin is 50% at 54-60ºC [194].

Even though the modification that had the best performance for this test was mod5φ we continued to consider mod1φ as the best modification overall because it had the best results for the systems studied here. For the testing of these hairpins in explicit solvent we chose mod1φ and compared its performance to ff99SB for chignolin and GB1m3.

3.11 Stability of hairpin simulations in explicit TIP3P solvent model

3.11.1 Chignolin

In the case of chignolin only the backbone RMSD of the MD trajectories was measured as a way to determine how the simulation was sampling the different RMSD values for 100 ns. This was done to the native and extended structures of both hairpins in order to determine how the backbone of the hairpin were behaving in this given time and not intended to define folded

states; instead this measurement was used to conclude the stability of the hairpin structure during this time. The simulations indicated that for one of the structures, mod1φ was able to find the simulated native conformation faster than ff99SB. This cannot be considered folded because it would require more simulation statistics in order conclude that the simulation has reached the native state, however it is encouraging that the simulations are sampling folded and unfolded conformations without preferring a single conformation. Furthermore, the starting native conformation simulation is not unfolding completely. More semi-extended conformations were evaluated in order to have more statistics where folded and unfolded conformations were happening.

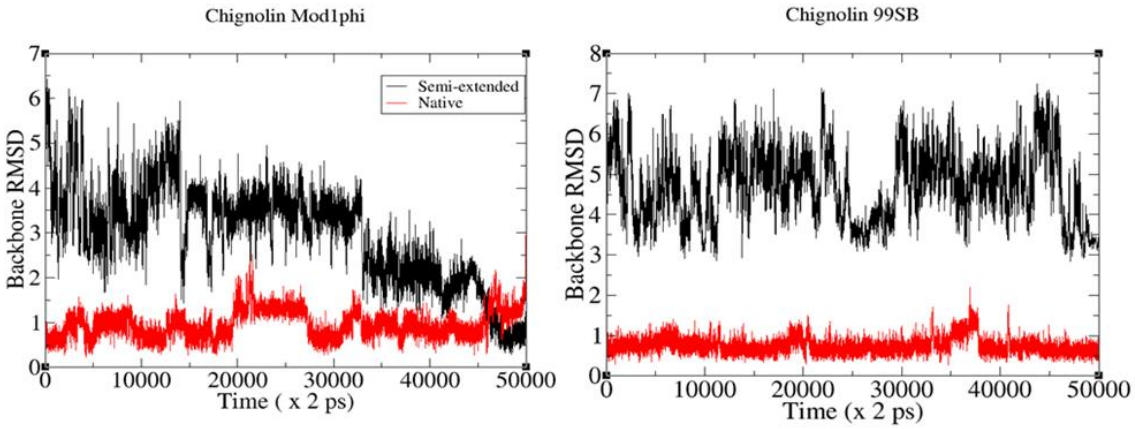


Figure 83 Comparison between mod1φ (left panel) and ff99SB (right panel) panel backbone RMSD values. Mod1φ appears to reach the same RMSD value for the native and semi-extended conformations faster than ff99SB. Simulations were run at 300K in explicit TIP3P waters. The backbone RMSD was calculated for the N, C_α, C atoms for residues 2-9. The termini residues were excluded from the calculation.

The results from the simulations of the multiple structures of chignolin indicated that 3 out of the 5 conformations simulated here reached a common point for backbone RMSD values faster than the simulation of ff99SB. It appears that the native conformation of ff99SB stays closer to 1Å than the native conformation of mod1φ (2Å or more). Furthermore, it is unknown if this effect could be observed for different conformations of ff99SB or if these conformations of ff99SB would reach the native conformation just as fast as mod1φ. We only simulated different conformations of mod1φ in order to have more than one simulation where this effect was observed, further analysis would require to extend the ff99SB simulation longer or to simulate the same conformations used for the simulations of mod1φ using ff99SB as force field.

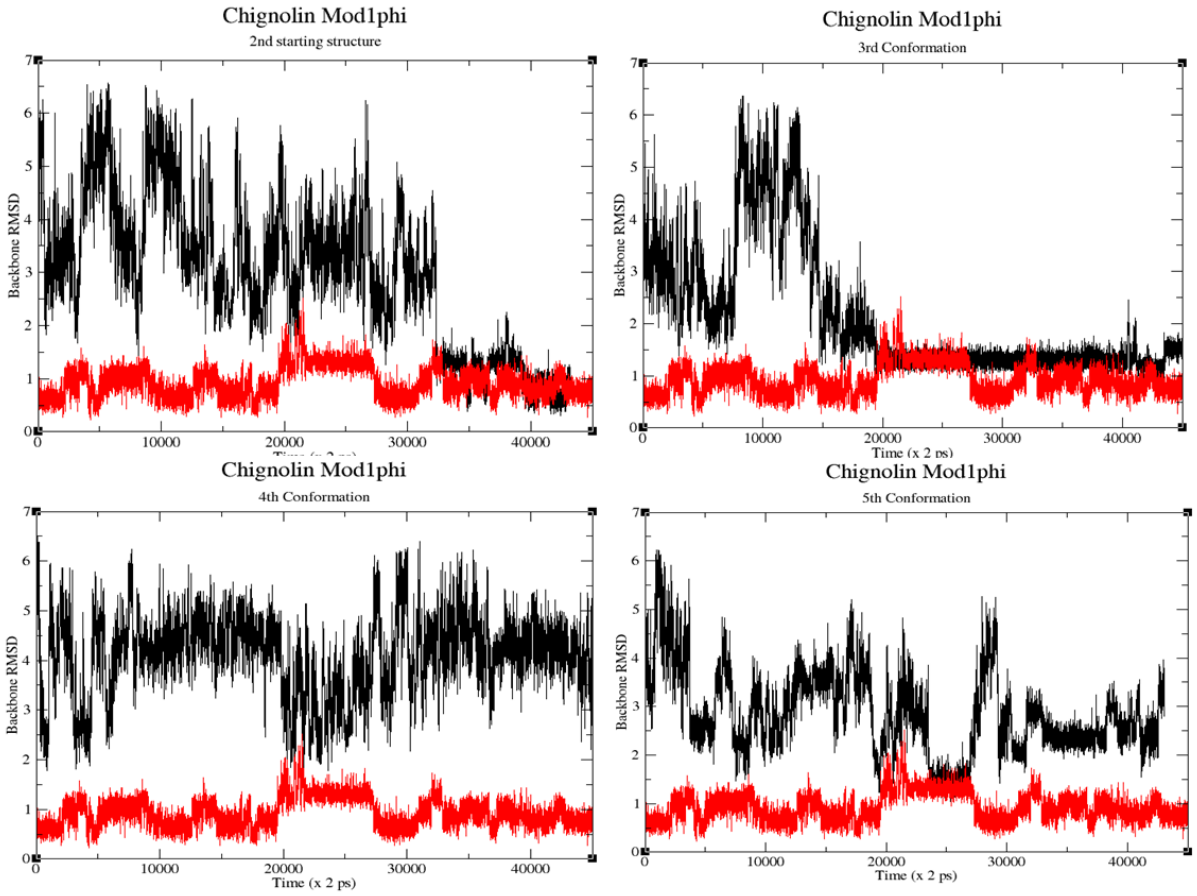


Figure 84 Different conformations for chignolin generated from snapshots of a trajectory generated at 400K (black curves). The native conformation (red) was the same for all graphs. Simulations were run at 300K in explicit TIP3P waters. The backbone RMSD was calculated for the N, C_α, C atoms for residues 2-9. The termini residues were excluded from the calculation.

The fact that mod1φ was able to reach low backbone RMSD values close to the native conformation was encouraging because it indicates that this modification can be applied to simulations of hairpins in explicit water. As simulations in the microsecond and millisecond time scales have become increasingly feasible, longer simulation times would be required to establish the ability of this modification to preserve the stability of this hairpin.

Since this metric is not quantitative enough to describe the conformations obtained from simulations of hairpins, other methods can be implemented for the validation of force field parameters as in the case of Best et al. [59] who compared experimental and calculated scalar coupling values from simulations of chignolin and Nuclear Overhauser effects. Although we know that calculating scalar coupling values from dihedral angles requires Karplus functions and that they have many caveats previously described this approach could be useful for force field

assessment; provided that the estimates for the uncertainties in the Karplus methods are accounted for in the calculations.

3.11.2 GB1m3

Similarly to the approach followed for the simulations of chignolin, we ran simulations of 4 semi-extended conformations of GB1m3 hairpin. As seen in figure 86, ff99SB seems to reach a state that resembles the native conformation faster than mod1φ. Furthermore, the native simulation of mod1φ appear to reach an alternative conformation around 3-4Å as seen in figure 87 while the ff99SB simulation did not display this behavior. When visually inspecting these structures we noticed the strands of the hairpin become curled up and the hydrogen bonding pattern is disrupted (figure 87).

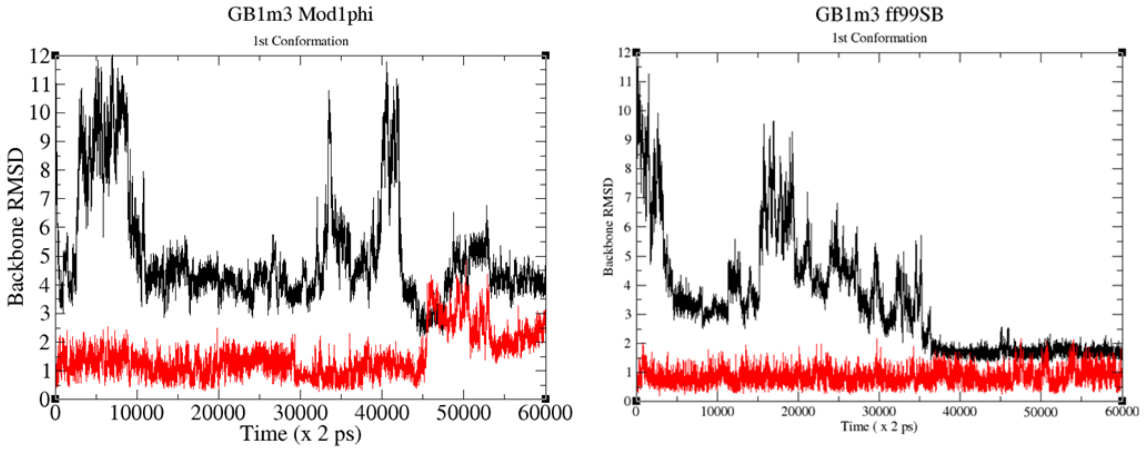


Figure 85 Results from MD simulations for semi-extended and native structures with mod1φ (left graph) and ff99SB (right graph). The backbone RMSD was calculated for the N, C_α, C atoms for residues 2-15. The termini residues were excluded from the calculation.

