Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15

Lee-Ping Wang, Keri A. McKiernan, Joe Gomes, Kyle A. Beauchamp, Teresa Head-Gordon, Julia E. Rice, William C. Swope, Todd J. Martínez, and Vijay S. Pande*

ABSTRACT

The increasing availability of high-quality experimental data and first-principles calculations creates opportunities for developing more accurate empirical force fields for simulation of proteins. We developed the AMBER-FB15 protein force field by building a high-quality quantum chemical data set consisting of comprehensive potential energy scans and employing the ForceBalance software package for parameter optimization. The optimized potential surface allows for more significant thermodynamic fluctuations away from local minima. In validation studies where simulation results are compared to experimental measurements, AMBER-FB15 in combination with the updated TIP3P-FB water model predicts equilibrium properties with equivalent accuracy, and temperature dependent properties with significantly improved accuracy, in comparison with published models. We also discuss the effect of changing the protein force field and water model on the simulation results.

INTRODUCTION

Molecular Dynamics (MD) simulations have demonstrated high utility for the functional study of biomolecular systems. The degree of spatial and temporal resolution afforded by this technique allows for atomic-scale analysis of structure, dynamics, and function. In order to achieve time scales relevant to biological processes, a classical interaction potential, or force field, is typically used. Although approximate, modeling of this variety has proven vital to the mechanistic, thermodynamic, and kinetic understanding of biological phenomena including but

* Email: pande@stanford.edu

not limited to enzyme catalysis[1-4], protein folding[5-9], protein-ligand binding,[10-13] and protein conformational change.[14-17] The results of these studies strongly depend on the accuracy of the underlying force field. While there have been noteworthy simulations on protein dynamics using a quantum chemical potential energy surface,[18-20] these are still incapable of realizing dynamics on the biologically relevant timescales (ns and beyond) for molecules of biologically relevant size (200 residues and beyond). Therefore, the development of accurate empirical force fields is of critical importance for computational biomolecular simulation.

The conceptual development of the consistent force field is credited to Lifson who proposed that the interactions between atoms could be described using an energy function and a small set of transferable empirical parameters. In 1967, Lifson, Warshel and Levitt successfully derived and parameterized the first force field.[21] In 1969, this idea was implemented by Levitt with the first computer simulation of a protein.[22]

Important to the future of condensed phase force fields was the development of the Optimized Potentials for Liquid Simulations (OPLS) model proposed by Jorgenson.[23] Here the nonbonded interactions were derived by fitting to experimental thermodynamic properties of organic liquids, a method which inspired parameterization methodologies of the first generation of all-atom protein force fields.

One of the first force fields capable of all-atom simulations of proteins in water is of the AMBER type and referred to as ff94.[24] This model approximates the energy of a system of molecules as a sum of terms including harmonic bonds, harmonic angles, electrostatic interactions, Lennard-Jones repulsion and dispersion interactions, and dihedral energy terms for adjusting the energy profiles of bond rotations. Harmonic bond and angle terms were optimized to reproduce experimental normal mode frequencies by fitting to structural and vibrational

frequency data on small molecule fragments of amino and nucleic acids. The atom-centered point charges were fit using Kollman's RESP method, which aims to reproduce the electrostatic potential of a target molecule to that calculated at the HF/6-31G* quantum level of theory.[25] The Lennard-Jones parameters were fit in order to reproduce densities and enthalpies of vaporization in simulations of organic liquids (as was done for OPLS). The dihedral parameters were fit using relative energies of alanine and glycine dipeptide conformers calculated via quantum mechanical (QM) methods at the MP2/6-31G* level.

The subsequent widely adopted major iterations of the AMBER type force field have carried over the functional form and most parameters from the original ff94 model. These more recent developments focused primarily on improving protein secondary structure representation via the successive refitting of the ff94 dihedral parameters. The torsions in ff94 applied equally to all quartets of atoms around a bond between two atom types; the parameters were fit to a set of experimental small molecule barrier heights. The ff99 force field[26] improved upon this approach by introducing explicit four atom dihedral terms that were fit to a larger set of small molecules, as well as a reference set of alanine tetrapeptide conformers. The ff99SB force field[27] was introduced by Hornak and Simmerling to improve conformational preferences for glycine and address known deficiencies of previous AMBER force fields such as over-stabilization of α-helices.[28] The amino acid backbone dihedrals for glycine and alanine were refit using a grid-based conformational scan of alanine and glycine tetrapeptides. In ff99 and ff99SB, the other protein parameter types were left unmodified from ff94. The ff99SB-ILDN[29] force field of Shaw and coworkers introduced explicit side chain parameters for four specific residue types (isoleucine, leucine, asparate, and asparagine). The explicit side chain parameters were fit to grid-based conformational scans calculated using second-order Møller-Plesset perturbation

theory with the resolution of the identity approximation[30] (RI-MP2) and a correlation-consistent augmented triple zeta basis set[31] (aug-cc-pvTZ), and validated by calculating NMR observables from simulation trajectories and comparing to experiment. The validation studies showed a significant improvement in the agreement between side-chain conformational states observed in simulations and those observed in NMR experiments. Although each successive modifications of the ff99 force field led to further improvements in secondary structure, the temperature dependence of partial folding remained a major limitation for these models.[32]

Today, researchers are looking in the directions of replacing the point charge model carried over from ff94 with new fixed-charge models and non-additive electrostatic potentials that include explicit polarization. The implicitly polarized charge model found in the ff15ipq[33] force fields treats the point charges of a target molecule as a sum of the charges calculated in vacuum and a perturbation of these charges caused by the presence of explicit solvent molecules, and in this way accounts for electrostatic polarization in a nonpolarizable model. The addition of polarizability in the form of Drude particles[34] or induced dipoles[35] produces a more physically realistic model of electrostatic polarization; however, these models incur a significantly greater computational cost which limits the timescales that are accessible compared with fixed-charge models. These new electrostatic models show great promise for improving the accuracy of the protein energy potential. However, protein force fields that incorporate these electrostatic models require refitting of the other bonded and nonbonded parameter types, and they have yet to be tested to the same extent as the RESP model. It is likely that fine-tuning of bonded and van der Waals interactions using high quality *ab initio* data will continue to be an essential part of developing future generations of protein force fields.

In this work our goal is to assess the limits of accuracy that can be attained by fitting intramolecular bond, angle, and dihedral parameters to QM calculations without modifying the functional form and nonbonded parameters, which we expect will complement efforts currently being undertaken to improve the nonbonded part. We systematically explore the modification of bond, angle, and dihedral parameters, taking the ff99SB functional form and parameter set as a starting point. We introduce a new potential energy scanning method to build an improved data set of dipeptide conformations and provide unprecedented coverage of the conformational space. The parameter optimization was done using ForceBalance,[36] an open-source software package designed to enable reproducible and systematic force field development.

The new parameter set is validated by calculating thermodynamic observables from protein simulations and comparing to experiment. We find that the new parameter set performs equally well as the previous models for equilibrium properties, where previous models gave good agreement with experiment, and gives superior performance for temperature dependence, where previous models perform poorly. Our main finding from the parameter re-optimization is that the ff99SB, and related similarly derived models, overestimate the steepness of potential energy basins, which explains why they predict the correct equilibrium structures, but may lead to problems when simulating conformational changes or deviations from these structures as observed in our subsequent validation studies.

Our validation testing includes a comparison of protein force fields combined with four water models: the TIP3P model most widely used in protein simulations, the updated and more accurate TIP3P-FB model[36], and the four-point TIP4P-Ew[37] and TIP4P-FB[36] models. TIP4P-Ew is a four-point water model developed for use with the particle mesh Ewald electrostatics method[38] that is ubiquitous today, and was among the first water models parameterized to

accurately reproduce the temperature dependence of the density.[37] The TIP3P-FB and TIP4P-FB models, developed ten years later, use the same functional form as TIP3P and TIP4P-Ew (respectively) and were systematically parameterized to reproduce the temperature and pressure dependence of a wide range of thermodynamic properties.[36] Despite the advances made in water models over the last two decades, the protein force fields have largely followed historical precedent in that they are developed and tested for use with the TIP3P model, which raises interesting questions of how the simulation accuracy may improve if the water model is changed. Here, our validation studies show that different force field / water model combinations produce widely varying temperature dependence properties of the protein, and combining AMBER-FB15 with TIP3P-FB produces the best agreement with experiment, despite the fact that the protein intermolecular parameters were not optimized. We discuss some interesting patterns in how different water models affect protein stability. We also describe common limitations of all tested models, which include underestimation of the slopes of protein melting curves and overly collapsed denatured state ensembles, highlighting the necessity of improved descriptions of nonbonded interactions.

The force field combination AMBER-FB15/TIP3P-FB is recommended for general-purpose simulations of proteins, particularly in situations where fluctuations away from equilibrium and temperature dependence are expected to play an important role. Additionally, the *ab initio* data set used to parameterize AMBER-FB15 has been made publicly available online, and we expect it to be useful for force field development efforts in the community.

THEORY

AMBER functional form. The AMBER99SB protein force field (abbreviated as A99SB) is the starting point of the parameterization in this work; it consists of the simple and well-known

functional form put forth in AMBER94 (here referred to as the AMBER functional form), the AMBER99 parameter set, and the "SB" correction to the protein backbone dihedral parameters.[27] In the AMBER functional form, the total potential energy of the system is written as a sum of bonded and nonbonded contributions:

$$
\begin{aligned}
E &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} + E_{\text{vdW}} + E_{\text{ele}}, \\
E_{\text{bond}} &= \sum_{i,j \in \text{bonds}} \frac{k_{ij}^{b}}{2}\left(r_{ij} - r_{ij}^{0}\right)^2, \quad E_{\text{angle}} = \sum_{i,j,k \in \text{angles}} \frac{k_{ijk}^{\theta}}{2}\left(\theta_{ijk} - \theta_{ijk}^{0}\right)^2, \\
E_{\text{dihedral}} &= \sum_{i,j,k,l \in \text{dihedrals}} \sum_{n=1}^{6} k_{ijkl,n}^{\phi}\left(1 + \cos\left(n\phi_{ijkl} - \phi_{ijkl,n}^{0}\right)\right), \\
E_{\text{improper}} &= \sum_{i,j,k,l \in \text{impropers}} k_{ijkl,2}^{\phi}\left(1 + \cos\left(2\phi_{ijkl}\right)\right), \\
E_{\text{vdW}} &= \sum_{i,j \in \text{nonbonded}} 4\varepsilon_{ij}\left[-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right] \\
E_{\text{ele}} &= \sum_{i,j \in \text{nonbonded}} \frac{q_i q_j}{r_{ij}},
\end{aligned}
\tag{1}
$$

where i, j, k, l are atomic indices and $r_{ij}, \theta_{ijk}, \phi_{ijkl}$ are functions of the atomic coordinates. The empirical parameters for bonded interactions are denoted as $k_{ij}^{b}, r_{ij}^{0}, k_{ijk}^{\theta}, \theta_{ijk}^{0}, k_{ijkl,n}^{\phi}, \phi_{ijkl,n}^{0}$ and depend on the *atom types* of the atoms involved. This work focuses on optimizing the bonded parameters, in some cases defining new atom types to increase the size of the parameter space.

