Building a More Predictive Protein Force Field: a Systematic and Reproducible Route to AMBER-FB15

Lee-Ping Wang[1], Keri A. McKiernan[2], Joseph Gomes[2], Kyle A. Beauchamp[3], Teresa Head-Gordon[4,5], Julia E. Rice[6], William C. Swope[6], Todd J. Martínez[2,7,8], and Vijay S. Pande[2,9*]

1. Department of Chemistry, University of California, Davis, California 95616, USA.
2. Department of Chemistry, Stanford University, Stanford, California 94305, USA.
3. Counsyl, Inc., South San Francisco, California 94080, USA.
4. Departments of Chemistry, Bioengineering, Chemical and Biomolecular Engineering, and Kenneth S. Pitzer Center for Theoretical Chemistry, University of California, Berkeley, California 94720, USA.
5. Chemical Sciences Division, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
6. IBM Almaden Research Center, IBM Research, San Jose, California 95120, United States
7. PULSE Institute, Stanford University, Stanford, California 94305, USA
8. SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA
9. Departments of Computer Science, Structural Biology, and Program in Biophysics, Stanford University, Stanford, California 94305, USA.

* Email: pande@stanford.edu .

Contents

* Email: pande@stanford.edu

1. Scatter plots of AMBER-FB15 force field parameters

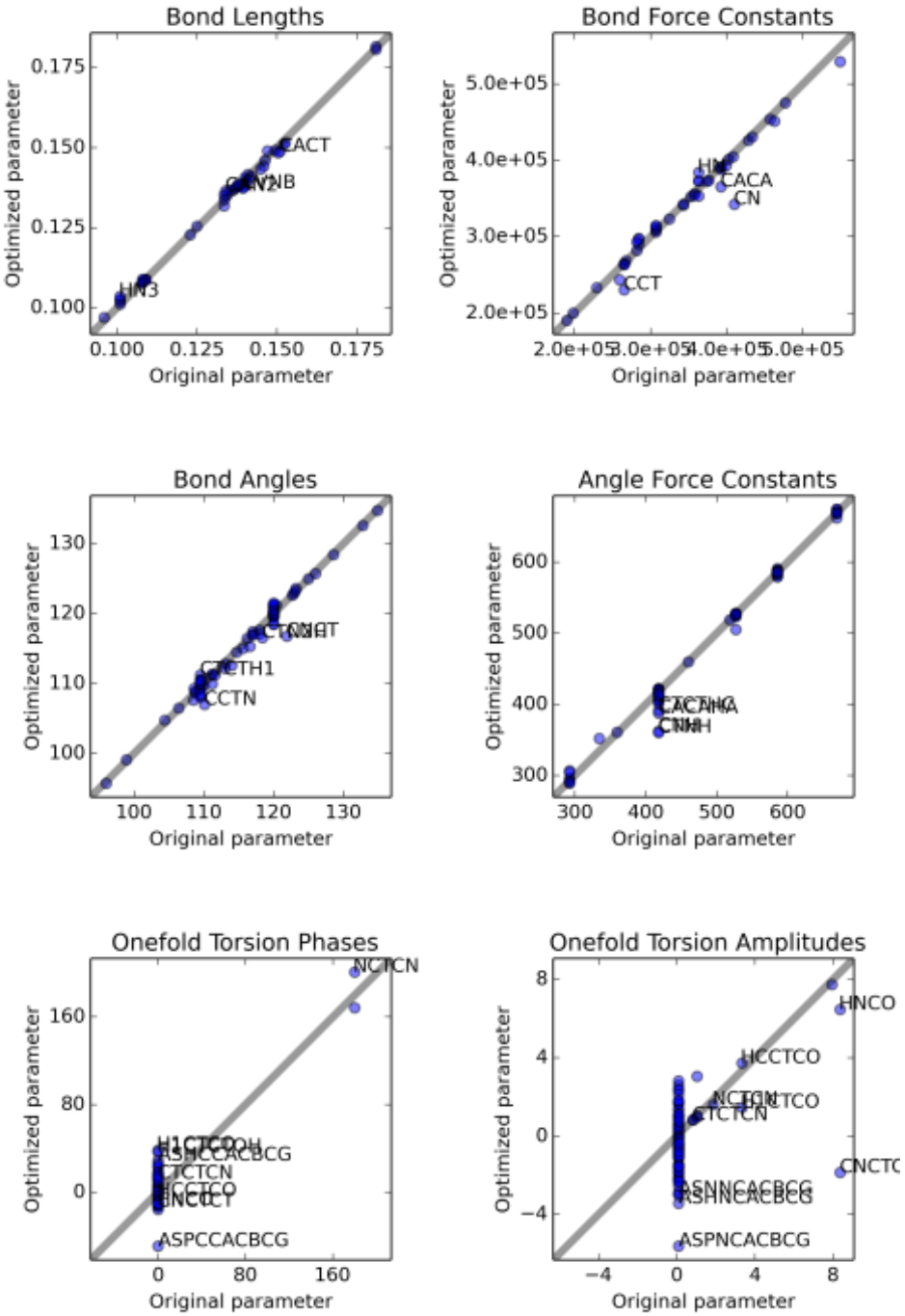


Figure S1. Scatter plots of initial and optimized force field parameters for AMBER-FB15. The GROMACS unit system is used (bond lengths in nm, bond force constants in kJ/mol/nm^2, angles in degrees, angle force constants in kJ/mol/rad^2, torsion phases in degrees, torsion amplitudes in kJ/mol.) Concatenated strings denoting the atom types in the interaction are displayed for selected force field parameters.