Nevertheless, it is unknown if the simulation would be trapped at this conformation, and how much more time it will be required to obtain enough folding and unfolding statistics. 99SB seems to need more time in order to have more folding-unfolding events. Other conformations were also simulated for shorter simulation time; however, results are not conclusive enough to determine which force field is better at folding this hairpin (see figure 88)

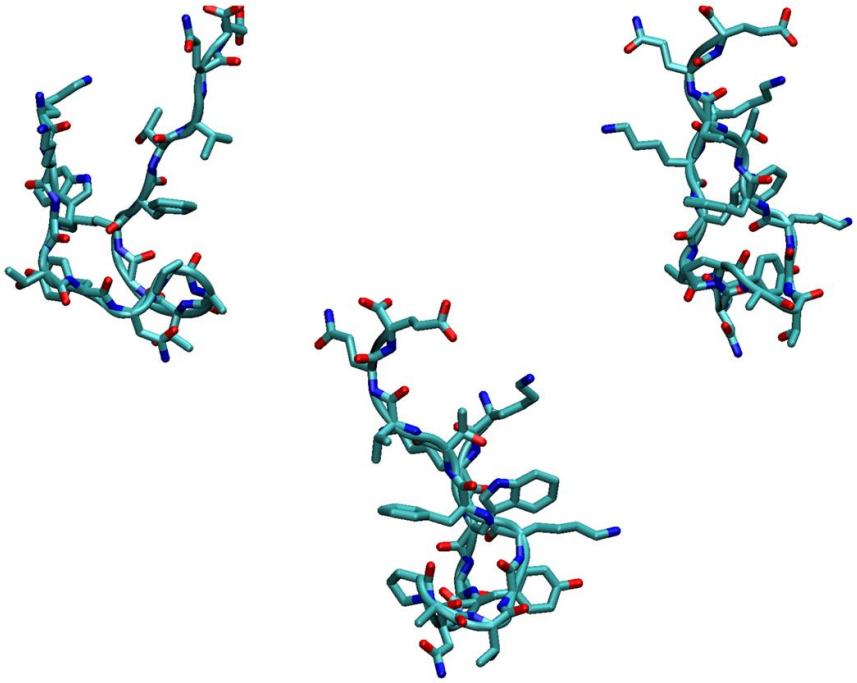


Figure 86 Snapshots of alternate native structure for GB1m3. The backbone RMSD for this structure is 3.26 Å

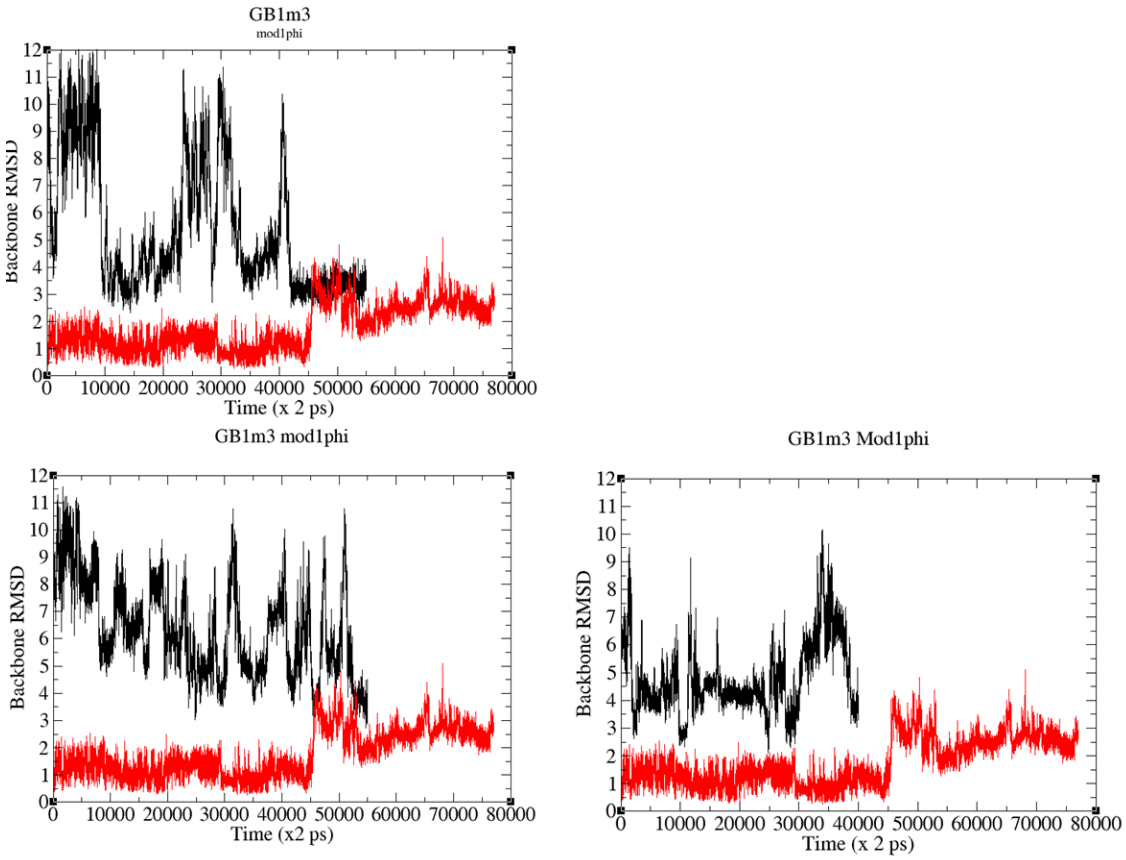


Figure 87 GB1m3 semi-extended conformations in explicit TIP3P waters using mod1φ. The backbone RMSD was calculated for the N, Cα, C atoms for residues 2-15. The termini residues were excluded from the calculation.

Although we know that the modifications are designed to increase the sampling of the ppII region which in turn would slightly decrease the sampling of the β region, we are not certain if this effect is only dependent on the modifications proposed here or if the side-chain parameters could play a role in the sampling of the backbone conformations. As previously described in the case of Val_3 the side-chain preferences of the force field influence the backbone conformations which justify the fitting of the side-chain parameters accordingly. Previous work by Lindorff-Larsen et al. [158] described re-parameterization of the side-chain dihedral functions of ff99SB for Leucine, Isoleucine, Asparagine and Aspartic acid as they were reportedly the amino-acids with the greatest deficiencies. However it would be good to evaluate all of the amino-acid side-chain parameters as they have not been re-fitted since Cornell et al. [134] did it with ff94. Recently, the work by Maier et al. [326] provides revisions to the backbone parameters for ff99SB that address these issues.

4. Discussion and Conclusions

Among the goals that drove the work presented here was to increase the agreement between simulated and experimental vicinal scalar couplings for poly-alanine peptides. In order to do this it was necessary to increase the energy barrier between the ppII and β secondary structure basins. Doing this would increase the sampling of the ppII dihedral region of the Ramachandran plot which would improve the results obtained from the simulations. Furthermore, we aimed at improving the helical propensities from simulations as well as applying modifications to the ψ/ψ' backbone energy terms. Increasing the energy barrier between the ppII and β secondary structure basins would also correct the energy function for the helical basin because the helical and ppII secondary structure basins are located in the $-60^{o} \geq \varphi \geq -90^{o}$ dihedral region. Therefore, augmenting the sampling of this region and the $60^{o} > \psi > -60^{o}$ region would achieve these two purposes.

Following this procedure we generated 29 modifications that were tested on a small peptide system known as Ala_5 for which experimental data has been reported by Graf et al. [172]. This peptide has been used previously by other groups[89, 173] for this type of testing. This homologous peptide is appropriate for this purpose because it has enough residues to match the turn of a helix (*i, i+4*). Since helical propensities for small peptides are primarily governed by dipole interactions, having the hydrogen bonding of a helical turn would define the dihedral populations for this secondary structure region more accurately than for shorter homologous peptides. Furthermore, it is known that poly-alanine peptides have mainly PPII character; although as the number of residues increases, the helical content increases as well[234].

Initially we concentrated in calculating the agreement for the $^{3}J(H_N, H_\alpha)$ coupling value because this is one of the scalar coupling values that simulations with ff99SB did not reproduce well. Then we continued at evaluating the agreement with other scalar couplings that Graf et al. [172] measured as well. From the modifications tested we noticed that mod1φ, mod2φ, mod1φ-

mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ improved the agreement for ^3J (H_N , H_α) and the rest of the scalar couplings measured by Graf et al. We calculated this agreement for all the scalar couplings in the form of χ^2 values using the original Karplus parameters described in table 7. The actual calculated values can be seen in tables 18 and 19. It is important to note that these calculated scalar coupling values are obtained from Karplus functions obtained by fitting NMR and X-ray crystallographic structural data. As mentioned previously, the Karplus function attempts to fit complex chemical effects such as spin-spin interactions that are dependent on many more variables than just the dihedral angles between the nuclei. Furthermore it was shown that some of the scalar coupling values in particular ^3J (H_N , H_α) that probes for the φ angle is dependent not only on the φ angle values but on the ψ angles as well, and the regions that define specific scalar coupling values are so broad that they can sometimes overlap different secondary structure regions (see figure 51). Primarily we used the ^3J (H_N , H_α) scalar coupling because it has a wide range of values which would allow to distinguish the different regions in φ dihedral space, mainly the ppII and β regions. However it is apparent that the relation between the dihedral angles and the scalar couplings taken from NMR spin-spin interactions is more complex than initially thought [323]. Besides this, as mentioned previously there are many different flavors of Karplus function parameters available for this type of calculations. Some experimentally derived[175] and some obtained from quantum calculations [235]. As illustrated in table 20 the results obtained depend on the set of parameters used for the calculations which could mislead the interpretation of the results obtained from simulations. For our purpose we chose the original Karplus parameters for the validation of the modifications because these parameters were obtained from fitting to experimental ubiquitin data. In the case of the other parameters known as DFT1 and DFT2 they were fitted to the backbone conformations of small alanine peptides where side-chain effects on the backbone dihedrals present in higher branched amino-acids are not being accounted for. A comparison between the χ^2 values obtained from DFT1, DFT2 and original parameters is shown in table 20 for all 29 modifications.

We also evaluated the modifications against Val_3 another small homologous peptide with vicinal scalar coupling data published by Graf et al [172]. We followed a similar approach as the one used for Ala_5 where initially we calculated ^3J (H_N , H_α) scalar coupling values from φ dihedral angles and then the agreement with the other scalar coupling values was reported in the

form of χ^2 values. In this case however the results seen for mod1φ, mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ were not indicative of the best agreement. Instead they were some of the force fields that had the worst agreement with experiments. The only modification that had fairly good agreement was mod2φ (χ^2 = 1.49±0.02). We decided to investigate this by examining the backbone dihedral propensities for all the amino-acids with the exception of glycine and proline based on the dihedral libraries taken from the protein data bank by Lovell et al. [149]. We discovered that the backbone dihedral propensities for most of the amino-acids are different than alanine, especially for higher branched amino-acids where the differences are significant, being valine one of these cases. Furthermore, when plotting the φ/ψ dihedral angles based on the χ_1 dihedral conformations defined by Lovell et al. as gauche-, gauche+ and trans we discovered that the results were remarkably different. To quantify this we re-calculated the ^3J (H_N , H_α) scalar coupling values and the results are seen in figure 67. This indicates that the backbone dihedrals are influenced by the side-chain propensities and this needs to be taken into account when evaluating the performance of force fields especially for higher branched amino-acids.

In terms of the helical propensities we simulated two helical peptides (Hydrogen Bond Surrogate Peptide and K19) and noticed that the modifications to the φ' dihedral term, particularly mod1φ, mod2φ and mod5φ increased the fraction helical propensities in comparison to ff99SB from approximately 0.06 to 0.25. Other modifications that performed well for Ala_5 showing improvement in the agreement with scalar couplings such as mod1φ-mod1ψ, mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ also demonstrated significant proliferation of the helical conformations (fraction helicities of 0.5 to 0.7); however the helical propensities were slightly larger than those observed from experiments (figures 68 and 74). Although the results shown here cannot be compared directly to the experimental values because our definition of helical content is based on the hydrogen bond definition given by Dictionary of Secondary Structure (DSSP) analysis whereas the experimental helical content definition is based on circular dichroism and chemical shift deviation analysis. Nevertheless it was encouraging to see that these modifications were performing better than ff99SB for these systems.

NMR order parameters for lysozyme were calculated from simulations in order to determine how the backbone flexibility was affected. It was observed that mod1φ-mod1ψ,

mod1φ-mod2ψ, mod1φ-mod3ψ and mod1φ-mod4ψ were adding significant flexibility to the backbone in regions where ff99SB had good agreement with experiments (figure 77). Overall, mod1φ had one of the best agreements closely followed by mod2φ and mod5φ was the last in the series. In the case of ubiquitin, mod1φ, mod2φ and mod5φ were used for the calculation of NMR order parameters from simulations. The results indicated that there is no significant difference between them, with the exception of mod2φ which appears to be slightly more flexible for residues 52 and 53 of the sequence as seen in figures 78 and 79.