In the nonbonded interactions involving pairs of atoms separated by 3 or more bonds, the pairwise Lennard-Jones parameters σ_{ij} and ε_{ij} are derived from those of individual atom types (σ_i and ε_i) via the Lorentz-Berthelot combining rules, and the atomic partial charges q_i are defined for each atom in each amino acid. These parameters are not modified in this work. The vdW and electrostatic interactions between pairs of atoms separated by exactly three bonds (i.e.

"1-4 pairs") are reduced by factors of 1.2 and 2.0, respectively; the 1-4 interactions are nominally considered to be a kind of bonded interaction, but they are also not modified in this work.

Reference Data	**# Calcs.**
Energy, gradients of 26 amino acids over (φ, ψ) (incl. ASH, CYM, GLH, HIE, HIP, LYN)	14,971
Energy, gradients of 21 amino acids over (χ_1, χ_2) (excluding ALA, CYM, GLY, PRO, VAL)	12,093
Energy, gradients of CYM, VAL over (φ, χ_1)	1,151
Vibrational frequencies and eigenvectors for 20 amino acids	20
Energy, gradients of MM-optimized structures	1,060

Table 1. Types of parameterization data for AMBER-FB15 force field.

Reference data set. We constructed a database of *ab initio* calculations consisting of single-point energies, nuclear gradients, and vibrational modes calculated for the blocked dipeptides ACE-X-NME containing one amino acid side chain. This database is further supplemented by additional single-point energies and gradients evaluated at optimized geometries using intermediate force field parameter sets, described later. All energy and gradient values in the database were respectively carried out at the RI-MP2/CBS and RI-MP2/aug-cc-pVTZ levels of theory; the calculations were performed in the gas phase. For each dipeptide, a 24x24 grid of structures was generated by constraining the backbone dihedral angles φ and ψ at 15-degree increments and minimizing the energy in the orthogonal degrees of freedom; a second grid for the side chain dihedral angles χ_1 and χ_2 was carried out for all amino acids with a side chain, except for valine and deprotonated cysteine (CYM) where the second grid uses φ and χ_1 instead.

Beyond the simplest dipeptides (i.e. glycine and alanine), the potential energy surfaces contain many local minima that cannot be comprehensively searched using local optimization methods. Moreover, a sequence of constrained geometry optimizations through a full rotation of a dihedral angle may not return to the starting structure, analogous to turning a corkscrew

embedded in a cork. This hysteresis is a consequence of the many orthogonal degrees of freedom that are only locally optimized using the previous structure as the initial guess. Our approach for scanning the potential surface attempts to find the lowest-energy local minimum in the orthogonal degrees of freedom, as they are likely to carry a higher thermodynamic weight in the protein. To this end we developed the following procedure to explore the conformational space using lower levels of theory:

1. Obtain a four-dimensional grid of structures using gas-phase simulated annealing simulations and the AMOEBA13 polarizable force field; the number of grid points was 12, 12, 6, 6 for φ, ψ, χ_1 and χ_2 respectively, giving a total of 5184 points for each amino acid.
2. For each structure on the four-dimensional grid, perform a MP2/6-31+G* geometry optimization with φ, ψ, χ_1 and χ_2 constrained.
3. For two chosen dihedral angles (e.g. φ, ψ), map the four-dimensional grid of structures to the two-dimensional grid and record the structure with the lowest energy, denoted by $q_{\varphi,\psi}, E_{\varphi,\psi}$. Note that after step 2, only one structure for each (φ, ψ) grid point is recorded out of a total of 36.
4. For each 2-D grid point (φ, ψ) containing a new lowest energy structure, initialize four MP2/6-31+G* geometry optimizations with new dihedral angle constraints $\varphi \pm \frac{\pi}{12}, \psi \pm \frac{\pi}{12}$.
5. Repeat steps 3 and 4 until no new lowest-energy structures are found. The end result is a 24x24 grid of structures with a resolution of $\frac{\pi}{12}$.

In each iteration of steps 3 and 4, each grid point with a new lowest-energy structure is used to launch four new geometry optimizations at the neighboring grid points, and the procedure is

carried out recursively until no more lowest-energy structures are found. As a result, each structure at the end of the procedure is minimized over the initial configurations of its four neighboring structures, and this condition is satisfied for the entire surface; the end result is a grid of structures with a continuous energy surface and discontinuities in the geometry. Following this, each structure is re-optimized at the RI-MP2/aug-cc-pVTZ level with the same dihedral constraints. A single RI-MP2/aug-cc-pVQZ calculation at the optimized geometry provides the means to estimate the energy in the MP2/CBS limit using Helgaker's two-point extrapolation.[39] We also carried out a frequency calculation at the RI-MP2/aug-cc-pVTZ level for the overall lowest-energy structure and scaled the frequencies using standard scaling factors.

***Parameter optimization*.** The parameters were optimized using the ForceBalance software package.[36,40,41] ForceBalance provides a framework where the differences between force field predictions and provided reference data are used to construct a weighted least-squares objective function and its derivatives. A regularization term (penalty function) is applied to prevent large parameter deviations where reference data is insufficient or the force field contains linear dependencies. The calculation is fully specified by:

(1) the functional form of the force field, parameter space (i.e. selection of which parameters to optimize and their interdependencies) and initial parameter values,

(2) the *targets* and their weights that contribute to the objective function,

(3) the *prior widths* that constrain the parameter deviations from their initial values, and

(4) the optimization algorithm that minimizes the objective function.

The main advantage of using ForceBalance is that the calculation is precisely specified and systematically carried out, ensuring that the results are reproducible and significantly reducing the effort involved when repeating the calculation with any component added or changed.

We used the functional form and initial parameters from the AMBER99SB force field; the choice of parameter space was decided by exploring the possible combinations of options in tuning the bond, angle, and torsional parameters, as described in the results and discussion section. We also explored defining independent dihedral parameters for amino acid side chains, which goes beyond the flexibility of the original AMBER99SB model. The extension of AMBER using side chain specific parameters has previously been explored in models such as AMBER99SB-ILDN, RSFF2[42] and AMBER14SB.[43]

The objective function is defined as a function of the differences between the force field predictions and the reference data, plus a regularization term that penalizes large parameter deviations from the initial values. The three types of targets and penalty term are combined as:

$$\chi^2 = \sum_{i=1}^{N_{\text{grids}}} \chi^2_{\text{grid};i} + \sum_{i=1}^{N_{\text{AA}}} \chi^2_{\text{vib};i} + 4\sum_{i=1}^{N'_{AA}} \chi^2_{\text{MM}_{\text{opt}};i} + \sum_{i=1}^{N_{\text{prm}}} \left(\frac{\alpha_i - \alpha_i^0}{p_i} \right)^2 . \quad (2)$$

The first term χ^2_{grid} represents the contributions from the energies and gradients evaluated over a two-dimensional dihedral grid:

$$\chi^2_{\text{grid}} = \left\langle w_E \frac{\left(E_{MM} - E_{QM}\right)^2}{\left\langle E^2_{QM} \right\rangle - \left\langle E_{QM} \right\rangle^2} + \frac{w_F}{3N_{\text{atom}}} \frac{\left|F_{MM} - F_{QM}\right|^2}{\left\langle \left|F_{QM}\right|^2 \right\rangle} \right\rangle \quad \textbf{(3)}$$

where E_{MM}, F_{MM} and E_{QM}, F_{QM} represent the energies and forces determined using the MM force field and QM reference set, respectively. The MM and QM energies are both referenced to the structure with the lowest QM energy. Similarly, $w_E = 1.0$ and $w_F = 0.1$ are weighting factors for the energy and force error terms. The angle brackets denote a weighted average over the points,

and the denominators ensure that the objective function has no physical units and the quantities are expressed as relative errors. The weighted average is given as:

$$\left\langle x \right\rangle \equiv \frac{\sum_{i=1}^{N_{pts}} w\left(E_{i,QM}\right) A\left(E_{i,MM} - E_{i,QM}\right) x_i}{\sum_{i=1}^{N_{pts}} w\left(E_{i,QM}\right) A\left(E_{i,MM} - E_{i,QM}\right)} , \tag{4}$$

where the factors $w\left(E\right)$ and $A\left(E_{MM} - E_{QM}\right)$ are given by:

$$w\left(E\right) = \begin{cases} D^{-1}, E \le D \\ \left(D^2 + \left(E - D\right)^2\right)^{-\frac{1}{2}}, E > D \\ 0, E > U \end{cases} \quad ; \quad A\left(\Delta\right) = \begin{cases} 1, \Delta \ge 0 \\ 100, \Delta < 0 \end{cases} \tag{5}$$

$w\left(E\right)$ is a decreasing function of the reference energy above the minimum, plotted in Figure 1. $D = 5$ kcal/mol is the energy threshold below which $w\left(E\right)$ is a constant; above the threshold, $w\left(E\right)$ becomes inversely proportional to the reference energy. $U = 20$ kcal/mol is the upper energy cutoff above which the weight is set to zero. $A\left(E_{MM} - E_{QM}\right)$ depends on the sign of the MM-QM energy difference and heavily penalizes MM energies that are lower than the QM energies. This reflects our experience that the positive and negative errors in the fit result in asymmetric effects on the simulations. Configurations with negative $E_{MM} - E_{QM}$ have a spuriously large thermodynamic weight and are more likely to appear during MM sampling, which could shift the peaks of the distribution and lead to severe errors such as incorrect equilibrium structures. On the other hand, configurations with positive $E_{MM} - E_{QM}$ have a spuriously small weight in the MM ensemble and underestimate portions of the distribution; this

could result in overestimation of barriers and underestimation of fluctuations, which are (in a sense) higher-order errors than incorrect equilibrium averages. Thus, enforcing $E_{MM} - E_{QM}$ to be nonnegative everywhere and using a weight function that decays with E_{QM} forces the fitting errors into the high-energy regions, where we expect the impacts on the thermodynamic properties to be the smallest.