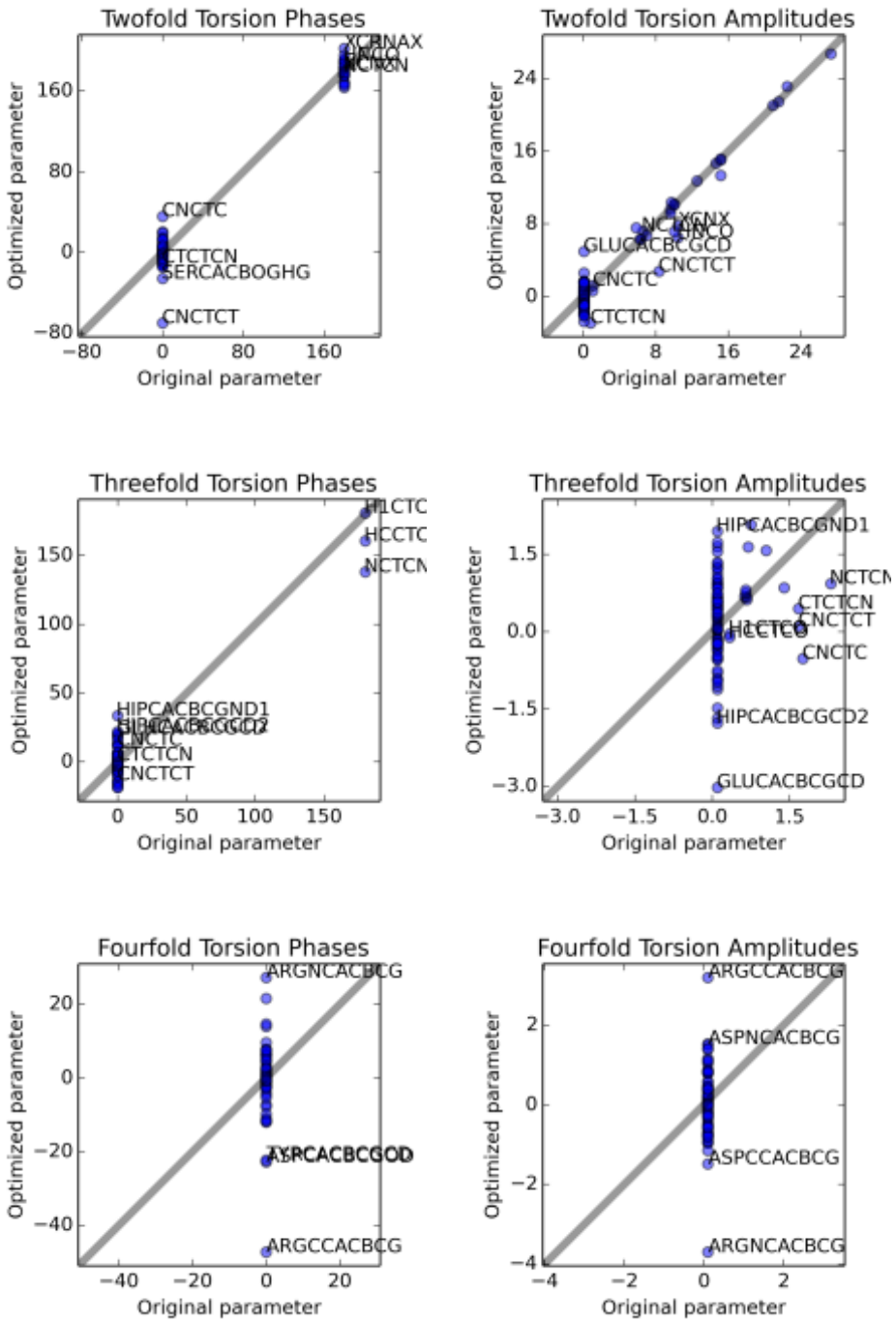


Figure S2. Scatter plots of initial and optimized force field parameters for AMBER-FB15. The GROMACS unit system is used (torsion phases in degrees, torsion amplitudes in kJ/mol.) Concatenated strings denoting the atom types in the interaction are displayed for selected force field parameters.