The results from the hairpin simulations (GB1m3 and HP5F) in implicit solvent reveal that the ff99SB had the closest agreement with experiments for the melting curves, followed by mod5φ, mod2φ and last mod1φ (see figures 82 and 83) . The simulations for the modifications also took longer time to converge between the native and extended conformations following the same trend as the melting curves populations (ff99SB > mod5φ > mod2φ > mod1φ). Although the experimental and simulation results cannot be directly compared because they do not have the same metrics in both cases; the melting curves were obtained from thermal denaturation followed by chemical shift deviations whereas in the case of the simulations the folded populations were obtained from RMSD calculations for the N, C_α, C backbone atoms taking the native conformation as the reference point. Nevertheless, the experimental results can serve as a qualitative metric to have an estimate as to how much fold the structures should be. Furthermore, the simulation results are dependent on many more factors than the force field backbone parameters being tested. As seen in figure 81 and discussed in section 3.10.1 the implicit solvent parameters play a crucial role in the results obtained. As illustrated in figure 81 changing parameters such as the atomic radii can stabilize the structure and yield significantly higher folded populations because the atomic radii defines the boundaries between the solute and solvent interactions. For the simulations shown, we did not include the solvent accessible surface area term in the calculation because it has been previously reported that this method does not accurately reproduce the intrinsic conformations of the systems being studied because they "bias" alternate conformations based on the parameters used for their fitting. However as shown in figure 81, including this term can significantly increase the fraction folded populations by a four-fold. Therefore careful attention must be paid when analyzing the results obtained from implicit solvent simulations.

We also tested other hairpin structures known as chignolin[185] and GB1m3 [194] in TIP3P waters to determine if the stability of the MD simulations would be compromised by using mod1φ. We chose this force field because it had the best performance for most of the test systems evaluated. We ran MD simulations of native and semi-extended structures of both hairpins in order to determine how the backbone of the hairpin was behaving in this given simulation time to determine the stability of the hairpin structure.

Surprisingly, when mod1φ was compared to ff99SB; the simulations with mod1φ reached a common state that resembled the folded structure faster than ff99SB. Furthermore, the native simulation did not completely unfold in the time scale simulated. We also simulated 4 other semi-extended conformations and two out of the four simulations had similar behavior. Since we did not simulated more semi-extended conformations for ff99SB we do not know if similar behavior could be seen in the simulations of ff99SB. However, it is encouraging to see that 3 out of the 5 conformations can reach a common state that resembles the folded structure. Furthermore, the simulations fold and unfold without preferring the random-coil state.

In the case of the GB1m3 simulations, mod1φ encountered alternative semi-folded conformations with 3-4Å backbone RMSD from native sequence as shown in figure 86. It is unknown if the simulations are trapped in this alternative conformation, therefore longer simulation time is required to make further determinations. Furthermore, the chignolin simulations should also be run longer in order to have robust statistics of behavior.

As mentioned previously, using calculations such as Root Mean Square Deviations from backbone atoms is not a robust metric that can quantitatively benchmark the quality of the modifications proposed. Recently, Best et al. [59] reported the comparison for vicinal scalar couplings between experiments and simulations for chignolin. They measured the same scalar coupling values that Graf et al. [172] had reported earlier for Ala_5 and reported the improvement in the agreement in the form of χ^2 values calculated from Karplus functions. Although we know that Karplus functions have many caveats that directly affect the results obtained, this method could be useful for benchmarking the results obtained from simulations in a more quantitative way, provided that the uncertainties in the Karplus functions are accounted for.

Overall, mod1φ showed improvement in the agreement with scalar couplings for Ala_5 and higher helical propensities for transient helical peptides such as Baldwin-type peptide known as K19[167] and a smaller peptide derived from the Bak BH3 domains [168] . Furthermore, improved agreement with NMR order parameters for Lysozyme was also observed. In the case of the hairpins, the implicit solvent simulations indicate that mod1φ had considerably less fraction folded than ff99SB. However the results are dependent on many more aspects than just the backbone parameters tested here, the parameters for the implicit GB solvent model also play an essential role in the calculations. In the case of the explicit TIP3P solvent simulations, they are not certain enough to determine if mod1φ is better or worse than ff99SB. Longer simulations are required to address these questions.

The force field is a key factor in the simulation of biomolecules since they are the description of the molecule being studied and careful attention must be paid to the choice of force field. We present a variant of the widely used ff99SB AMBER force field. The recommended force field only includes marginal changes to the existing energy function of ff99SB aiming to preserve some of the qualities of this force field while enhancing the areas in which the force field was lacking. The criteria for improvement were based on the agreement with scalar couplings for Ala_5 and helical content for HBS as well as K19. The results for mod1φ showed improvement in the agreement with scalar couplings of Ala_5 and helical content propensity of the force field, furthermore it also showed better agreement with the results obtained from NMR relaxation experiments for lysozyme. The results from hairpin simulations are not yet conclusive to determine if mod1φ performs better or worse than ff99SB. However, in the case of Ala_5, Hydrogen Bond Surrogate, K19, lysozyme and ubiquitin mod1φ showed significant increased ppII and helical propensities when compared to ff99SB. Furthermore, the calculated vicinal scalar coupling values of Ala_5 from simulation improved significantly.

Although the modifications presented here are based on changing the backbone dihedral parameters of ff99SB there are other terms in the force field that are coupled to the dihedral propensities such as the non-bonded electrostatics and van der Waals terms that have been recently re-fitted. Nerenberg et al. [157] recently reported development of the van der Waals terms by re-parameterization to small hydrocarbon molecules without changing the electrostatics that were published by Cornell et al. [134]. Cerutti et al. [321] also reported the development of a

new polarizable charge model for amino-acids. These are few efforts attempting to revise these terms as they have been left untouched for many years; most of the efforts for improving force fields including the AMBER ones have been concentrated in improving the backbone dihedral terms while leaving the non-bonded terms untouched. Future improvement would require to couple the effects of these terms to the dihedral parameters as well as developing better methods to benchmark the results obtained from simulations since the current methods have many limitations that can hinder the correct assessment of force field quality.

References

1. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. **28**(1): p. 235-242.
2. Binnig, G., C.F. Quate, and C. Gerber, *Atomic Force Microscope.* Physical Review Letters, 1986. **56**(9): p. 930-933.
3. Rostislav, V.L., *Feature-oriented scanning methodology for probe microscopy and nanotechnology.* Nanotechnology, 2004. **15**(9): p. 1135.
4. Buck, M., et al., *Structural Determinants of Protein Dynamics: Analysis of 15N NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg White Lysozyme.* Biochemistry, 1995. **34**(12): p. 4041-4055.
5. Roberts, S.T., et al., *A fast-scanning Fourier transform 2D IR interferometer.* Optics Communications, 2011. **284**(4): p. 1062-1066.
6. Ebbinghaus, S., et al., *Protein Folding Stability and Dynamics Imaged in a Living Cell.* Nature Methods, 2010. **7**(4): p. 319-323.
7. *Ab initio molecular dynamics calculations with simple, localized, orthonormal real-space basis sets.* Physical Review B, 2003. **68**(12): p. 125110.
8. *Ab Initio Calculation of Hyperfine Interaction Parameters: Recent Evolutions, Recent Examples.* Hyperfine Interactions, 2005: p. 1.
9. Alexey Onufriev, D.B., David A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model.* Proteins: Structure, Function, and Bioinformatics, 2004. **55**(2): p. 383-394.
10. *On-the-fly localization of electronic orbitals in Car–Parrinello molecular dynamics.* The Journal of Chemical Physics, 2004. **120**(5): p. 2169.
11. *Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit.* The Journal of Chemical Physics, 2006. **125**(15): p. 154507.
12. E., T.M., *Ab initio molecular dynamics: basic concepts, current trends and novel applications.* Journal of Physics: Condensed Matter, 2002. **14**(50): p. 1297-1355.
13. Feynman, R. and A.R. Hibbs, *Quantum Mechanics and Path Integrals.* 1965.
14. Feynman, R., *Statistical Mechanics.* 1972.
15. Piana, S., et al., *Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site.* Journal of the American Chemical Society, 2001. **123**(36): p. 8730-8737.
16. Hutter, J., P. Carloni, and M. Parrinello, *Nonempirical Calculations of a Hydrated RNA Duplex.* Journal of the American Chemical Society, 1996. **118**(36): p. 8710-8712.
17. Rothlisberger, U. and P. Carloni, *Ab initio molecular dynamics studies of a synthetic biomimetic model of galactose oxidase.* International Journal of Quantum Chemistry, 1999. **73**(2): p. 209-218.
18. Andreoni, W., A. Curioni, and T. Mordasini, *DFT-based molecular dynamics as a new tool for computational biology: First applications and perspective.* IBM Journal of Research and Development, 2001. **45**(3.4): p. 397-407.
19. Rovira, C. and M. Parrinello, *First-principles molecular dynamics simulations of models for the myoglobin active center.* International Journal of Quantum Chemistry, 2000. **80**(6): p. 1172-1180.
20. Rovira, C., et al., *Influence of the Heme Pocket Conformation on the Structure and Vibrations of the Fe-CO Bond in Myoglobin: A QM/MM Density Functional Study.* Biophysical journal, 2001. **81**(1): p. 435-445.

21. Lee, Y.T. and J.S. Lin, *Ab Initio molecular dynamics study of ethylene adsorption onto Si(001) surface: Short-time fourier transform analysis of structural coordinate autocorrelation function.* Journal of Computational Chemistry, 2013. **34**(31): p. 2697-2706.
22. Zhao, J., et al., *Ab initio molecular dynamics studies on the growth of ammonium chloride clusters.* Theoretical Chemistry Accounts, 2013. **132**(9): p. 1-11.
23. Jakse, N. and A. Pasturel, *Liquid Aluminum: Atomic diffusion and viscosity from ab initio molecular dynamics.* Sci. Rep., 2013. **3**: p. 1-8.
24. Liu, Z., L.E. Carter, and E.A. Carter, *Full Configuration Interaction Molecular Dynamics of Na2 and Na3.* The Journal of Physical Chemistry, 1995. **99**(13): p. 4355-4359.
25. Di Martino, B., M. Celino, and V. Rosato, *An high performance Fortran implementation of a Tight-Binding Molecular Dynamics simulation.* Computer Physics Communications, 1999. **120**(2–3): p. 255-268.
26. Friesner, R.A., *Solution of self-consistent field electronic structure equations by a pseudospectral method.* Chemical Physics Letters, 1985. **116**(1): p. 39-43.
27. Lippert, B.G., J.H. Parrinello, and Michele, *A hybrid Gaussian and plane wave density functional scheme.* Molecular Physics, 1997. **92**(3): p. 477-488.
28. Kadowaki, H., Nishitoh, H., Urano, F., Sadamitsu, C., Matsuzawa, A., Takeda, K., Masutani, H., Yodoi, J., Urano, Y., Nagano, T. and Ichijo, H., *Amyloid [beta] induces neuronal cell death through ROS-mediated ASK1 activation.* Cell Death Differ, 2005. **12**(1): p. 19-24.
29. Smith, A., *Protein Misfolding.* Nature, 2003. **426**: p. 883-909.
30. Tang, Y., et al., *Multistate Folding of the Villin Headpiece Domain.* Journal of Molecular Biology, 2006. **355**(5): p. 1066-1077.
31. Ng, S.P., Randles, L. G. and Clarke, J., *Single molecule studies of protein folding using atomic force microscopy.* Methods Mol. Biol., 2007(350): p. 139-167.
32. *Scanning Electron Microscope: Advantages and Disadvantages in Imaging.*
33. Leach, A., Molecular Modeling, 1998. **3rd Edition**.
34. Simmerling, C., B. Strockbine, and A.E. Roitberg, *All-Atom Structure Prediction and Folding Simulations of a Stable Protein.* Journal of the American Chemical Society, 2002. **124**(38): p. 11258-11259.
35. Hornak, V., et al., *HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(4): p. 915-920.
36. Ferrara, P. and A. Caflisch, *Folding simulations of a three-stranded antiparallel β-sheet peptide.* Proceedings of the National Academy of Sciences, 2000. **97**(20): p. 10780-10785.
37. Ferrara, P. and A. Caflisch, *Native topology or specific interactions: what is more important for protein folding?* Journal of Molecular Biology, 2001. **306**(4): p. 837-850.
38. Berhanu, W.M. and U.H.E. Hansmann, *The stability of cylindrin β-barrel amyloid oligomer models—A molecular dynamics study.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(9): p. 1542-1555.
39. Beck, D.A.C., G.W.N. White, and V. Daggett, *Exploring the energy landscape of protein folding using replica-exchange and conventional molecular dynamics simulations.* Journal of Structural Biology, 2007. **157**(3): p. 514-523.
40. Brooks, C.L., III, et al., *Chemical physics of protein folding.* Proceedings of the National Academy of Sciences, 1998. **95**(19): p. 11037-11038.
41. Fersht, A. and V. Dagget, *Protein folding and unfolding at atomic resolution.* Cell, 2002. **108**: p. 573 - 582.
42. Guha, J., V. Vishal, and S.P. Vijay, *Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece.* The Journal of Chemical Physics, 2006. **124**(16): p. 164902.