The second term χ^2_{vib} represents the contributions from the vibrational frequencies evaluated over the 20 standard amino acids:

$$\chi^2_{\text{vib}} = \frac{1}{N_{\text{modes}}} \sum_{i=1}^{N_{\text{modes}}} \left(\frac{\nu_{i,MM} - \nu_{i,QM}}{100\,\text{cm}^{-1}} \right)^2 \tag{6}$$

The QM vibrational modes are ordered by increasing frequency, whereas the corresponding MM vibrational mode is chosen to have the largest absolute value of the dot product with the QM vibrational eigenvector. Unlike the energy and gradient calculations, the MM energy is minimized prior to calculating the vibrational modes.

The third term in the objective function addresses the appearance of spurious energy minima in the MM force field in parts of configuration space not covered by the grid of structures. This term consists of energies and gradients evaluated at MM-optimized structures as in Equation (4), but without energy-dependent weights. We fully optimized each structure on the grid using the MM force field without constraints and clustered the structures with a heavy-atom root-mean-square deviation (RMSD) cutoff of 0.1 Angstrom, leading to a small number of cluster centers for each amino acid (< 50). These structures were used to calculate MP2/CBS energies and MP2/aTZ gradients that were added to the objective function. Because updating the force field

parameters changes the MM energy surface and the locations of minima, this cycle can be repeated to eliminate spurious minima that appear for the new parameter set.

Parameter Type	Prior Width
Bond length	0.01 nm
Bond force constant	10^5 kJ mol^{-1} nm^{-2}
Bond angle	5°
Angle force constant	100 kJ mol^{-1} rad^{-2}
Dihedral phase	π rad
Dihedral amplitude	10 kJ mol^{-1}

Table 2. Prior width values for each parameter type.

The fourth term in the objective function is the regularization term that penalizes parameter deviations from their initial values. Since the force field parameters have different physical unit systems, the parameter deviations must be placed on the same footing by rescaling prior to computing the penalty function. The penalty function corresponds to a prior distribution in a Bayesian interpretation, and thus the rescaling factors for parameter deviations are equivalent to the prior widths. The results of the optimization does depend on the choice of prior widths, but in a much less sensitive way compared to the force field parameters themselves.

The objective function was minimized using a variation of the Levenberg trust-radius method[44-47] implemented in ForceBalance. A parameter update $\left(\mathbf{k}_{n+1} - \mathbf{k}_n\right)$ is calculated as:

$$\mathbf{k}_{n+1} - \mathbf{k}_n = \left[\mathbf{H}\big|_{\mathbf{k}_n} + (\lambda - 1)^2 \mathbf{I} \right]^{-1} \mathbf{G}\big|_{\mathbf{k}_n} \tag{7}$$

where $G_i \equiv \frac{\partial}{\partial k_i} \chi^2$ and $H_{ij} \equiv \frac{\partial^2}{\partial k_i \partial k_j} \chi^2$ are the gradient and Hessian matrix of the objective function in parameter space, and λ is a parameter that affects the length of the optimization step. The Hessian is approximated using the Gauss-Newton method. For the calculations in this paper, the objective function is much less expensive to evaluate than its derivatives – so a line search over λ is performed rather than taking an optimization step of a fixed length. ForceBalance uses

the Brent method as implemented in SciPy to perform the line search. The quadratic form of the coefficient $(\lambda-1)^2$ maps all λ values on the real line to nonnegative values and ensures that the line search is well behaved.

COMPUTATIONAL METHODS

The *ab initio* reference calculations were carried out in a workflow involving several software packages. The initial high-dimensional dihedral grid of structures was generated from restrained simulated annealing simulations using the AMOEBA protein force field as implemented in TINKER.[48] The recursive search over the two-dimensional dihedral grids was performed using a Python program that interfaces with the Q-Chem 4.1 quantum chemistry package[49,50] and uses the Work Queue distributed computing library[51] to manage a large number of Q-Chem calculations running in parallel. The calculations of final optimized structures, energies and gradients were performed in Psi4.[52] Frequencies were obtained in Psi4 via numerical differentiation of the analytic gradients.

The parameterization calculations were performed using ForceBalance via an interface to GROMACS 4.6.5,[53] and contained two fundamental types of MM calculations – single-point energy / gradient evaluations, and frequency calculations. In the frequency calculations, the MM energy was fully minimized using the L-BFGS algorithm prior to calculating the Hessian. ForceBalance also uses the Work Queue library to evaluate individual *targets* in parallel, providing a significant speed-up compared to running all of the MM calculations sequentially.

The validation calculations were performed using multiple software packages. The equilibrium sampling simulations initialized from the crystal structure were carried out using GROMACS 4.6.5 running on standard Linux HPC hardware. The analyses of the equilibrium simulations to calculate RMSD from the crystal structure and NMR scalar couplings / chemical

shifts were carried out using the GROMACS analysis tools, the MDTraj trajectory analysis package,[54] and the ShiftX2 chemical shift prediction software[55].

The temperature replica exchange simulations were carried out using the GPU-accelerated version of AMBER14 running on the OLCF Titan supercomputer, and analysis was performed using the cpptraj[56] and MDTraj[57] software packages. The simulations of the denatured state ensemble were carried out on the Open Science Grid (OSG), a distributed computing network that utilizes donated idle CPU cycles from research computing facilities.[58]

RESULTS AND DISCUSSION

Model	Bond	Angle	Dihedrals	Side Chains	*N* (Params)	MUE (kcal/mol)	Objective Function χ^2
	Initial Parameters (A99SB)					2.78	38.0
Prelim 1	No	No	k_φ	No	69	1.69	16.8
Prelim 2	No	No	k_φ, φ_0	No	138	1.65	12.6
Prelim 3	k_b	k_θ	k_φ	No	210	1.34	11.9
A99SB-V	k_b, b_0	k_θ, θ_0	k_φ, φ_0	No	420	1.00	6.0
AMBER-FB15	k_b, b_0	k_θ, θ_0	k_φ, φ_0	k_φ, φ_0	1406	0.80	4.1

Table 3. Optimized objective function (χ^2) values in a preliminary run of ForceBalance using only the energies from the first two rows of Table 1. The results indicate that all parameter types have a significant impact on lowering the objective function. The final two lines are A99SB-V and AMBER-FB15 respectively.

Choice of optimization parameters. In order to assess the significance of optimizing different types of parameters on the final result, we tested several combinations of the following binary choices: (1) including bond and angle parameters, (2) allowing side chain torsions to take on distinct parameter values, and (3) including equilibrium geometry parameters in addition to the force constants and amplitudes. We ran several optimizations using a simplified version of the objective function where only the (φ, ψ) dihedral scans were included (first row of Table 1); the results for different choices of parameters are given in Table 3. Our results show that tuning the backbone dihedral parameters have a significant effect on decreasing the objective function (not surprising since the target data involves scanning the energy over these degrees of freedom).

Perhaps more surprising is the effect of including bond and angle parameters in the optimization; comparing the first and third rows of Table 3 show that including the bond and angle force constants have an effect of lowering the objective function by ~30% compared to using only the dihedral force constants. Allowing the equilibrium geometry parameters to be

optimized results in a further 50% decrease in the objective function as shown in the fifth row of Table 3. Based on these results, we decided to allow all parameter types to vary in our optimizations.

The model named A99SB-V is the optimized result using all of the A99SB bonded parameters and the data in Table 1. After adding some parameters corresponding to alternative protonation states of amino acids, the total number of adjustable parameters in A99SB-V was 434, and the mean unsigned error (MUE) of the potential across all of the dihedral scans was 1.90 kcal/mol. We also developed a variant of this force field where the side chain torsion parameters for different amino acids were all allowed to vary independently; this led to a decrease of 37% in the objective function, but the number of parameters increased greatly to 1406. Because this force field reproduced experimental results more accurately than A99SB-V and several other models in the validation calculations, we named it AMBER-FB15 and recommend it here for broader use.

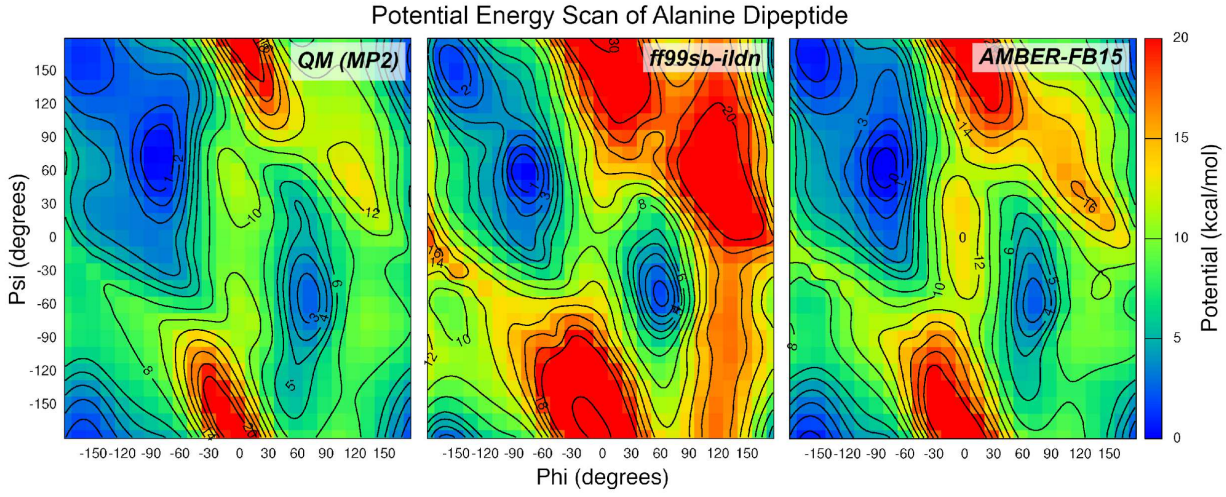

Figure 1. Plot of the potential energy in alanine dipeptide calculated for energy-minimized structures at the MP2/aug-cc-pVTZ level with the (φ,ψ) dihedral angles constrained. Color indicates the relative potential energy with respect to the minimum.