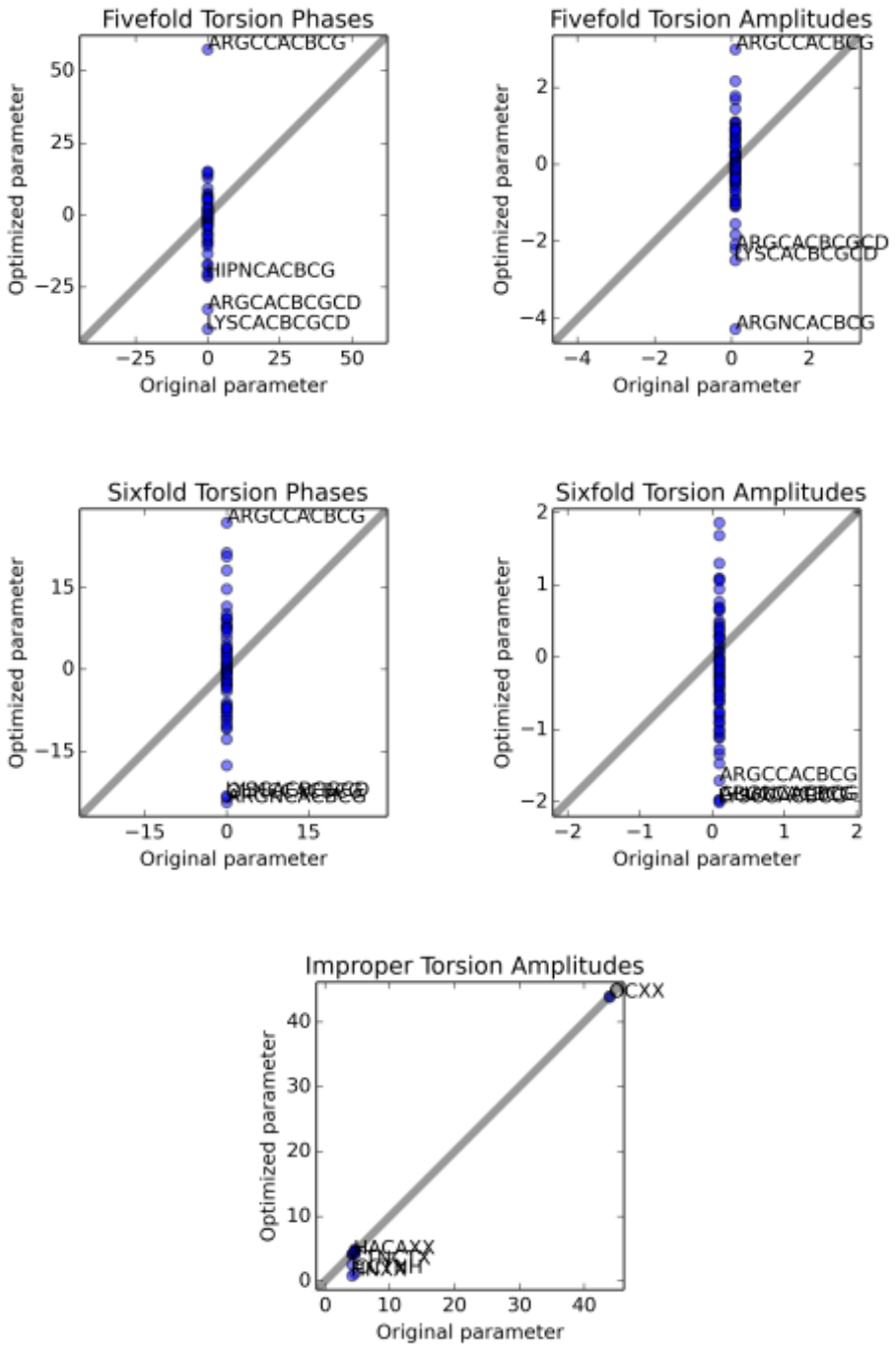


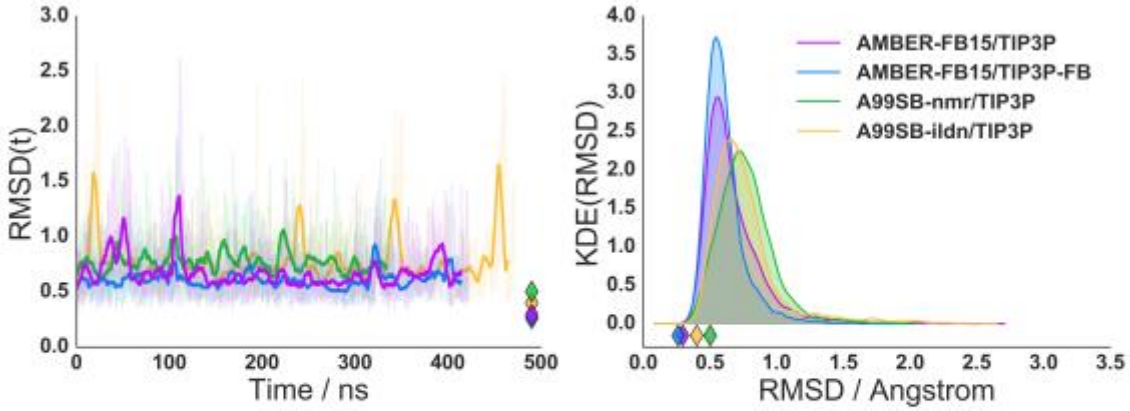
Figure S3. Scatter plots of initial and optimized force field parameters for AMBER-FB15. The GROMACS unit system is used (torsion phases in degrees, torsion amplitudes in kJ/mol.) Concatenated strings denoting the atom types in the interaction are displayed for selected force field parameters.