43. Kamirski, S., et al., *Protein folding from a highly disordered denatured state: The folding pathway of chymotrypsin inhibitor 2 at atomic resolution.* Proc Natl Acad Sci U S A, 2001. **98**: p. 4349 - 4354.
44. Min-yi Shen, K.F.F., *All-atom fast protein folding simulations: The villin headpiece.* Proteins: Structure, Function, and Genetics, 2002. **49**(4): p. 439-445.
45. Piana, S., K. Lindorff-Larsen, and David E. Shaw, *How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization?* Biophysical Journal, 2011. **100**(9): p. L47-L49.
46. Rhee, Y.M. and V.S. Pande, *Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation.* Biophys. J., 2003. **84**(2): p. 775-786.
47. Spoel, D., A. Patriksson, and M.M. Seibert, *Protein Folding Properties from Molecular Dynamics Simulations*, in *Applied Parallel Computing. State of the Art in Scientific Computing*, B. Kågström, et al., Editors. 2007, Springer Berlin Heidelberg. p. 109-115.
48. Andrec, M., et al., *Chemical Theory and Computation Special Feature: Protein folding pathways from replica exchange simulations and a kinetic network model.* Proceedings of the National Academy of Sciences, 2005. **102**(19): p. 6801-6806.
49. Chowdhury, S., Lee, M. C., Xiong, G., Duan Y., *Ab initio folding simulation of the Trp-Cage mini protein approaches NMR resolution.* J. Mol. Biol., 2003. **327**: p. 711-7.
50. Day, R., D. Paschek, and A.E. Garcia, *Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(8): p. 1889-1899.
51. Ding, F., S.V. Buldyrev, and N.V. Dokholyan, *Folding Trp-Cage to NMR Resolution Native Structure Using a Coarse-Grained Protein Model.* Biophys. J., 2005. **88**(1): p. 147-155.
52. Duan, Y., L. Wang, and P.A. Kollman, *The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(17): p. 9897-9902.
53. Enemark, S. and R. Rajagopalan, *Turn-directed folding dynamics of [small beta]-hairpin-forming de novo decapeptide Chignolin.* Physical Chemistry Chemical Physics, 2012. **14**(36): p. 12442-12450.
54. Freddolino, P.L., et al., *Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain.* Biophysical Journal, 2008. **94**(10): p. L75-L77.
55. Fuzo, C. and L. Degrève, *Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin.* Journal of Molecular Modeling, 2012. **18**(6): p. 2785-2794.
56. Irbäck, A. and S. Mohanty, *Folding Thermodynamics of Peptides.* Biophysical Journal, 2005. **88**(3): p. 1560-1569.
57. Juraszek, J. and P.G. Bolhuis, *Sampling the multiple folding mechanisms of Trp-cage in explicit solvent.* Proceedings of the National Academy of Sciences, 2006. **103**(43): p. 15859-15864.
58. Kubelka, J., W.A. Eaton, and J. Hofrichter, *Experimental Tests of Villin Subdomain Folding Simulations.* Journal of Molecular Biology, 2003. **329**(4): p. 625-630.
59. Kührová, P., et al., *Force-Field Dependence of Chignolin Folding and Misfolding: Comparison with Experiment and Redesign.* Biophysical journal, 2012. **102**(8): p. 1897-1906.
60. Lee, I.-H. and S.-Y. Kim, *Dynamic Folding Pathway Models of the Trp-Cage Protein.* BioMed Research International, 2013. **2013**: p. 9.
61. Lei, H., et al., *Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations.* Proceedings of the National Academy of Sciences, 2007. **104**(12): p. 4925-4930.
62. Marinelli, F., et al., *A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations.* PLoS Comput Biol, 2009. **5**(8): p. e1000452.
63. Meuzelaar, H., et al., *Folding Dynamics of the Trp-Cage Miniprotein: Evidence for a Native-Like Intermediate from Combined Time-Resolved Vibrational Spectroscopy and Molecular Dynamics Simulations.* The Journal of Physical Chemistry B, 2013.
64. Nicolae-Viorel, B. and H. Gerhard, *Peptide folding kinetics from replica exchange molecular dynamics.* Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 2008. **77**(3): p. 030902.

65. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nature Reviews Drug Discovery, 2002. **1**(4): p. 309-315.
66. Wang, Y.-Y., J.C. Nacher, and X.-M. Zhao, *Predicting drug targets based on protein domains.* Molecular BioSystems, 2012. **8**(5): p. 1528-1534.
67. Bakheet, T.M. and A.J. Doig, *Properties and identification of human protein drug targets.* Bioinformatics, 2009. **25**(4): p. 451-457.
68. Helin, K. and D. Dhanak, *Chromatin proteins and modifications as drug targets.* Nature, 2013. **502**(7472): p. 480-488.
69. Showalter, S.A., et al., *Toward Quantitative Interpretation of Methyl Side-Chain Dynamics from NMR by Molecular Dynamics Simulations.* Journal of the American Chemical Society, 2007. **129**(46): p. 14146-14147.
70. Wickstrom, L., et al., *Reconciling the Solution and X-ray Structures of the Villin Headpiece Helical Subdomain: Molecular Dynamics Simulations and Double Mutant Cycles Reveal a Stabilizing Cation−π Interaction†.* Biochemistry, 2007. **46**(12): p. 3624-3634.
71. Simmerling, C., B. Strockbine, and A. Roitberg, *All-Atom structure prediction and folding simulations of a stable protein.* J Am Chem Soc, 2002. **124**: p. 11258 - 11259.
72. Showalter, S.A. and R. Brüschweiler, *Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field.* Journal of Chemical Theory and Computation, 2007. **3**(3): p. 961-975.
73. Lindorff-Larsen, K., et al., *Systematic Validation of Protein Force Fields against Experimental Data.* PLoS ONE, 2012. **7**(2): p. e32131.
74. Lange, O.F., D. van der Spoel, and B.L. de Groot, *Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data.* Biophysical Journal, 2010. **99**(2): p. 647-655.
75. Shao, Q., L. Yang, and Y.Q. Gao, *Structure change of β-hairpin induced by turn optimization: An enhanced sampling molecular dynamics simulation study.* The Journal of Chemical Physics, 2011. **135**(23): p. -.
76. Paschek, D., H. Nymeyer, and A.E. García, *Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water.* Journal of Structural Biology, 2007. **157**(3): p. 524-533.
77. Shen, M.a.F., K., *All-Atom Fast Protein Folding Simulations: The Villin Headpiece.* Proteins: Structure, Function, and Genetics, 2002. **49**: p. 439-445.
78. Kumar, A., et al., *Molecular Dynamic Simulation Reveals Damaging Impact of RAC1 F28L Mutation in the Switch I Region.* PLoS ONE, 2013. **8**(10): p. e77453.
79. Petrov, D., et al., *A Systematic Framework for Molecular Dynamics Simulations of Protein Post-Translational Modifications.* PLoS Comput Biol, 2013. **9**(7): p. e1003154.
80. Yuan, S., et al., *Lipid Receptor S1P1/ Activation Scheme Concluded from Microsecond All-Atom Molecular Dynamics Simulations.* PLoS Comput Biol, 2013. **9**(10): p. e1003261.
81. Chang, W.E., et al., *Molecular Dynamics Simulations of Anti-Aggregation Effect of Ibuprofen.* Biophysical journal, 2010. **98**(11): p. 2662-2670.
82. Shang, Y. and C. Simmerling, *Molecular Dynamics Applied in Drug Discovery: The Case of HIV-1 Protease #*, in *T Computational Drug Discovery and Design*. p. 527-549.
83. Borhani, D. and D. Shaw, *The future of molecular dynamics simulations in drug discovery.* Journal of Computer-Aided Molecular Design, 2012. **26**(1): p. 15-26.
84. Okimoto, N., et al., *High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations.* PLoS Comput Biol, 2009. **5**(10): p. e1000528.
85. Duan, Y., 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**(16): p. 1999-2012.
86. Warshel, A., M. Kato, and A.V. Pisliakov, *Polarizable Force Fields: History, Test Cases, and Prospects.* Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2034-2045.

87. Halgren, T.A. and W. Damm, *Polarizable force fields.* Current Opinion in Structural Biology, 2001. **11**(2): p. 236-242.
88. Rick, S.W. and S.J. Stuart, Rev. Comput. Chem., 2002. **18**.
89. Best, R.B. and G. Hummer, *Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides.* The Journal of Physical Chemistry B, 2009. **113**(26): p. 9004-9015.
90. Li, D.-W. and R. Brüschweiler, *Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins.* Journal of Chemical Theory and Computation, 2011: p. null-null.
91. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles.* Journal of chemical theory and computation, 2012. **8**(9): p. 3257-3273.
92. Gellman, S.H. and D.N. Woolfson, *Mini-proteins Trp the light fantastic.* Nat Struct Mol Biol, 2002. **9**(6): p. 408-410.
93. Martin, L., et al., *Engineering Novel Bioactive Mini-Proteins on Natural Scaffolds.* Tetrahedron, 2000. **56**(48): p. 9451-9460.
94. Qiu, L., et al., *Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 microseconds.* Journal of the American Chemical Society, 2002. **124**(44): p. 12952-12953.
95. Wang, M., et al., *Dynamic NMR Line-Shape Analysis Demonstrates that the Villin Headpiece Subdomain Folds on the Microsecond Time Scale.* J. Am. Chem. Soc., 2003. **125**(20): p. 6032-6033.
96. Shakhnovich, E., et al., *Protein folding bottlenecks: A lattice Monte Carlo simulation.* Physical Review Letters, 1991. **67**(12): p. 1665.
97. Tang, Y., et al., *Peptide Models Provide Evidence for Significant Structure in the Denatured State of a Rapidly Folding Protein: The Villin Headpiece Subdomain.* Biochemistry, 2004. **43**(11): p. 3264-3272.
98. Ali, M.H., et al., *X-ray structure analysis of a designed oligomeric miniprotein reveals a discrete quaternary architecture.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(33): p. 12183-12188.
99. Haubertin, D.Y., et al., *Molecular Dynamics Simulations of E. coli MsbA Transmembrane Domain: Formation of a Semipore Structure.* Biophysical Journal, 2006. **91**(7): p. 2517-2531.
100. Lazaridis, T., *Effective energy function for proteins in lipid membranes.* Proteins: Structure, Function, and Bioinformatics, 2003. **52**(2): p. 176-192.
101. Jones, Peter M. and Anthony M. George, *Molecular-Dynamics Simulations of the ATP/apo State of a Multidrug ATP-Binding Cassette Transporter Provide a Structural and Mechanistic Basis for the Asymmetric Occluded State.* Biophysical Journal, 2011. **100**(12): p. 3025-3034.
102. St-Pierre, J.-F., et al., *Molecular Dynamics Simulations of the Bacterial ABC Transporter SAV1866 in the Closed Form.* The Journal of Physical Chemistry B, 2012. **116**(9): p. 2934-2942.
103. Allinger, N.L., *Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms.* Journal of the American Chemical Society, 1977. **99**(25): p. 8127-8134.
104. Allinger, N.L., Y.H. Yuh, and J.H. Lii, *Molecular mechanics. The MM3 force field for hydrocarbons. 1.* Journal of the American Chemical Society, 1989. **111**(23): p. 8551-8566.
105. Lifson, S. and A. Warshel, *Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules.* The Journal of Chemical Physics, 1968. **49**(11): p. 5116-5129.
106. Nemethy, G., Gibson, K. D., Palmer, K. A., Chang, N. Y., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H. A., *Energy Parameters in Polypeptides. 10. Improved Geometrical Parameters and Nonbonded Interactions for Use in the ECEPP/3 Algorithm, with Application to Proline-Containing Peptides.* Journal of Physical Chemistry, 1992. **96**: p. 6472-6484.
107. Nemethy, G., Pottle M. S. and Scheraga, H. A., *Energy Parameters in Polypeptides. 9. Updating the Geometrical Parameters, Nonbonded Interactions, and Hydrogen Bond Interactions for the Naturally Occurring Amino Acids.* Journal of Physical Chemistry, 1983. **87**: p. 1883-1887.