Optimized parameter values. Figures S1-S3 in the Supporting Information show the original and optimized parameter values in AMBER-FB15 grouped by parameter type. The optimized equilibrium bond and angle parameters are all within 5% of their initial values and fall very close to the straight line. Bond and angle force constants show slightly larger deviations; some force constants involving the amide bond are reduced by up to 10% from their initial values. The torsion phases and amplitudes are more widely distributed, largely because the initial guesses for side-chain parameters are set to zero. With few exceptions, the equilibrium torsion phases fall within $\frac{\pi}{6}$ radians 30 degrees of their initial values, and the torsion amplitudes seldom change by more than 4 kJ/mol (1.0 kcal/mol). The largest parameter deviations are observed for arginine and lysine, which possess charged side chains; this is expected due to the especially strong electrostatic interactions in the gas-phase QM calculations, which contributes large terms to the objective function. We note in passing that the usage of gas-phase QM data is most likely to fail for charged systems, but choosing the most appropriate QM method to fit a condensed-phase fixed charge model remains an important challenge.[59]

Quality of fit. Figure 1 shows the potential surface of alanine dipeptide evaluated at the constrained energy minima. As expected, AMBER-FB15 produces a closer fit to the QM energy surface relative to A99SB-ILDN (equivalent to A99SB for alanine). Comparison of the QM and MM surfaces reveals that A99SB-ILDN fits the low-energy regions (in blue) much more accurately than the high-energy regions (in red), and high-energy regions are systematically overpredicted. By contrast, AMBER-FB15 significantly reduces (but does not eliminate) the overprediction of the energy, and the low-energy basins with energy less than 5 kcal/mol above the minimum are significantly broadened. We expect that the broader energy basins in low-

energy regions will lead to larger thermodynamic fluctuations at finite temperatures, which may result in more accurate predictions (as explored in the validation simulations).

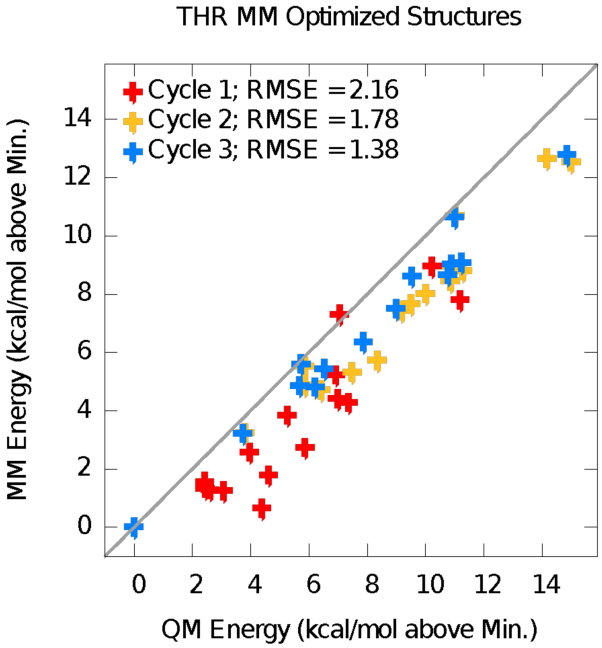

Figure 2. MM vs. QM potential energies for MM-optimized geometries of threonine dipeptide. Each data point corresponds to a local energy minimum predicted by the force field. The Cycle 1 parameters were fitted to QM data from the torsion scans only. The QM data points at the local minima of Cycle 1 are added to the optimization of the Cycle 2 parameters. Cycle 3 is the final parameter set. The spurious MM energy minima (points far below the diagonal line) are eliminated in later cycles.

Figure 2 shows the comparison of QM and MM energies at the local minima of the optimized force field for threonine dipeptide. The initial parameter set (red crosses) predicts the relative energies with a RMS error of 2.16 kcal/mol, and several local minima are within 2 kcal/mol of the lowest-energy structure; by contrast the QM relative energies are significantly higher, ranging between 2 and 6 kcal/mol above the minimum. These local minima with spuriously low relative energies are biased towards higher probability in finite-temperature simulations, which could adversely perturb the equilibrium structure. These local minima are added to the objective function (third term in Equation (2)) to obtain a new set of parameters, which predicts a new set

of local minima with relative energies that match the QM calculations much more closely (yellow crosses). Repeating the addition of local minima to the objective function leads to smaller improvements in the predicted relative energies (blue crosses), and the resulting parameter set is kept as the final version.

Equilibrium properties.

In order to assess the ability of AMBER-FB15 to reproduce equilibrium properties of folded proteins, we ran simulations of 8 proteins: the third IgG-binding domain from streptococcal protein G, abbreviated as *GB3* (PDB ID: 1IGD), *acetyltransferase* from the COG2388 family (2EVN), *lambda repressor* taken from the repressor-operator complex (1LMB), *lysozyme* from bacteriophage lambda (1AM7), N-terminal Domain of Ribosomal Protein L9 or *NTL9* (2HBA), a variant of the *Trp-cage* miniprotein (2JOF), *ubiquitin* (1UBQ), and chicken villin subdomain HP-35 or *villin headpiece* (2F4K). Each protein was simulated at 298.15 K using 7 force fields and 4 water models. For each simulation, the RMSD of the protein backbone to the PDB reference was computed using the residue intervals specified in Table S1, and the RMSD probability density function estimated via a kernel density estimate (KDE). RMSD of the averaged structure was also computed from these simulations. This data is illustrated for three proteins in Figure 3, and the rest are provided in Figure S4. The diamond markers denote the RMSD of the averaged Cartesian coordinates over the whole trajectory. Two protonation states of lysozyme were considered – one state is determined using the pKa values of the amino acids, and the other is determined using the H++ pKa prediction software. Each of the 252 simulations was performed for at least 300 ns with an average trajectory length of 500 ns.

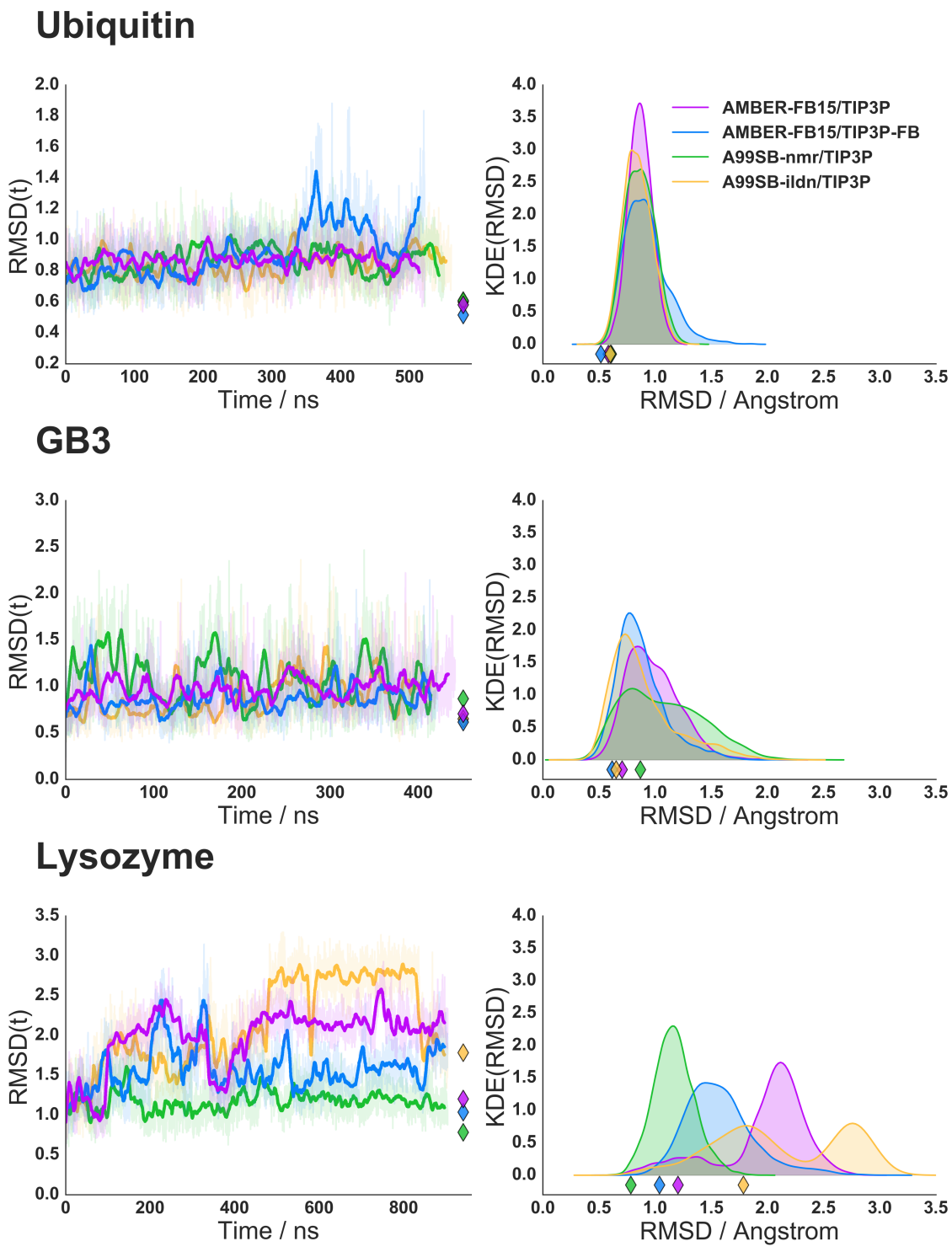

Figure 3. Time series of RMSD for three proteins and four simulations. The diamond markers represent the RMSD of the Cartesian average of the protein backbone conformations.