2. RMSD time series of simulated proteins

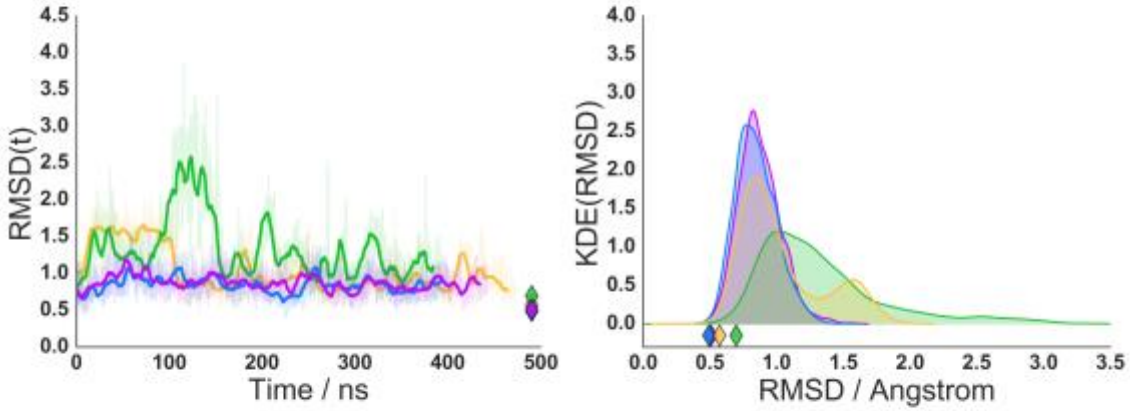
Protein	PDB ID	RMSD Start	RMSD End
Acetyl	2EVN	9	94
GB3	1IGD	6	60
Lambda	1LMB	4	80
Lysozyme	1AM7	6	150
NTL9	2HBA	2	38
TrpCage	2JOF	2	14
Ubiq	1UBQ	2	71
Villin	2F4K	3	32

Table S1. PDB IDs and starting / ending residue numbers (numbered from 1, including endpoints) for computing the RMSD of the protein backbone to the crystal structure.