108. Dunfield, L.G., Burguess, A. W. and Scheraga H. A., *Energy Parameters in Polypeptides. 8. Empirical Potential Energy Algorithm for the Conformational Analysis of Large Molecules.* Journal of Physical Chemistry, 1978. **82**(24): p. 2609-2616.
109. Momany, F.A., et al., *Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids.* The Journal of Physical Chemistry, 1975. **79**(22): p. 2361-2381.
110. McGuire, R.F.M., F. A. and Scheraga H. A., *Energy Parameters in Polypeptides. V. An Empirical Bond Potential Function Based on Molecular Orbital Calculations.* Journal of Physical Chemistry, 1972. **76**(3): p. 375-393.
111. Scheraga, H.A., et al., *Energy parameters in polypeptides. IV. Semiempirical molecular orbital calculations of conformational dependence of energy and partial charge in di- and tripeptides.* The Journal of Physical Chemistry, 1971. **75**(15): p. 2286-2297.
112. Momany, F.A., et al., *Energy parameters in polypeptides. III. Semiempirical molecular orbital calculations for hydrogen-bonded model peptides.* The Journal of Physical Chemistry, 1970. **74**(12): p. 2424-2438.
113. Yan, J.F., Momany, F. A., Hoffman, R and Scheraga, H. A., *Energy Parameters in Polypeptides II. Semiempirical Molecular Orbital Calculations for MOdel Peptides.* Journal of Physical Chemistry, 1970. **74**(2): p. 420-433.
114. Poland, D.a.S., H., *Energy Parameters in Polypeptides. I. Charge Distributions and the Hydrogen Bond.* Journal of the American Chemical Society, 1967. **6**(12): p. 3791-3800.
115. Lewis, P.N., Momany, F. A. & Scheraga, H. A. (1973), Biochim. Biophys. Acta, 1973. **303**: p. 211-229.
116. Ponder, J.W. and D.A. Case, *Force Fields for Protein Simulations*, in *Advances in Protein Chemistry*, D. Valerie, Editor. 2003, Academic Press. p. 27-85.
117. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
118. van Gusteren, W.F. and H. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual.* 1987.
119. David A. Case, et al., *The Amber biomolecular simulation programs.* Journal of Computational Chemistry, 2005. **26**(16): p. 1668-1688.
120. Pitzer, K.S., *Potential energies for rotation about single bonds.* Discussions of the Faraday Society, 1951. **10**(0): p. 66-73.
121. Lennard-Jones, J. and J.A. Pople, *A survey of the principles determining the structure and properties of molecules. Part 1.-The factors responsible for molecular shape and bond energies.* Discussions of the Faraday Society, 1951. **10**(0): p. 9-18.
122. Lennard-Jones, J. and G.G. Hall, *A survey of the principles determining the structure and properties of molecules. Part II.-The ionization potentials and resonance energies of hydrocarbons.* Discussions of the Faraday Society, 1951. **10**(0): p. 18-26.
123. Weiner, P.K. and P.A. Kollman, *AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions.* Journal of Computational Chemistry, 1981. **2**(3): p. 287-303.
124. Pearlman, D.A., 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**(1-3): p. 1-41.
125. Case, D.A.C., T. E., et al., *The Amber biomolecular simulation programs.* J. Computat. Chem., 2005. **26**: p. 1668-1688.
126. Hagler, A.T., E. Huler, and S. Lifson, *Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals.* Journal of the American Chemical Society, 1974. **96**(17): p. 5319-5327.

127. Hagler, A.T. and S. Lifson, *Energy functions for peptides and proteins. II. Amide hydrogen bond and calculation of amide crystal properties.* Journal of the American Chemical Society, 1974. **96**(17): p. 5327-5335.
128. Allinger, N.L., *Calculation of Molecular Structure and Energy by Force-Field Methods*, in *Advances in Physical Organic Chemistry*, V. Gold and D. Bethell, Editors. 1976, Academic Press. p. 1-82.
129. Weiner, S.J., et al., *A new force field for molecular mechanical simulation of nucleic acids and proteins.* Journal of the American Chemical Society, 1984. **106**(3): p. 765-784.
130. Jorgensen, W.L., *Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water.* Journal of the American Chemical Society, 1981. **103**(2): p. 335-340.
131. Williams, D.E., Reviews in Computational Chemistry, 1991. **2**: p. 219-271.
132. Williams, D.E., *Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals.* Journal of Computational Chemistry, 2001. **22**(11): p. 1154-1166.
133. Weiner, S.J., et al., *An all atom force field for simulations of proteins and nucleic acids.* Journal of Computational Chemistry, 1986. **7**(2): p. 230-252.
134. Wendy D. Cornell, P.C., Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* J. Am. Chem. Soc., 1995. **117**(19): p. 5179-5197.
135. Jorgensen, W.L.C., J.; Madura, J. D.; Impey, R. W.; Klein, M. L. , *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys., 1983. **79**: p. 926-935.
136. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model.* The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
137. Cornell, W.D., et al., *Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation.* Journal of the American Chemical Society, 1993. **115**(21): p. 9620-9631.
138. Beachy, M.D., et al., *Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields.* Journal of the American Chemical Society, 1997. **119**(25): p. 5908-5920.
139. Ösapay, K., et al., *Dielectric Continuum Models for Hydration Effects on Peptide Conformational Transitions.* The Journal of Physical Chemistry, 1996. **100**(7): p. 2698-2705.
140. Okur, A., et al., *Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins.* Journal of Computational Chemistry, 2003. **24**(1): p. 21-31.
141. García, A.E. and K.Y. Sanbonmatsu, *α-Helical stabilization by side chain shielding of backbone hydrogen bonds.* Proceedings of the National Academy of Sciences, 2002. **99**(5): p. 2782-2787.
142. Viktor Hornak, R.A., Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling, *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins: Structure, Function, and Bioinformatics, 2006. **65**(3): p. 712-725.
143. Pendley, S.S., Y.B. Yu, and T.E. Cheatham, *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**(3): p. 612-629.
144. Kollman P, D.R., Cornell W, Fox T, Chipot C, Pohorille A, *The development/application of the ''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 simulations of biomolecular systems, 1997. **3**: p. 83-96.
145. Sorin, E.J. and V.S. Pande, *Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations.* Biophysical Journal, 2005. **88**(4): p. 2472-2493.

146. Demarest, S.J., Y. Hua, and D.P. Raleigh, *Local Interactions Drive the Formation of Nonnative Structure in the Denatured State of Human α-Lactalbumin: A High Resolution Structural Characterization of a Peptide Model in Aqueous Solution†,‡.* Biochemistry, 1999. **38**(22): p. 7380-7387.
147. Cochran, A.G., N.J. Skelton, and M.A. Starovasnik, *Tryptophan zippers: Stable, monomeric β-hairpins.* Proceedings of the National Academy of Sciences, 2001. **98**(10): p. 5578-5583.
148. Mackerell, A.D., *Empirical force fields for biological macromolecules: Overview and issues.* Journal of Computational Chemistry, 2004. **25**(13): p. 1584-1604.
149. Lovell, S.C., et al., *Structure validation by Cα geometry: ϕ,ψ and Cβ deviation.* Proteins: Structure, Function, and Bioinformatics, 2003. **50**(3): p. 437-450.
150. 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. 11258-11259.
151. Wickstrom, L., A. Okur, and C. Simmerling, *Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data.* Biophysical Journal, 2009. **97**(3): p. 853-856.
152. Best, R.B., N.-V. Buchete, and G. Hummer, *Are Current Molecular Dynamics Force Fields too Helical?* Biophysical Journal, 2008. **95**(1): p. L07-L09.
153. *Correction.* Biophysical Journal, 2008. **95**(9): p. 4494.
154. Nerenberg, P.S. and T. Head-Gordon, *Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides.* Journal of Chemical Theory and Computation, 2011. **7**(4): p. 1220-1230.
155. Li, D.-W. and R. Brüschweiler, *NMR-Based Protein Potentials.* Angewandte Chemie International Edition. **49**(38): p. 6778-6780.
156. Sakae, Y. and Y. Okamoto, *Optimizations of force-field parameters for protein systems with the secondary-structure stability and instability.* arXiv.org, 2013. **v1-physics**(1301.1169).
157. Chapman, D.E., J.K. Steck, and P.S. Nerenberg, *Optimizing Protein–Protein van der Waals Interactions for the AMBER ff9x/ff12 Force Field.* Journal of Chemical Theory and Computation, 2013. **10**(1): p. 273-281.
158. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins: Structure, Function, and Bioinformatics. **78**(8): p. 1950-1958.
159. Best, R.B. and J. Mittal, *Free energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences.* Proteins: Structure, Function, and Bioinformatics: p. n/a-n/a.
160. Best, R.B. and J. Mittal, *Balance between α and β Structures in Ab Initio Protein Folding.* The Journal of Physical Chemistry B, 2010. **114**(26): p. 8790-8798.
161. Mittal, J. and R.B. Best, *Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water.* Biophysical Journal. **99**(3): p. L26-L28.
162. Nerenberg, P.S., et al., *Evaluation and Improvement of the Amber ff99SB Force Field with an Advanced Water Model.* Biophysical Journal. **100**(3, Supplement 1): p. 311a-311a.
163. Alexander, D.M., Jr., *Empirical force fields for biological macromolecules: Overview and issues.* Journal of Computational Chemistry, 2004. **25**(13): p. 1584-1604.
164. Beauchamp, K.A., et al., *Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements.* Journal of Chemical Theory and Computation, 2012. **8**(4): p. 1409-1414.
165. Freddolino, P.L., et al., *Force Field Bias in Protein Folding Simulations.* Biophysical Journal, 2009. **96**(9): p. 3772-3780.
166. Kortemme, T., A.V. Morozov, and D. Baker, *An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein–Protein Complexes.* Journal of Molecular Biology, 2003. **326**(4): p. 1239-1259.
167. Kun Song, J.S., Matthew Fesinmeyer, Niels H. Andersen and Carlos Simmerling, *Structural insights for designed alanine-rich helices: Comparing NMR helicity measures and conformational ensembles from molecular dynamics simulation.* Biopolymers, 2008. **89**(9): p. 747-760.