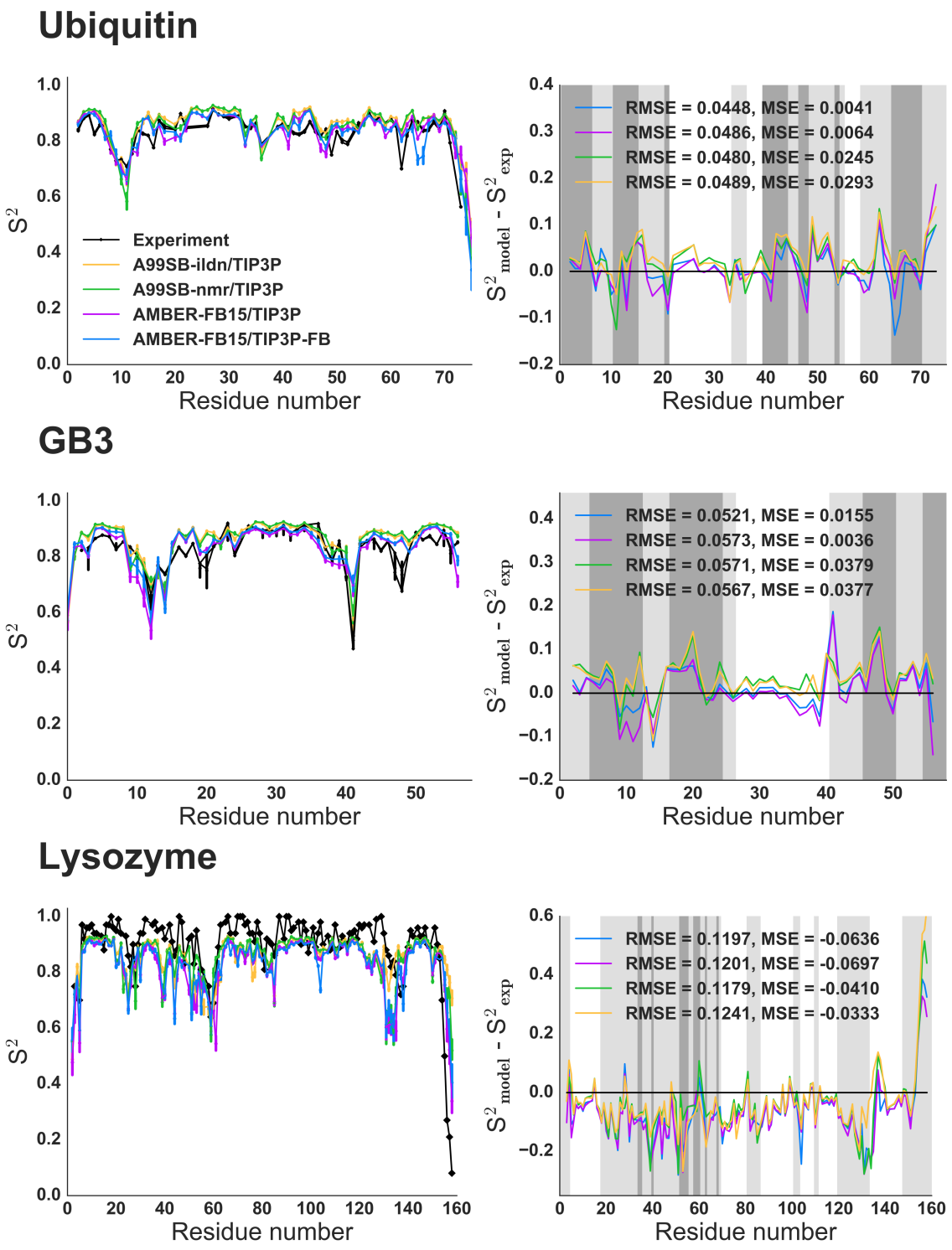

Figure 4. Lipari-Szabo S^2 order parameters and error residuals compared to experimental NMR measurements. The root-mean-squared error (RMSE) and mean signed error (MSE) of the simulated observables with respect to experiment are given in the legends. The background of the error residual plots are colored according to secondary structure as determined by DSSP

analysis. White denotes helix, light gray denotes coil, and dark gray denotes strand secondary structure classification.

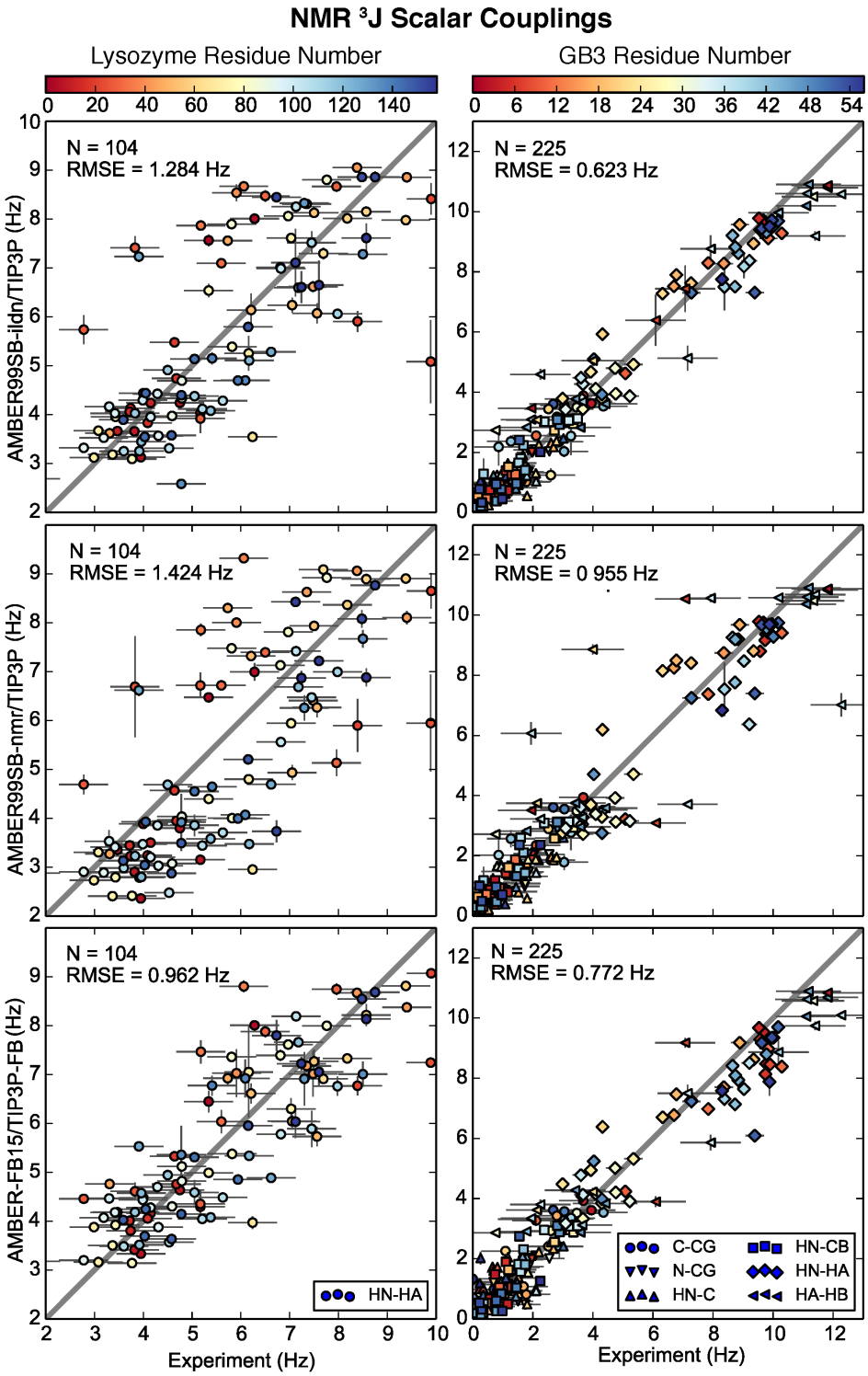

Figure 5. Scatter plots of experimental vs. calculated NMR three-bond scalar couplings. Two proteins are shown (left: bacteriophage lysozyme, PDB ID 1AM7, right: GB3, PDB ID 1IGD) and three models (top, AMBER99SB-ildn/TIP3P; middle, AMBER99SB-nmr/TIP3P; bottom,

AMBER-FB15/TIP3P-FB from this work.) Symbols represent the atom pair involved in the coupling, and colors represent the position of the residue in the protein sequence.

Simulating a protein in water at ambient conditions may not reproduce the crystallographic structure exactly, due to differences in the environment and thermodynamic ensemble. However, crystal structures are often the best structural data available, and it is reasonable to assume that proteins in water stay reasonably close to the crystal structure unless experiments show strong evidence to the contrary. Thus, the RMSD of the simulation trajectory to the crystal structure is routinely considered as an important qualitative validation test of a protein force field, and simulations that deviate significantly from the crystal structure in a short time (i.e. on the sub-microsecond timescale) are interpreted as evidence of force field errors.

Figure 3 shows the RMSD time series for four proteins simulated using four combinations of the protein force field and water model. The protein remains folded in all simulations, and the simulations differ in terms of the overall RMSD to the crystal structure. In the case of ubiquitin, all models have nearly identical RMSD distributions, except for AMBER-FB15/TIP3P-FB which has a small shoulder in the distribution indicating more flexibility in the backbone (also see Supporting Figure S4). In lysozyme and GB3, AMBER-FB15 predicts an RMSD value in between that of A99SB-ILDN and A99SB-NMR; when the water model is changed to TIP3P-FB, the RMSD distribution is shifted to lower values. In all of the simulations except for acetyltransferase, the averaged backbone Cartesian coordinates of the AMBER-FB15/TIP3P-FB remains very close to the crystal structure with a RMSD of 1.0 Å or less. The RMSD distribution for lysozyme is significantly broader than ubiquitin and GB3, and the A99SB-ILDN simulation possesses some bimodal character; this may indicate larger conformational changes on timescales exceeding microseconds that have not been fully sampled in our calculations.

To enrich our understanding of model dependence on equilibrium stability, Lipari-Szabo S2 order parameters were computed for proteins and compared to the available experimental data. Previous studies have shown that simulation lengths exceeding 100 ns are required for accurate estimation of these order parameters,[60] a condition that is satisfied by our calculations. These simulated observables were determined from the trajectories using the isotropic reorientational eigenmode dynamics[61] (iRED) as implemented in the cpptraj program for windows of length 2, 4, and 8 ns. The per-residue deviation from the experimental NMR measurements are shown in Figure 3. These order parameters measure the orientational disorder of the protein backbone N-H vectors on the sub-nanosecond timescale. For all three proteins simulated, we found AMBER-FB15 to produce lower S^2 values by 0.02–0.03 compared to A99SB-ILDN and A99SB-NMR. In the cases of ubiquitin and GB3, AMBER-FB15 predicts significantly lower mean signed errors (MSE), indicating that the increased disorder is consistent with experiment. AMBER-FB15/TIP3P-FB predicts the smallest root mean squared errors (RMSE) for these two proteins. In the case of lysozyme, the experimental measurements have many S^2 values in excess of 0.9, higher than all of the simulated values; here AMBER-FB15 predicts the largest MSE although the RMSE is still very close to those of A99SB-ILDN and A99SB-NMR. An earlier study by Smith and coworkers applied an upper threshold of 0.9 to the experimental order parameters;[62,63] when using this threshold, all of the RMSE values are significantly reduced with AMBER-FB15 producing the lowest error (Figure S5).