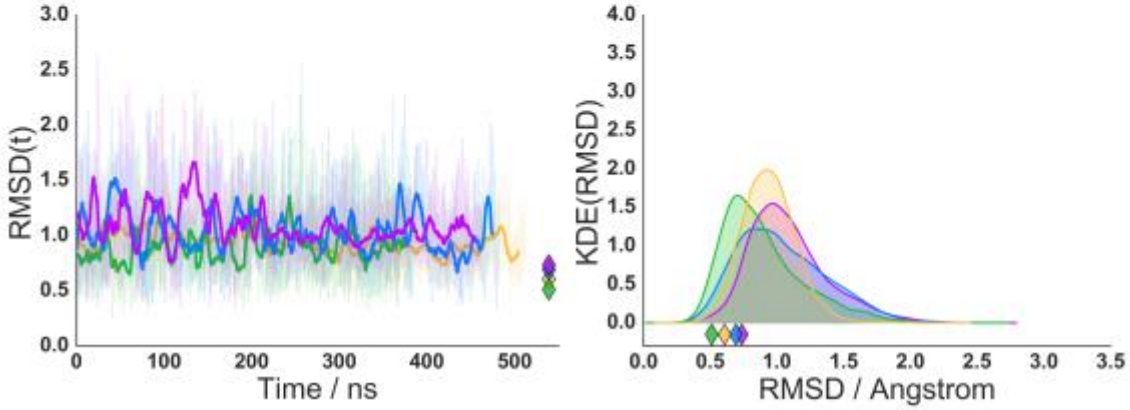
NTL9



Lambda



Villin



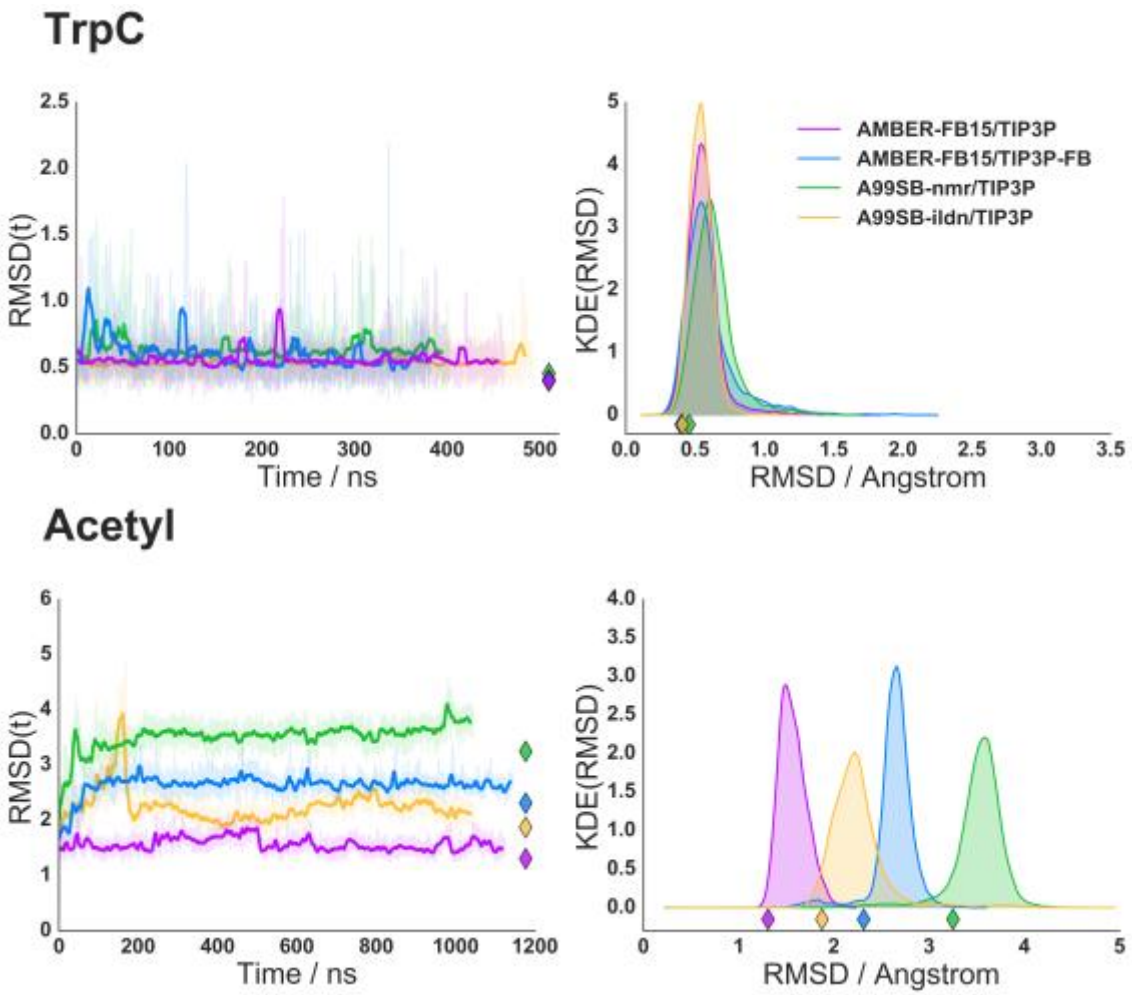


Figure S4. Backbone RMSD time series for five simulated proteins (NTL9, lambda repressor, villin headpiece, TrpCage, and acetyltransferase) using four models. The symbol indicates the backbone RMSD calculated using the averaged Cartesian coordinates.