168. Wang, D., et al., *Evaluation of Biologically Relevant Short alpha-Helices Stabilized by a Main-Chain Hydrogen-Bond Surrogate.* Journal of the American Chemical Society, 2006. **128**(28): p. 9248-9256.
169. Xu, L., et al., *Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models.* The Journal of Physical Chemistry B, 2013. **117**(28): p. 8408-8421.
170. Cino, E.A., W.-Y. Choy, and M. Karttunen, *Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations.* Journal of Chemical Theory and Computation, 2012. **8**(8): p. 2725-2740.
171. Patapati, Kalliopi K. and Nicholas M. Glykos, *Three Force Fields' Views of the 310 Helix.* Biophysical journal, 2011. **101**(7): p. 1766-1771.
172. Graf, J., et al., *Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study.* Journal of the American Chemical Society, 2007. **129**(5): p. 1179-1189.
173. Nerenberg, P.S. and T. Head-Gordon, *Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides.* Journal of Chemical Theory and Computation: p. 1220-1230.
174. Karplus, M., *Vicinal Proton Coupling in Nuclear Magnetic Resonance.* Journal of the American Chemical Society, 1963. **85**(18): p. 2870-2871.
175. Hu, J.-S. and A. Bax, *Determination of φ and χ1 Angles in Proteins from 13C−13C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond?* Journal of the American Chemical Society, 1997. **119**(27): p. 6360-6368.
176. Pérez, C., et al., *Self-Consistent Karplus Parametrization of 3J Couplings Depending on the Polypeptide Side-Chain Torsion χ1.* Journal of the American Chemical Society, 2001. **123**(29): p. 7081-7093.
177. Pardi, A., M. Billeter, and K. Wüthrich, *Calibration of the angular dependence of the amide proton-Cα proton coupling constants, 3JHNα, in a globular protein: Use of 3JHNα for identification of helical secondary structure.* Journal of Molecular Biology, 1984. **180**(3): p. 741-751.
178. Hu, J.-S. and A. Bax, *Measurement of Three-Bond 13Ca-13C J Couplings between Carbonyl and Carbonyl/Carboxyl Carbons in Isotopically Enriched Proteins.* Journal of the American Chemical Society, 1996. **118**(34): p. 8170-8171.
179. Wang, A.C. and A. Bax, *Determination of the Backbone Dihedral Angles φ in Human Ubiquitin from Reparametrized Empirical Karplus Equations.* Journal of the American Chemical Society, 1996. **118**(10): p. 2483-2494.
180. Schmidt, J.M., et al., *Self-consistent 3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles.* Journal of Biomolecular NMR, 1999. **14**(1): p. 1-12.
181. Schmidt, J., *Asymmetric Karplus curves for the protein side-chain 3 J couplings.* Journal of Biomolecular NMR, 2007. **37**(4): p. 287-301.
182. Brueschweiler, R. and D.A. Case, *Adding Harmonic Motion to the Karplus Relation for Spin-Spin Coupling.* Journal of the American Chemical Society, 1994. **116**(24): p. 11199-11200.
183. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-2637.
184. Song, K., et al., *Structural insights for designed alanine-rich helices: Comparing NMR helicity measures and conformational ensembles from molecular dynamics simulation.* Biopolymers, 2008. **89**(9): p. 747-760.
185. Honda, S., et al., *10 Residue Folded Peptide Designed by Segment Statistics.* Structure, 2004. **12**(8): p. 1507-1518.
186. Honda, S., et al., *Crystal Structure of a Ten-Amino Acid Protein.* Journal of the American Chemical Society, 2008. **130**(46): p. 15327-15331.

187. Okumura, H., *Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method.* Proteins: Structure, Function, and Bioinformatics, 2012. **80**(10): p. 2397-2416.
188. Astrakas, L.G., C. Gousias, and M. Tzaphlidou, *Structural destabilization of chignolin under the influence of oscillating electric fields.* Journal of Applied Physics, 2012. **111**(7): p. -.
189. Enemark, S., N.A. Kurniawan, and R. Rajagopalan, *β-hairpin forms by rolling up from C-terminal: Topological guidance of early folding dynamics.* Sci. Rep., 2012. **2**.
190. Satoh, D., et al., *Folding free-energy landscape of a 10-residue mini-protein, chignolin.* FEBS letters, 2006. **580**(14): p. 3422-3426.
191. Fesinmeyer, R.M., F.M. Hudson, and N.H. Andersen, *Enhanced Hairpin Stability through Loop Design: The Case of the Protein G B1 Domain Hairpin.* Journal of the American Chemical Society, 2004. **126**(23): p. 7238-7243.
192. Blanco, F.J., G. Rivas, and L. Serrano, *A short linear peptide that folds into a native stable [beta]-hairpin in aqueous solution.* Nat Struct Mol Biol, 1994. **1**(9): p. 584-590.
193. Nguyen, H., D.R. Roe, and C. Simmerling, *Improved Generalized Born Solvent Model Parameters for Protein Simulations.* Journal of Chemical Theory and Computation, 2013. **9**(4): p. 2020-2034.
194. Olsen, K.A., et al., *Hairpin folding rates reflect mutations within and remote from the turn region.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15483-15487.
195. Eker, F., et al., *Tripeptides Adopt Stable Structures in Water. A Combined Polarized Visible Raman, FTIR, and VCD Spectroscopy Study.* Journal of the American Chemical Society, 2002. **124**(48): p. 14330-14341.
196. Eker, F., K. Griebenow, and R. Schweitzer-Stenner, *Stable Conformations of Tripeptides in Aqueous Solution Studied by UV Circular Dichroism Spectroscopy.* Journal of the American Chemical Society, 2003. **125**(27): p. 8178-8185.
197. Urs, S. and F.v.G. Wilfred, *Molecular dynamics simulation of hen egg white lysozyme: A test of the GROMOS96 force field against nuclear magnetic resonance data.* Proteins: Structure, Function, and Genetics, 2000. **40**(1): p. 145-153.
198. Koller, A.N., H. Schwalbe, and H. Gohlke, *Starting Structure Dependence of NMR Order Parameters Derived from MD Simulations: Implications for Judging Force-Field Quality.* Biophysical Journal, 2008. **95**(1): p. L04-L06.
199. Genheden, S., et al., *Starting-Condition Dependence of Order Parameters Derived from Molecular Dynamics Simulations.* Journal of Chemical Theory and Computation, 2010. **6**(7): p. 2176-2190.
200. Cross, A.J. and G.R. Fleming, *Influence of inhibitor binding on the internal motions of lysozyme.* Biophysical Journal, 1986. **50**(3): p. 507-512.
201. Mine, S., et al., *Analysis of the internal motion of free and ligand-bound human lysozyme by use of 15N NMR relaxation measurement: A comparison with those of hen lysozyme.* Protein Science, 2000. **9**(9): p. 1669-1684.
202. Lorieau, J.L. and A.E. McDermott, *Conformational Flexibility of a Microcrystalline Globular Protein: Order Parameters by Solid-State NMR Spectroscopy.* Journal of the American Chemical Society, 2006. **128**(35): p. 11505-11512.
203. Lange, O.F., et al., *Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution.* Science, 2008. **320**(5882): p. 1471-1475.
204. Lee, A. and A.J. Wand, *Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation.* Journal of Biomolecular NMR, 1999. **13**(2): p. 101-112.
205. Lee, A.L., P.F. Flynn, and A.J. Wand, *Comparison of 2H and 13C NMR Relaxation Techniques for the Study of Protein Methyl Group Dynamics in Solution.* Journal of the American Chemical Society, 1999. **121**(12): p. 2891-2902.

206. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.* Journal of the American Chemical Society, 1982. **104**(17): p. 4559-4570.
207. Case, D.A., *NMR lecture Dynamics.* PDF.
208. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity.* Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.
209. Wallach, D., *Effect of Internal Rotation on Angular Correlation Functions.* The Journal of Chemical Physics, 1967. **47**(12): p. 5258-5268.
210. Wittebort, R.J. and A. Szabo, *Theory of NMR relaxation in macromolecules: Restricted diffusion and jump models for multiple internal rotations in amino acid side chains.* The Journal of Chemical Physics, 1978. **69**(4): p. 1722-1736.
211. Brink, D.M. and G.R. Satchler, *"Angular Momentum".* Clarendon Press: Oxford, 1968.
212. Bernstein, M.A., K.F. King, and X.J. Zhou, *Handbook of MRI pulse sequences.* Elsevier Academic Press, 2004: p. 960.
213. Weast, R.C. and M.J. Astle, *Handbook of Chemistry and Physics.* CRC Press, 1982: p. E66.
214. Showalter, S., et al., *Iron Responsive Element RNA Flexibility Described by NMR and Isotropic Reorientational Eigenmode Dynamics.* Journal of Biomolecular NMR, 2005. **32**(3): p. 179-193.
215. Showalter, S.A. and K.B. Hall, *Isotropic Reorientational Eigenmode Dynamics Complements NMR Relaxation Measurements for RNA*, in *Methods in Enzymology*, L.J. Thomas, Editor. 2005, Academic Press. p. 465-480.
216. Lienin, S.F. and R. Brüschweiler, *Characterization of Collective and Anisotropic Reorientational Protein Dynamics.* Physical Review Letters, 2000. **84**(23): p. 5439-5442.
217. Clore, G.M., et al., *Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins.* Journal of the American Chemical Society, 1990. **112**(12): p. 4989-4991.
218. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.* Biochemistry, 1989. **28**(23): p. 8972-8979.
219. Clore, G.M., et al., *Low resolution structure of interleukin-1β in solution derived from 1H・15N heteronuclear three-dimensional nuclear magnetic resonance spectroscopy.* Journal of Molecular Biology, 1990. **214**(4): p. 811-817.
220. Peng, J.W. and G. Wagner, *Mapping of the spectral densities of nitrogen-hydrogen bond motions in Eglin c using heteronuclear relaxation experiments.* Biochemistry, 1992. **31**(36): p. 8571-8586.
221. Woessner, D.E., *Spin Relaxation Processes in a Two-Proton System Undergoing Anisotropic Reorientation.* The Journal of Chemical Physics, 1962. **36**(1): p. 1-4.
222. Daragan, V.A. and K.H. Mayo, *Motional model analyses of protein and peptide dynamics using 13C and 15N NMR relaxation.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1997. **31**(1): p. 63-105.
223. Prompers, J.J. and R. Brüschweiler, *Thermodynamic Interpretation of NMR Relaxation Parameters in Proteins in the Presence of Motional Correlations.* The Journal of Physical Chemistry B, 2000. **104**(47): p. 11416-11424.
224. Prompers, J.J., S.F. Lienin, and R. Bruschweiler, Biocomputing, Proceedings of the 2001 Pacific Symposium, 2001. **Altman, R. B., Dunker, A.K., Hunter, L., Lauderdale, K., Klein, T. E., Eds.;**(World Scientific: Singapore): p. 79-88.
225. Abraham, A., *Principles of Nuclear Magnetism.* Clarendon Press: Oxford, 1961.
226. Wangsness, R.K. and F. Bloch, *The Dynamical Theory of Nuclear Induction.* Physical Review, 1953. **89**(4): p. 728-739.
227. Redfield, A.G., *On the Theory of Relaxation Processes.* IBM Journal of Research and Development, 1957. **1**(1): p. 19-31.