The protein structure from equilibrium MD can also be related to NMR experiments using empirical relations to map the three-dimensional structure to the NMR observable. Three-bond J-couplings are often used to compare simulated dihedral angles to experiment; this requires the use of an empirical Karplus relation, which is developed by fitting the crystal structure backbone

and side chain dihedral angles to the NMR observable. The comparison of calculated to experimental NMR observables is an important validation test, but perfect agreement is not expected due to the assumptions and residual errors of the empirical model. Furthermore, because the Karplus relations implicitly include some effects of dynamics in mapping the crystal structure to the solution NMR experiment, using molecular dynamics snapshots as an input to this mapping results in double-counting the effects of dynamics[64,65] which may lead to additional errors.

Figure 5 shows the RMS error of the computed NMR three-bond J-couplings compared to experiments for two proteins, bacteriophage lysozyme and GB3. The recommended model in this work (AMBER-FB15/TIP3P-FB) is compared to A99SB-ildn and A99SB-nmr, both with the TIP3P water model. From examining the left column, the AMBER-FB15/TIP3P-FB model predicts the backbone J-couplings of bacteriophage lysozyme in closer agreement with experiment. The right column shows that A99SB-ildn and AMBER-FB15 both have improved results over A99SB-nmr, which could be explained by the explicit parameterization of side chain torsional potentials. We also calculated J-couplings for two other proteins (ubiquitin and NTL9, Figure S6), and found small differences between the RMSE values compared to experiment on the order of 0.1 – 0.2 Hz. Although the J-couplings shown here were calculated using the Karplus parameters of Ruterjans and coworkers,[66,67] we note that the RMSE values change on the order of 0.1 when using the parameters of Bax and coworkers[68,69] and does not affect the qualitative interpretation of the results.

The NMR chemical shifts on ^{1}H, ^{13}C and ^{15}N can be predicted from MD trajectories using empirical models such as SHIFTX2,[55] which take into account a rather large number of geometric features and fitting parameters to represent the local chemical environment. Similar to

Karplus relations for J-couplings, the chemical shift models are fitted using structural input from crystallography. The RMSE of the predicted chemical shifts are plotted in Figures S7 and S8. We observed that the prediction quality depends heavily on the protein, in contrast to the case of three-bond J-couplings. The RMSE is often within the range of the intrinsic error of SHIFTX2 itself; in an extreme case, the RMSE for ubiquitin is smaller than the SHIFTX2 intrinsic error, which corresponds to χ^2 statistics of less than one and does not reflect the differences between force fields in a meaningful way. From this, we concluded that the chemical shift predictions were insufficient to distinguish AMBER-FB15 from the literature models.

The results in this section show that AMBER-FB15 / TIP3P-FB does not degrade the accuracy of simulating proteins in their native structure at ambient conditions, which is an important validation test for any modern protein force field. Our claim is limited to the systems and time scales studied in this paper, but it lends important credibility to this model for future simulations of interesting biomolecular problems. Furthermore, equilibrium properties are no longer a frontier for protein force field development, with temperature dependence and characterization of the denatured state ensemble being much more important. We will focus our discussion on these important frontiers in the next section.

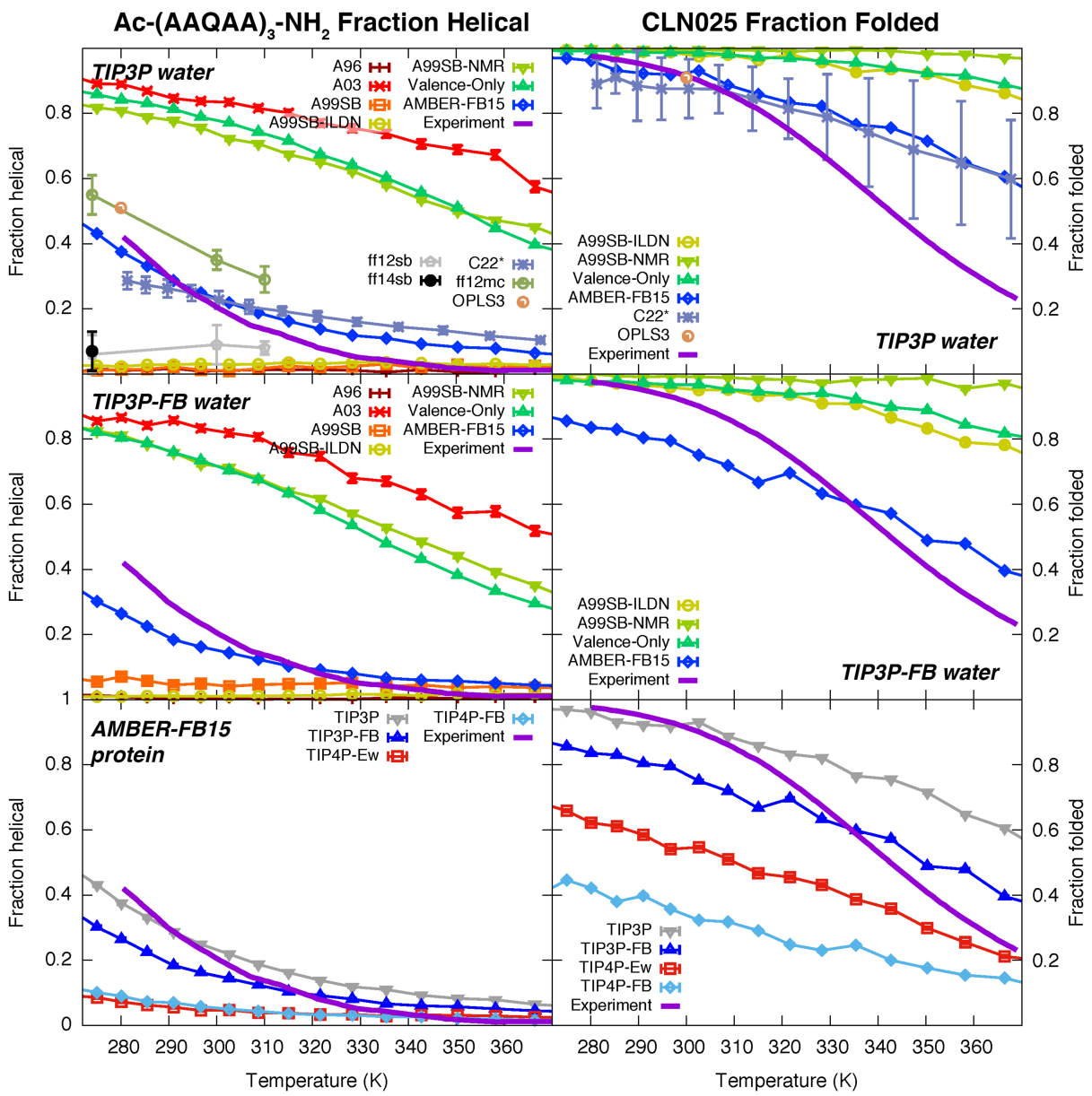

Figure 6. Temperature dependence of secondary structure for two small peptides as a function of temperature and several force field / water model combinations. The performance of the AMBER-FB15 / TIP3P-FB model combination is the dark blue trace in the middle row. Left column: The helical fraction of Ac-(AAQAA)$_3$-NH$_2$. Right column: The fraction folded of CLN025. Top row: Comparison of multiple protein force fields using TIP3P water model. Middle row: Same comparison using TIP3P-FB water model. Bottom row: Comparison of four water models using AMBER-FB15 protein force field.

Temperature dependence. Several of the most popular protein force fields in the past ten years have succeeded at reproducing equilibrium structures of folded proteins but failed to predict an

accurate temperature dependence of the structural ensemble. In previous work, Best and Hummer proposed the A03* and A99SB* models which were directly fitted to reproduce helical fractions at finite temperature;[70] more recently, Wu and coworkers showed improved performance for temperature dependence adding extra 1-5 and 1-6 Lennard-Jones interaction terms and fitting the potentials to experimentally derived free energy distributions.[42] Here we consider the predicted temperature dependence of AMBER-FB15 for two model systems; Ac-$(AAQAA)_3$-NH_2 (abbreviated here as $AAQAA_3$), a 15-residue peptide with partial a-helical character at room temperature, and CLN025, a 10-residue peptide with mostly b-hairpin structure. These two proteins have a significant temperature dependence of the folded fraction in the range 280 – 370 C as measured by circular dichroism[71] and temperature-dependent infrared spectroscopy experiments.[72] The results presented in this section are taken from NVT replica exchange simulations as implemented in AMBER.[73,74]

The left column of Figure 6 shows the temperature dependence of Ac-$(AAQAA)_3$-NH_2 for combinations of protein force fields and water models compared to experiment. Our results for published models show a high degree of consistency compared with existing protein force field validation studies of temperature dependence performed by Lindorff-Larsen and coworkers.[75] The top left and middle left panels compare seven protein force fields using the TIP3P and TIP3P-FB models respectively. The data indicates that protein force fields developed to reproduce equilibrium properties of folded proteins may fail to describe the temperature dependence of partially folded proteins; the A99SB and A99SB-ildn force fields significantly underestimate the α-helical fraction whereas A99SB-nmr significantly overestimates it. The two parameter sets discussed in this paper, A99SB-V and AMBER-FB15, also differ significantly in their temperature dependence. A99SB-V overestimates the helical fraction and behaves similarly

to A99SB-nmr whereas AMBER-FB15 has a temperature dependence mostly consistent with the experiment.

The right column of Figure 6 shows temperature dependence plots for the CLN025 peptide, a small model of a beta hairpin. Due to the high cost of these simulations, we skipped the older models (A96, A03 and A99SB) and compared four protein force fields only. The top right and middle right panels show that A99SB-ildn and A99SB-nmr both overestimate the folded fraction, in contrast to the results for Ac-$(AAQAA)_3$-NH_2 where A99SB-ildn and A99SB-nmr are on either side of the correct result. AMBER-FB15 again comes closest to reproducing the experimental result.

The bottom left panel of Figure 6 compares temperature trends of Ac-$(AAQAA)_3$-NH_2 using the AMBER-FB15 protein force field and four different water models. The choice of water model affects the helical content; the simulations using TIP3P predict the most helical content, followed by TIP3P-FB; the best agreement with experiment is given by TIP3P below 300K and TIP3P-FB above 300K. By contrast, the simulations using TIP4P-Ew and TIP4P-FB predict a much lower helical content. In a similar fashion, the bottom right panel of Figure 6 shows the temperature trends in the folded fraction of CLN025 using the AMBER-FB15 force field and four water models. The TIP3P simulations predict the highest folded fraction, followed by TIP3P-FB, then TIP4P-Ew and TIP4P-FB. Figure S9 shows that using the TIP4P-Ew and TIP4P-FB water models have the effect of decreasing the amount of protein structure for all seven protein force fields. The best overall agreement with experiment is given by the AMBER-FB15/TIP3P-FB simulations.

The effect of changing the water model on peptide stability is an interesting feature of the simulations. Clearly, the accuracy of the protein temperature dependence does not depend

strongly on the accuracy of the water model, as both the TIP4P-Ew and TIP4P-FB models are highly accurate for computing the properties of water. The ability of protein simulations to accurately reproduce temperature dependence with TIP4P-Ew has been shown for model peptides where abundant NMR data is available, requiring changes in only one backbone dihedral parameter.[76] Moreover, the nonbonded protein parameters of the protein were not optimized, which will certainly have a strong effect on the temperature dependence. We expect that improved derivations of point charge models from quantum chemistry calculations[77] and accompanying reparameterization of the Lennard-Jones interaction terms[78] will produce more accurate descriptions of temperature dependence for realistic water models. In light of all these considerations, it is still instructive to search for other trends in the water models that correlate well with the temperature dependence trends observed here.

When CLN025 is simulated with AMBER-FB15 (and when $AAQAA_3$ is simulated with A99SB-V), the helical/folded fraction takes on a wide range of values between 0.1 and 0.9; there is also a clear trend of peptide stability that goes as TIP3P > TIP3P-FB > TIP4P-Ew > TIP4P-FB. We could not find a significant correlation between the peptide stability and the basic properties of the water models, such as the internal energy or magnitude of the dipole moment. On the other hand, the peptide stability was significantly correlated with the average interaction energy between protein and water (Figure 7).

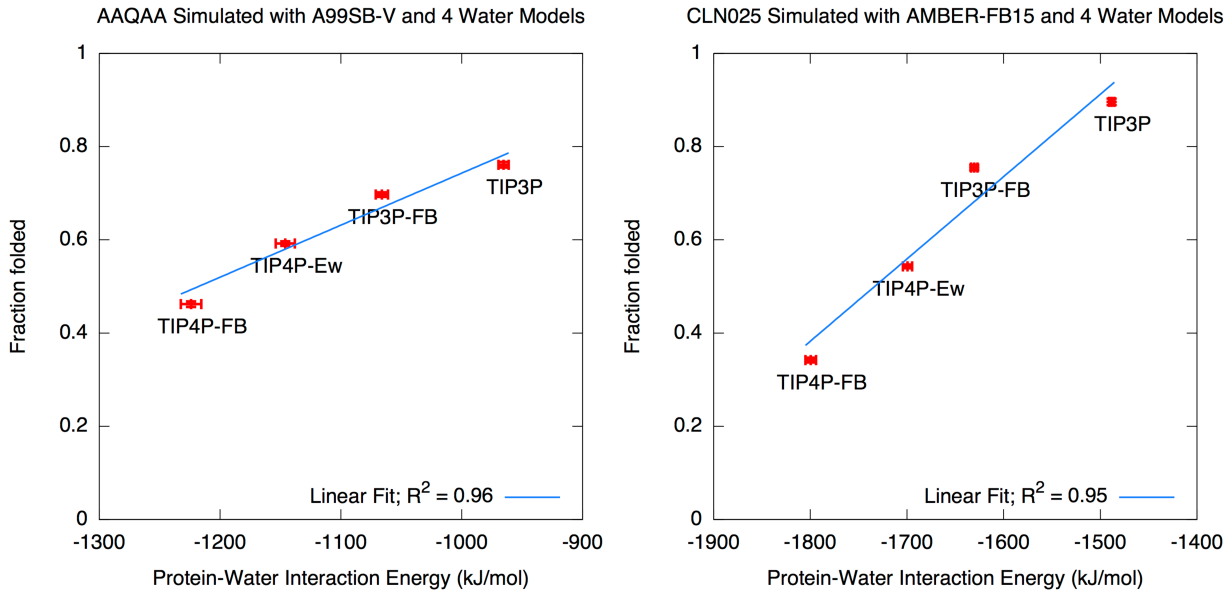

Figure 7. Correlation between average protein-water interaction energy and fraction of secondary structure. Left: AAQAA simulated with A99SB-V (left). Right: CLN025 simulated with AMBER-FB15 (right). Each plot contains four simulations with four water models. Error bars represent one standard error.

We tested the effects of changing the water model "in-place" by replacing the water model in the simulation trajectory, creating a 4x4 grid where the simulation trajectory using model X was used to calculate the protein-water interaction using model Y. We found that (1) changing the water model from TIP3P→TIP3P-FB→TIP4P-EW→TIP4P-FB increased the protein-water interaction strength independent of which trajectory was used, and (2) the conformational ensembles from TIP3P→TIP3P-FB→TIP4P-EW→TIP4P-FB had increasingly strong protein-water interactions independent of which water model was used (Figure S10). Our analysis indicates that having stronger water-protein interactions *causes* proteins to become less stable.

While this is an encouraging sign of progress, we also note that all of the potentials underestimate the slope of the temperature dependence. One possible reason is that the simulated and experimental ensembles are different; the experiment is performed at constant pressure whereas the replica exchange simulations could only be done in the NVT ensemble. If the

simulations had been run in the NPT ensemble instead, the density of water would have decreased at higher temperatures, which may have an effect on the helical fraction. Another possibility is the pairwise additive approximation from the force field, which neglects many-body effects such as those arising from explicit electronic polarization. Including the electronic polarizability may increase the cooperativity of helix formation and lead to a steeper temperature dependence.[79] We intend to apply this parameterization strategy and *ab initio* data set toward the parameterization of a polarizable force field in forthcoming work.

	TIP3P	**TIP3P-FB**	**TIP4P-EW**	**TIP4P-FB**
A96	1.80 (0.10)	**1.99 (0.16)**	**1.92 (0.10)**	1.88 (0.08)
A03	1.42 (0.06)	1.77 (0.15)	1.65 (0.08)	1.79 (0.12)
A99SB	1.69 (0.11)	1.83 (0.14)	1.75 (0.09)	1.84 (0.11)
A99SB-ILDN	1.70 (0.13)	1.77 (0.12)	1.82 (0.11)	**2.01 (0.11)**
A99SB-NMR	1.68 (0.10)	1.82 (0.10)	1.75 (0.11)	1.79 (0.08)
A99SB-V	1.64 (0.08)	**1.91 (0.13)**	1.80 (0.10)	1.75 (0.08)
AMBER-FB15	1.47 (0.07)	1.77 (0.10)	1.75 (0.08)	1.88 (0.10)

Table 4. Average radius of gyration (R_g) of the denatured state ensemble of GB3 simulated using seven protein models and four water models. Bold entries denote average R_g values in excess of 1.9 nm. The experimental measurements are 2.2 nm (FRET) and 2.6 nm (SAXS) from Ref. 79.

Denatured state ensemble. A current frontier in protein simulations is the description of the denatured state ensemble (DSE), a vast conformational space where protein conformations are extended relative to the native state.[80] The DSE is closely connected with intrinsically disordered proteins,[81] which do not possess a well-defined native state and may play important roles in neurological disorders.[82] Experimentally, the average radius of gyration of denatured proteins may be inferred from Forster resonance energy transfer (FRET) and small angle X-ray scattering (SAXS) data.[83]

Here we simulated the DSE of GB3 by first denaturing the protein by running 20 ns simulations at 600 K for all 28 protein / water model combinations; we then extracted five

snapshots at 1 ns intervals from the end of each trajectory, creating 140 initial structures in total. We then launched 10 ns simulations for each of the 140 initial conditions for all 28 protein / water model combinations, a total of 3,960 simulations total. 2,226 of these 3,960 simulations ran to completion, representing about 800 ns of simulation time for each protein / water model combination. Our results are summarized in Table 4, which shows that all tested protein models systematically underestimate the radius of gyration in comparison with experiment. Although the short simulation time of 10 ns is not enough to fully sample this large ensemble,[84] the R_g values demonstrate some significant trends with respect to the water model; for example, the TIP3P simulations predict significantly more compact distributions (on average 0.2 nm smaller than TIP3P-FB). TIP3P-FB, TIP4P-Ew and TIP4P-FB produce average R_g values that are within the margin of statistical error, but they are all significantly less than the experimental values derived from FRET or SAXS measurements, which are 2.2 and 2.6 nm respectively. Thus, we conclude that none of the protein/water model combinations are able to accurately describe the DSE. Efforts to increase the strength of protein/water interactions by increasing the van der Waals ε parameters of water have shown some promising results,[85] though more studies are needed to assess whether this approach applies equally well to the large sequence space of IDPs.[86]

Conclusion

The AMBER-FB15 protein force field combines the well-established model of intermolecular interactions from AMBER94 with a systematic and thorough optimization of the intramolecular terms. The key difference in the optimized result is a significant lowering of the potential in regions away from the energy minima, which is expected to yield greater flexibility in finite temperature simulations. We validated the new model with extensive simulations on multiple proteins; we found that the predictions of equilibrium thermodynamic properties were equivalent in accuracy to published models, and the predictions of temperature dependence were significantly improved. Replacing the TIP3P water model with the updated TIP3P-FB model resulted in overall improved accuracy of the temperature dependence predictions. Supported by the evidence in this paper, we are optimistic that the model combination AMBER-FB15/TIP3P-FB will yield accurate predictions in simulations of proteins, particularly when fluctuations away from equilibrium, conformational changes and/or temperature dependence are expected to play important roles. Our work also highlights the limits of reparameterizing the intramolecular part of the potential, as the predicted properties of the denatured state ensemble are still significantly different from experimental measurements. Future improvement of intermolecular interactions in force fields should focus on improving the description of this important aspect of protein chemistry.