228. Redfield, A.G., *The Theory of Relaxation Processes*, in *Advances in Magnetic and Optical Resonance*, S.W. John, Editor. 1965, Academic Press. p. 1-32.
229. Prompers, J.J. and R. Bruschweiler, *Reorientational Eigenmode Dynamics: A Combined MD/NMR Relaxation Analysis Method for Flexible Parts in Globular Proteins.* Journal of the American Chemical Society, 2001. **123**(30): p. 7305-7313.
230. Prompers, J.J. and R. Brüschweiler, *General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation.* Journal of the American Chemical Society, 2002. **124**(16): p. 4522-4534.
231. Musselman, C., H.M. Al-Hashimi, and I. Andricioaei, *iRED Analysis of TAR RNA Reveals Motional Coupling, Long-Range Correlations, and a Dynamical Hinge.* Biophysical Journal, 2007. **93**(2): p. 411-422.
232. Van Gusteren, W.F., et al., *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Vdf Hochschulverlag AG and der ETH Zurich, 1996. **Zurich**(Switzerland): p. 1-1042.
233. Horn, H.W., et al., *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew.* The Journal of Chemical Physics, 2004. **120**(20): p. 9665-9678.
234. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(14): p. 5286-5290.
235. Case, D.A., C. Scheurer, and R. Bruschweiler, *Static and Dynamic Effects on Vicinal Scalar J Couplings in Proteins and Peptides: A MD/DFT Analysis.* Journal of the American Chemical Society, 2000. **122**(42): p. 10390-10397.
236. Wirmer, J. and H. Schwalbe, *Angular dependence of 1J(Ni,Cα i) and 2J(Ni,Cα (i − 1)) coupling constants measured in J-modulated HSQCs.* Journal of Biomolecular NMR, 2002. **23**(1): p. 47-55.
237. Hennig, M., et al., *Determination of ψ Torsion Angle Restraints from 3J(Cα,Cα) and 3J(Cα,HN) Coupling Constants in Proteins.* Journal of the American Chemical Society, 2000. **122**(26): p. 6268-6277.
238. Holinger, E.P., T. Chittenden, and R.J. Lutz, *Bak BH3 Peptides Antagonize Bcl-xL Function and Induce Apoptosis through Cytochrome c-independent Activation of Caspases.* Journal of Biological Chemistry, 1999. **274**(19): p. 13298-13304.
239. Chittenden, T., et al., *A conserved domain in Bak, distinct from BH1 and BH2, mediates cell death and protein binding functions.* The EMBO Journal, 1995. **15**: p. 14.
240. Lifson, S. and A. Roig, *On the Theory of Helix—Coil Transition in Polypeptides.* The Journal of Chemical Physics, 1961. **34**(6): p. 1963-1974.
241. Zimm, B.H. and J.K. Bragg, *Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains.* The Journal of Chemical Physics, 1959. **31**(2): p. 526-535.
242. Rohl, C.A., A. Chakrabartty, and R.L. Baldwin, *Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol.* Protein Science, 1996. **5**(12): p. 2623-2637.
243. Yang, J., et al., *α-Helix Nucleation Constant in Copolypeptides of Alanine and Ornithine or Lysine.* Journal of the American Chemical Society, 1998. **120**(41): p. 10646-10652.
244. Wang, D., W. Liao, and P.S. Arora, *Enhanced Metabolic Stability and Protein-Binding Properties of Artificial α Helices Derived from a Hydrogen-Bond Surrogate: Application to Bcl-xL.* Angewandte Chemie International Edition, 2005. **44**(40): p. 6525-6529.
245. Kushal, S., et al., *Protein domain mimetics as in vivo modulators of hypoxia-inducible factor signaling.* Proceedings of the National Academy of Sciences, 2013.
246. Wang, D., M. Lu, and P.S. Arora, *Inhibition of HIV-1 Fusion by Hydrogen-Bond-Surrogate-Based α Helices.* Angewandte Chemie International Edition, 2008. **47**(10): p. 1879-1882.
247. Henchey, L.K., et al., *Inhibition of Hypoxia Inducible Factor 1—Transcription Coactivator Interaction by a Hydrogen Bond Surrogate α-Helix.* Journal of the American Chemical Society, 2009. **132**(3): p. 941-943.
248. Henchey, L.K., et al., *High Specificity in Protein Recognition by Hydrogen-Bond-Surrogate α-Helices: Selective Inhibition of the p53/MDM2 Complex.* ChemBioChem, 2010. **11**(15): p. 2104-2107.

249. Hansmann, U., *Parallel tempering algorithm for conformational studies of biological molecules.* Chemical Physics Letters, 1997. **281**(1-3): p. 140-150.
250. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding.* Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.
251. Zhou, R., *Replica Exchange Molecular Dynamics Method for Protein Folding Simulation*, in *Protein Folding Protocols*, Y. Bai and R. Nussinov, Editors. 2006, Humana Press. p. 205-223.
252. Chapman, R.N., G. Dimartino, and P.S. Arora, *A Highly Stable Short α-Helix Constrained by a Main-Chain Hydrogen-Bond Surrogate.* Journal of the American Chemical Society, 2004. **126**(39): p. 12252-12253.
253. Bao, J., et al., *Dynamical Binding of Hydrogen-Bond Surrogate Derived Bak Helices to Antiapoptotic Protein Bcl-xL.* The Journal of Physical Chemistry B, 2009. **113**(11): p. 3565-3571.
254. Scheraga, H.A., J.A. Vila, and D.R. Ripoll, *Helix–coil transitions re-visited.* Biophysical Chemistry, 2002. **101–102**(0): p. 255-265.
255. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proceedings of the National Academy of Sciences, 1989. **86**(14): p. 5286-5290.
256. Huang, C.-Y., et al., *Temperature-Dependent Helix−Coil Transition of an Alanine Based Peptide.* Journal of the American Chemical Society, 2001. **123**(38): p. 9235-9238.
257. Doig, A.J. and R.L. Baldwin, *N- and C-capping preferences for all 20 amino acids in α-helical peptides.* Protein Science, 1995. **4**(7): p. 1325-1336.
258. Andersen, N.H. and H. Tong, *Empirical parameterization of a model for predicting peptide helix/coil equilibrium populations.* Protein Science, 1997. **6**(9): p. 1920-1936.
259. Jas, G.S. and K. Kuczera, *Equilibrium Structure and Folding of a Helix-Forming Peptide: Circular Dichroism Measurements and Replica-Exchange Molecular Dynamics Simulations.* Biophysical Journal, 2004. **87**(6): p. 3786-3798.
260. Zhang, W., et al., *Fs-21 Peptides Can Form Both Single Helix and Helix−Turn−Helix.* The Journal of Physical Chemistry B, 2004. **108**(22): p. 7479-7489.
261. Nymeyer, H. and A.E. García, *Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent.* Proceedings of the National Academy of Sciences, 2003. **100**(24): p. 13934-13939.
262. Robustelli, P., K.A. Stafford, and A.G. Palmer, *Interpreting Protein Structural Dynamics from NMR Chemical Shifts.* Journal of the American Chemical Society, 2012. **134**(14): p. 6365-6374.
263. Soares, T.A., et al., *Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme.* Journal of Biomolecular NMR, 2004. **30**(4): p. 407-422.
264. Stocker, U. and W.F. van Gunsteren, *Molecular dynamics simulation of hen egg white lysozyme: A test of the GROMOS96 force field against nuclear magnetic resonance data.* Proteins: Structure, Function, and Bioinformatics, 2000. **40**(1): p. 145-153.
265. Buck, M. and M. Karplus, *Internal and Overall Peptide Group Motion in Proteins: Molecular Dynamics Simulations for Lysozyme Compared with Results from X-ray and NMR Spectroscopy.* Journal of the American Chemical Society, 1999. **121**(41): p. 9645-9658.
266. Buck, M., et al., *Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme.* Biophysical journal, 2006. **90**(4): p. L36-L38.
267. Case, D.A., *Molecular Dynamics and NMR Spin Relaxation in Proteins.* Accounts of Chemical Research, 2001. **35**(6): p. 325-331.
268. Viktor, H., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins: Structure, Function, and Bioinformatics, 2006. **65**(3): p. 712-725.
269. Tjandra, N., et al., *Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation.* Journal of the American Chemical Society, 1995. **117**(50): p. 12562-12566.
270. Linge, J.P. and M. Nilges, *Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation.* Journal of Biomolecular NMR, 1999. **13**(1): p. 51-59.

271. Hatfield, M.P.D., R.F. Murphy, and S.n. Lovas, *The CLN025 Decapeptide Retains a β-Hairpin Conformation in Urea and Guanidinium Chloride.* The Journal of Physical Chemistry B, 2011. **115**(17): p. 4971-4981.
272. Davis, C.M., et al., *Raising the Speed Limit for β-Hairpin Formation.* Journal of the American Chemical Society, 2012. **134**(35): p. 14476-14482.
273. Hatfield, M.P.D., R.F. Murphy, and S. Lovas, *VCD spectroscopic properties of the β-hairpin forming miniprotein CLN025 in various solvents.* Biopolymers, 2010. **93**(5): p. 442-450.
274. Rodriguez, A., et al., *Computational Study of the Free Energy Landscape of the Miniprotein CLN025 in Explicit and Implicit Solvent.* The Journal of Physical Chemistry B, 2011. **115**(6): p. 1440-1449.
275. Zhao, G.-J. and C.-L. Cheng, *Molecular dynamics simulation exploration of unfolding and refolding of a ten-amino acid miniprotein.* Amino Acids, 2012. **43**(2): p. 557-565.
276. Li, W. and S. Takada, *Characterizing Protein Energy Landscape by Self-Learning Multiscale Simulations: Application to a Designed β-Hairpin.* Biophysical Journal, 2010. **99**(9): p. 3029-3037.
277. Henry, E.R., R.B. Best, and W.A. Eaton, *Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations.* Proceedings of the National Academy of Sciences, 2013.
278. Okur, A., et al., *Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir.* J. Chem. Theory Comput., 2007. **3**(2): p. 557-568.
279. Okur, A., et al., *Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model.* Journal of Chemical Theory and Computation, 2006. **2**(2): p. 420-433.
280. Bashford, D.a.C., D., *Generalized Born Models of Macromolecular Solvation Effects.* Annual Reviews of Physical Chemistry, 2000. **51**: p. 129-152.
281. Geney, R., et al., *Investigation of Salt Bridge Stability in a Generalized Born Solvent Model.* J. Chem. Theory Comput., 2006. **2**(1): p. 115-127.
282. Roe, D.R., et al., *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**(7): p. 1846-1857.
283. Onufriev, A., *The Generalized Born Model: its foundation, applications and limitations.* Review, 2010.
284. Wang, T. and R.C. Wade, *Implicit solvent models for flexible protein–protein docking by molecular dynamics simulation.* Proteins: Structure, Function, and Bioinformatics, 2003. **50**(1): p. 158-169.
285. Polydorides, S. and T. Simonson, *Monte carlo simulations of proteins at constant pH with generalized born solvent, flexible sidechains, and an effective dielectric boundary.* Journal of Computational Chemistry, 2013. **34**(31): p. 2742-2756.
286. Fan, H., et al., *Comparative study of generalized Born models: Protein dynamics.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6760-6764.
287. Scarsi, M., J. Apostolakis, and A. Caflisch, *Comparison of a GB Solvation Model with Explicit Solvent Simulations: Potentials of Mean Force and Conformational Preferences of Alanine Dipeptide and 1,2-Dichloroethane.* The Journal of Physical Chemistry B, 1998. **102**(18): p. 3637-3641.
288. Munoz, V., et al., *Folding dynamics and mechanism of [beta]-hairpin formation.* Nature, 1997. **390**(6656): p. 196-199.
289. Honda, S., N. Kobayashi, and E. Munekata, *Thermodynamics of a β-hairpin structure: evidence for cooperative formation of folding nucleus.* Journal of Molecular Biology, 2000. **295**(2): p. 269-278.
290. Watanabe, H., K. Yamasaki, and S. Honda, *Tracing Primordial Protein Evolution Through Structurally Guided Stepwise Segment Elongation.* Journal of Biological Chemistry, 2013.
291. Chen, J., W. Im, and C.L. Brooks, *Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field.* Journal of the American Chemical Society, 2006. **128**(11): p. 3728-3736.

292. Weinstock, D., *Replica Exchange Molecular Dynamics.* Simulations in Computational Biophysics, 2010. **Rutgers the State University of New Jersey**.
293. De Simone, A., et al., *Insights into Stability and Toxicity of Amyloid-Like Oligomers by Replica Exchange Molecular Dynamics Analyses.* Biophysical Journal, 2008. **95**(4): p. 1965-1973.
294. Xavier, P. and E.M. Alan, *Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent.* The Journal of Chemical Physics, 2007. **126**(1): p. 014903.
295. Kokubo, H., T. Tanaka, and Y. Okamoto, *Two-dimensional replica-exchange method for predicting protein–ligand binding structures.* Journal of Computational Chemistry, 2013. **34**(30): p. 2601-2614.
296. Wallqvist, A., et al., *Detecting Native Protein Folds among Large Decoy Sets with the OPLS All-Atom Potential and the Surface Generalized Born Solvent Model*, in *Computational Methods for Protein Folding*. 2002, John Wiley & Sons, Inc. p. 459-486.
297. Dominy, B.N. and C.L. Brooks, *Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids.* The Journal of Physical Chemistry B, 1999. **103**(18): p. 3765-3773.
298. Srinivasan, J., et al., *Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects.* Theoretical Chemistry Accounts, 1999. **101**(6): p. 426-434.
299. Tsui, V. and D.A. Case, *Theory and applications of the generalized born solvation model in macromolecular simulations.* Biopolymers, 2000. **56**(4): p. 275-291.
300. Kadirvelraj, R., et al., *Understanding the bacterial polysaccharide antigenicity of Streptococcus agalactiae versus Streptococcus pneumoniae.* Proceedings of the National Academy of Sciences, 2006. **103**(21): p. 8149-8154.
301. Yuzlenko, O. and T. Lazaridis, *Membrane protein native state discrimination by implicit membrane models.* Journal of Computational Chemistry, 2013. **34**(9): p. 731-738.
302. Richards, F.M., *Areas, Volumes, Packing, and Protein Structure.* Annual Review of Biophysics and Bioengineering, 1977. **6**(1): p. 151-176.
303. Roux, B.t. and T. Simonson, *Implicit solvent models.* Biophysical Chemistry, 1999. **78**(1–2): p. 1-20.
304. Still, W.C., et al., *Semianalytical treatment of solvation for molecular mechanics and dynamics.* Journal of the American Chemical Society, 1990. **112**(16): p. 6127-6129.
305. Edinger, S.R., et al., *Solvation Free Energies of Peptides: Comparison of Approximate Continuum Solvation Models with Accurate Solution of the Poisson−Boltzmann Equation.* The Journal of Physical Chemistry B, 1997. **101**(7): p. 1190-1197.
306. Qiu, D., et al., *The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii.* The Journal of Physical Chemistry A, 1997. **101**(16): p. 3005-3014.
307. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium.* The Journal of Physical Chemistry, 1996. **100**(51): p. 19824-19839.
308. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Pairwise solute descreening of solute charges from a dielectric medium.* Chemical Physics Letters, 1995. **246**(1–2): p. 122-129.
309. Jones, J.E., *On the Determination of Molecular Fields. II. From the Equation of State of a Gas.* Proceedings of the Royal Society of London. Series A, 1924. **106**(738): p. 463-477.
310. Berendsen, H., et al., *MD with coupling to an external bath.* J Phys Chem, 1984. **81**: p. 3684 - 3690.
311. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
312. Bergonzo, C., et al., *A partial nudged elastic band implementation for use with large or explicitly solvated systems.* International Journal of Quantum Chemistry, 2009. **109**(15): p. 3781-3790.
313. Kumar, S., et al., *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* Journal of Computational Chemistry, 1992. **13**(8): p. 1011-1021.

314. Cies-gravela, M., et al., *Synchronization induced by Langevin dynamics.* Physical Review E, 2001. **63**(6): p. 065202.
315. Duke, R.E. and L. Pedersen, *PMEMD 3.* University of North Carolina-Chapel HIll, 2003.
316. Clarke, D.T., et al., *The α-helix folds on the millisecond time scale.* Proceedings of the National Academy of Sciences, 1999. **96**(13): p. 7232-7237.
317. Sauter, C., et al., *Structure of tetragonal hen egg-white lysozyme at 0.94 A from crystals grown by the counter-diffusion method.* Acta Crystallographica Section D, 2001. **57**(8): p. 1119-1126.
318. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature.* The Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.
319. Vijay-Kumar, S., C.E. Bugg, and W.J. Cook, *Structure of ubiquitin refined at 1.8Åresolution.* Journal of Molecular Biology, 1987. **194**(3): p. 531-544.
320. Nerenberg, P.S., et al., *Optimizing Solute–Water van der Waals Interactions To Reproduce Solvation Free Energies.* The Journal of Physical Chemistry B, 2012. **116**(15): p. 4524-4534.
321. Cerutti, D.S., et al., *Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization.* The Journal of Physical Chemistry B, 2013. **117**(8): p. 2328-2338.
322. Betancourt, M.R. and J. Skolnick, *Local Propensities and Statistical Potentials of Backbone Dihedral Angles in Proteins.* Journal of Molecular Biology, 2004. **342**(2): p. 635-649.
323. Salvador, P., I.H. Tsai, and J.J. Dannenberg, *J-coupling constants for a trialanine peptide as a function of dihedral angles calculated by density functional theory over the full Ramachandran space.* Physical Chemistry Chemical Physics, 2011. **13**(39): p. 17484-17493.
324. Ruscio, J.Z. and A. Onufriev, *A Computational Study of Nucleosomal DNA Flexibility.* Biophysical Journal, 2006. **91**(11): p. 4121-4132.
325. Onufriev, A., Bashford, Donald, Case, David A., *Modification of the Generalized Born Model Suitable for Macromolecules.* The Journal of Physical Chemistry B, 2000. **104**(15): p. 3712-3720.
326. Maier, J., C. Martinez, and L. Wickstrom, *A comprehensive revision of AMBER protein dihedral corrections.* in press, 2014.
327. Chen, J. and C.L. Brooks Iii, *Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions.* Physical Chemistry Chemical Physics, 2008. **10**(4): p. 471-481.
328. Levy, R.M., et al., *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**(31): p. 9523-9530.
329. Wagoner, J.A. and N.A. Baker, *Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms.* Proceedings of the National Academy of Sciences, 2006. **103**(22): p. 8331-8336.
330. Chen, J. and C.L. Brooks, *Critical Importance of Length-Scale Dependence in Implicit Modeling of Hydrophobic Interactions.* Journal of the American Chemical Society, 2007. **129**(9): p. 2444-2445.
331. Bondi, A., *van der Waals Volumes and Radii.* The Journal of Physical Ch
emistry, 1964. **68**(3): p. 441-451.

Appendix

Table 27 ppII secondary basin populations for the central residue of Val_3

Exp = 29	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	48.95±3.73	50.45±2.11	47.97±1.50	45.40±0.93	43.69±0.95
Mod1φ	68.59±2.41	69.07±0.84	67.34±1.07	65.64±2.80	59.17±1.13
Mod2φ	63.42±1.01	59.78±1.37	60.55±2.33	60.27±2.67	56.31±0.50
Mod3φ	57.97±0.75	58.21±2.75	57.66±0.40	53.26±0.00	52.50±0.85
Mod4φ	53.90±1.17	51.28±0.17	52.47±0.09	50.40±0.75	48.71±0.63
Mod5φ	52.95±1.83	55.11±4.20	54.20±3.28	54.35±0.65	47.42±1.10

Table 27 ppII secondary basin populations for the central residue of Val_3. Experimental data obtained from fitting to vicinal scalar couplings by Graf et al. Results taken from revised GB neck model simulations. Uncertainties calculated from native and extended simulations

Table 28 β secondary basin populations for the central residue of Val_3

Exp = 52	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	40.17±2.22	39.76±0.03	39.26±0.59	38.82±1.31	38.76±0.69
Mod1φ	22.42±0.13	20.78±0.35	21.85±1.67	20.26±0.30	18.87±0.81
Mod2φ	25.82±0.97	23.95±2.91	24.59±0.30	25.32±0.56	23.29±2.57
Mod3φ	32.91±0.01	28.41±2.24	30.66±1.16	30.22±0.00	26.54±1.51
Mod4φ	36.18±0.51	32.21±2.02	34.69±2.38	34.98±2.33	33.18±0.45
Mod5φ	26.29±2.07	21.27±1.02	22.02±0.26	24.20±1.21	21.57±2.63

Table 28 β secondary basin populations for the central residue of Val_3. Experimental data obtained from fitting to vicinal scalar couplings by Graf et al. Results taken from revised GB neck model simulations. Uncertainties calculated from native and extended simulations

Table 29 α_R basin populations for the central residue of Val_3

Exp = 19	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψi
99SB	6.91±1.31	6.97±2.59	8.83±1.76	11.52±0.01	13.60±1.91
Mod1φ	5.15±2.62	7.85±1.10	7.65±0.74	9.28±3.40	17.55±1.53
Mod2φ	6.59±1.27	13.75±4.17	11.05±2.05	10.40±2.73	16.35±1.84
Mod3φ	5.61±0.90	10.91±0.50	8.64±0.04	10.95±0.00	16.05±2.71
Mod4φ	6.22±0.99	12.33±1.31	9.15±1.72	10.77±0.23	14.10±0.04
Mod5φ	16.02±0.45	20.59±1.21	20.50±1.3	17.73±2.60	26.49±4.24

Table 29 α_R basin populations for the central residue of Val_3. Experimental data obtained from fitting to vicinal scalar couplings by Graf et al. Results taken from revised GB neck model simulations. Uncertainties calculated from native and extended simulations.

Table 30 α_L basin populations for the central residue of Val_3

Exp = 0	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	0.01±0.09	0.03±0.03	0.04±0.05	0.04±0.02	0.02±0.01
Mod1φ	0.01±0.02	0.00±0.01	0.00±0.00	0.08±0.08	0.00±0.03
Mod2φ	0.03±0.00	0.00±0.00	0.04±0.03	0.00±0.03	0.00±0.01
Mod3φ	0.00±0.00	0.00±0.01	0.00±0.02	0.15±0.00	0.15±0.08
Mod4φ	0.01±0.14	0.13±0.07	0.03±0.03	0.00±0.19	0.00±0.00
Mod5φ	0.96±0.13	0.61±1.39	0.00±1.99	0.28±0.67	1.11±0.35

Table 30 α_L basin populations for the central residue of Val_3. Experimental data obtained from fitting to vicinal scalar couplings by Graf et al. Results taken from revised GB neck model simulations.

Table 31 ^3J (H_N, H_α) values for central residue of Val_3

Exp 7.94±0.02	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	7.69±0.02	7.63±0.03	7.68±0.04	7.80±0.00	7.83±0.03
Mod1φ	6.36±0.00	6.34±0.08	6.32±0.02	6.48±0.01	6.53±0.03
Mod2φ	6.87±0.00	6.84±0.01	6.74±0.06	6.88±0.08	6.96±0.01
Mod3φ	6.90±0.00	6.85±0.02	6.84±0.03	6.90±0.01	6.93±0.01
Mod4φ	7.32±0.03	7.32±0.03	7.25±0.02	7.39±0.04	7.50±0.04
Mod5φ	7.11±0.02	7.03±0.08	7.03±0.01	7.11±0.01	7.18±0.00

Table 31 ^3J (H_N, H_α) values for the central residue of Val_3. Uncertainties calculated from native and extended simulations.

Table 32 ^3J (H_N, C_β) values for central residue of Val_3

	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	2.22±0.02	2.19±0.01	2.21±0.00	2.20±0.00	2.21±0.00
Mod1φ	1.83±0.00	1.79±0.00	1.79±0.00	1.83±0.02	1.82±0.00
Mod2φ	1.96±0.00	1.92±0.02	1.94±0.00	1.95±0.00	1.95±0.01
Mod3φ	1.97±0.00	1.92±0.01	1.94±0.01	2.04±0.00	1.99±0.02
Mod4φ	2.10±0.02	2.11±0.01	2.08±0.01	2.10±0.03	2.07±0.00
Mod5φ	3.14±0.17	3.33±0.17	2.98±0.14	3.01±0.07	2.87±0.17

Table 32 ^3J (H_N, C_β) values for the central residue of Val_3. Uncertainties calculated from native and extended simulations

Table 33 ^3J (H_α, C') values for central residue of Val_3

Exp 1.38±0.06	99SB	Mod1ψ	Mod2ψ	Mod3ψ	Mod4ψ
99SB	1.24±0.03	1.28±0.00	1.27±0.00	1.28±0.01	1.26±0.00
Mod1φ	1.72±0.00	1.78±0.00	1.76±0.02	1.75±0.00	1.77±0.01
Mod2φ	1.60±0.01	1.63±0.03	1.64±0.00	1.61±0.01	1.62±0.02
Mod3φ	1.48±0.00	1.56±0.02	1.54±0.01	1.51±0.01	1.57±0.02
Mod4φ	1.38±0.00	1.44±0.02	1.39±0.02	1.41±0.02	1.41±0.01
Mod5φ	1.12±0.03	1.08±0.02	1.08±0.02	1.15±0.00	1.11±0.04

Table 33 ^3J (H_α, C') values for the central residue of Val_3. Uncertainties calculated from native and extended simulations