Acknowledgements

We acknowledge Dave Case and Dave Cerutti for insightful discussions. We acknowledge Christina Redfield for sharing the order parameter data for bacteriophage lysozyme. The simulations of protein temperature dependence were performed on the Titan supercomputer at the Oak Ridge Leadership Computing Facility (OLCF). The simulations of the denatured state ensemble used computational resources provided by the Open Science Grid (OSG). We acknowledge Miron Livny and the OSG Team for providing this resource and for their support. KAB acknowledges support from NIH grant P30CA008747, the Sloan Kettering

Institute, and Starr Foundation grant I8-A8-058. We acknowledge members of the Chodera lab for helpful discussions.

Disclosure

KAB is currently an employee of Counsyl, Inc.
VSP is a consultant and SAB member of Schrodinger, LLC and Globavir, sits on the Board of Directors of Omada Health, and is a General Partner at Andreessen Horowitz.

Supporting Information Available. Scatter plots of original and optimized force field parameters for AMBER-FB15; RMSD time series for simulated proteins not shown in the main text; NMR S^2 order parameter calculation for lysozyme with thresholded experimental value; NMR ^3J couplings for ubiquitin and NTL9; summaries of RMS errors in predicted chemical shifts; temperature dependence plots for AAQAA and CLN025 with 4-point water models; additional tables and figures describing protein-water interaction energies.

Bibliography

(1) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry-Us* **1991**, *30*, 5826.
(2) Gao, J. *Current Opinion in Structural Biology* **2003**, *13*, 184.
(3) Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C. M.; Vervoort, J. *J. Am. Chem. Soc.* **2000**, *122*, 8728.
(4) Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. *Proceedings of the National Academy of Sciences* **2000**, *97*, 2017.
(5) Ferrara, P.; Apostolakis, J.; Caflisch, A. *The Journal of Physical Chemistry B* **2000**, *104*, 5000.
(6) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proceedings of the National Academy of Sciences* **1992**, *89*, 8721.
(7) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. *Folding and Design* **1996**, *1*, 441.
(8) Šali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248.
(9) Šali, A.; Shakhnovich, E.; Karplus, M. *Journal of Molecular Biology* **1994**, *235*, 1614.
(10) Gallicchio, E.; Levy, R. M. In *Computational chemistry methods in structural biology*; Elsevier BV: 2011, p 27.
(11) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(12) Marelius, J.; Hansson, T.; Åqvist, J. *Int. J. Quant. Chem.* **1998**, *69*, 77.
(13) Wereszczynski, J.; McCammon, J. A. *Quarterly Reviews of Biophysics* **2011**, *45*, 1.
(14) Böckmann, R. A.; Grubmüller, H. *Nature Structural Biology* **2002**.

(15) Kruger, P. *Protein Science* **2001**, *10*, 798.
(16) Ma, J.; Sigler, P. B.; Xu, Z.; Karplus, M. *Journal of Molecular Biology* **2000**, *302*, 303.
(17) Yang, W.; Gao, Y. Q.; Cui, Q.; Ma, J.; Karplus, M. *Proceedings of the National Academy of Sciences* **2003**, *100*, 874.
(18) Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. *The Journal of Physical Chemistry B* **2012**, *116*, 12501.
(19) Rudberg, E.; Rubensson, E. H.; Sałek, P. *The Journal of Chemical Physics* **2008**, *128*, 184106.
(20) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1789.
(21) Lifson, S. *The Journal of Chemical Physics* **1968**, *49*, 5116.
(22) Levitt, M.; Lifson, S. *Journal of Molecular Biology* **1969**, *46*, 269.
(23) Jorgensen, W. L.; Tirado-Rives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657.
(24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(25) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *The Journal of Physical Chemistry* **1993**, *97*, 10269.
(26) Wang, J.; Cieplak, P.; Kollman, P. A. *Journal of Computational Chemistry* **2000**, *21*, 1049.
(27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712.
(28) Garcia, A. E.; Sanbonmatsu, K. Y. *Proceedings of the National Academy of Sciences* **2002**, *99*, 2782.
(29) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Structure, Function, and Bioinformatics* **2010**, NA.
(30) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chemical Physics Letters* **1993**, *208*, 359.
(31) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J Chem Phys* **1992**, *96*, 6796.
(32) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341.
(33) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. *J. Chem. Theory Comput.* **2016**, *12*, 3926.
(34) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2013**, *9*, 5430.
(35) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, *9*, 4046.
(36) Wang, L.-P.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. Lett.* **2014**, *5*, 1885.
(37) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J Chem Phys* **2004**, *120*, 9665.
(38) Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089.
(39) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J Chem Phys* **1997**, *106*, 9639.
(40) Wang, L.-P.; Chen, J.; Van Voorhis, T. *J. Chem. Theory Comput.* **2013**, *9*, 452.
(41) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *The Journal of Physical Chemistry B* **2013**, *117*, 9956.

(42) Zhou, C. Y.; Jiang, F.; Wu, Y. D. *J Phys Chem B* **2015**, *119*, 1035.
(43) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696.
(44) Dennis, J. E.; Gay, D. M.; Walsh, R. E. *ACM Transactions on Mathematical Software* **1981**, *7*, 348.
(45) Marquardt, D. W. *Journal of the Society for Industrial and Applied Mathematics* **1963**, *11*, 431.
(46) Moré, J. J. In *Lecture Notes in Mathematics*; Springer Nature: 1978, p 105.
(47) Moré, J. J.; Sorensen, D. C. *SIAM Journal on Scientific and Statistical Computing* **1983**, *4*, 553.
(48) Ponder, J. W.; Washington University School of Medicine: St. Louis, MO, 2004; Vol. 2016.
(49) Krylov, A. I.; Gill, P. M. W. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *3*, 317.
(50) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kus, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L., III; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A., Jr.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A. *Molecular Physics* **2015**, *113*, 184.
(51) Albrecht, M.; Rajan, D.; Thain, D. In *2013 IEEE International Conference on Cluster Computing (CLUSTER)*; Institute of Electrical and Electronics Engineers (IEEE): 2013.
(52) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *2*, 556.
(53) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1-2*, 19.
(54) McGibbon, Robert T.; Beauchamp, Kyle A.; Harrigan, Matthew P.; Klein, C.; Swails, Jason M.; Hernández, Carlos X.; Schwantes, Christian R.; Wang, L.-P.; Lane, Thomas J.; Pande, Vijay S. *Biophysical Journal* **2015**, *109*, 1528.
(55) Han, B.; Liu, Y. F.; Ginzinger, S. W.; Wishart, D. S. *Journal of Biomolecular Nmr* **2011**, *50*, 43.
(56) Roe, D. R.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084.

(57) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. *Biophysical Journal* **2015**, *109*, 1528.

(58) Altunay, M.; Avery, P.; Blackburn, K.; Bockelman, B.; Ernst, M.; Fraser, D.; Quick, R.; Gardner, R.; Goasguen, S.; Levshina, T.; Livny, M.; McGee, J.; Olson, D.; Pordes, R.; Potekhin, M.; Rana, A.; Roy, A.; Sehgal, C.; Sfiligoi, I.; Wuerthwein, F.; Open Sci Grid Executive, B. *Journal of Grid Computing* **2011**, *9*, 201.

(59) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. *J Phys Chem B* **2013**, *117*, 2328.

(60) Koller, A. N.; Schwalbe, H.; Gohlke, H. *Biophysical Journal* **2008**, *95*, L4.

(61) Prompers, J. J.; Bruschweiler, R. *Journal of the American Chemical Society* **2002**, *124*, 4522.

(62) Smith, L. J.; Bowen, A. M.; Di Paolo, A.; Matagne, A.; Redfield, C. *Chembiochem* **2013**, *14*, 1780.

(63) Smith, L. J.; van Gunsteren, W. F.; Hansen, N. *Eur. Biophys. J. Biophys. Lett.* **2015**, *44*, 235.

(64) Case, D. A. *Current Opinion in Structural Biology* **2000**, *10*, 197.

(65) Case, D. A.; Scheurer, C.; Bruschweiler, R. *Journal of the American Chemical Society* **2000**, *122*, 10390.

(66) Perez, C.; Lohr, F.; Ruterjans, H.; Schmidt, J. M. *Journal of the American Chemical Society* **2001**, *123*, 7081.

(67) Schmidt, J. M.; Blumel, M.; Lohr, F.; Ruterjans, H. *Journal of Biomolecular Nmr* **1999**, *14*, 1.

(68) Vuister, G. W.; Bax, A. *Journal of the American Chemical Society* **1993**, *115*, 7772.

(69) Vuister, G. W.; Wang, A. C.; Bax, A. *Journal of the American Chemical Society* **1993**, *115*, 5334.

(70) Best, R. B.; Hummer, G. *J Phys Chem B* **2009**, *113*, 9004.

(71) Shalongo, W.; Dugad, L.; Stellwagen, E. *Journal of the American Chemical Society* **1994**, *116*, 8288.

(72) Davis, C. M.; Xiao, S. F.; Raeigh, D. P.; Dyer, R. B. *Journal of the American Chemical Society* **2012**, *134*, 14476.

(73) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668.

(74) Pitera, J. W.; Swope, W. *P Natl Acad Sci USA* **2003**, *100*, 7587.

(75) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *Plos One* **2012**, *7*.

(76) Nerenberg, P. S.; Head-Gordon, T. *Journal of Chemical Theory and Computation* **2011**, *7*, 1220.

(77) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J Phys Chem B* **1998**, *102*, 3586.

(78) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. *J Phys Chem B* **2012**, *116*, 4524.
(79) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. *Chemical Reviews* **2016**, *116*, 4983.
(80) Tompa, P. *Current Opinion in Structural Biology* **2011**, *21*, 419.
(81) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J Am Chem Soc* **2008**, *130*, 6145.
(82) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. In *Annual Review of Biophysics* 2008; Vol. 37, p 215.
(83) Skinner, J. J.; Yu, W.; Gichana, E. K.; Baxa, M. C.; Hinshaw, J. R.; Freed, K. F.; Sosnick, T. R. *P Natl Acad Sci USA* **2014**, *111*, 15975.
(84) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *Journal of the American Chemical Society* **2012**, *134*, 3787.
(85) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J Phys Chem B* **2015**, *119*, 5113.
(86) Henriques, J.; Skepo, M. *Journal of Chemical Theory and Computation* **2016**, *12*, 3407.