3. Bacteriophage lysozyme S2 threshold

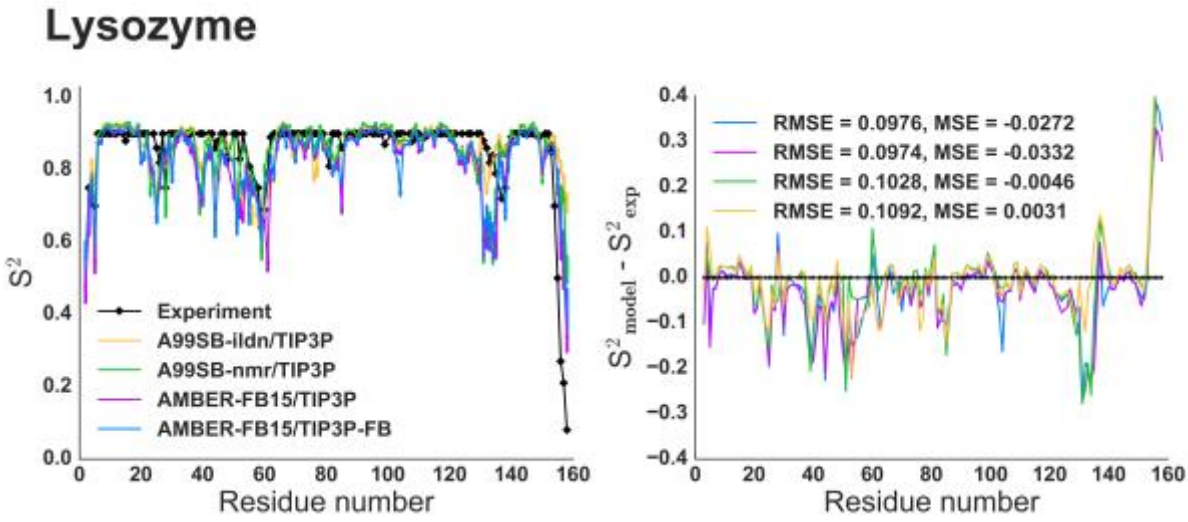


Figure S5. Backbone RMSD time series for five simulated proteins (NTL9, lambda repressor, villin headpiece, TrpCage, and acetyltransferase) using four models. The symbol indicates the backbone RMSD calculated using the averaged Cartesian coordinates.

4. NMR ^{3}J couplings of ubiquitin and NTL9

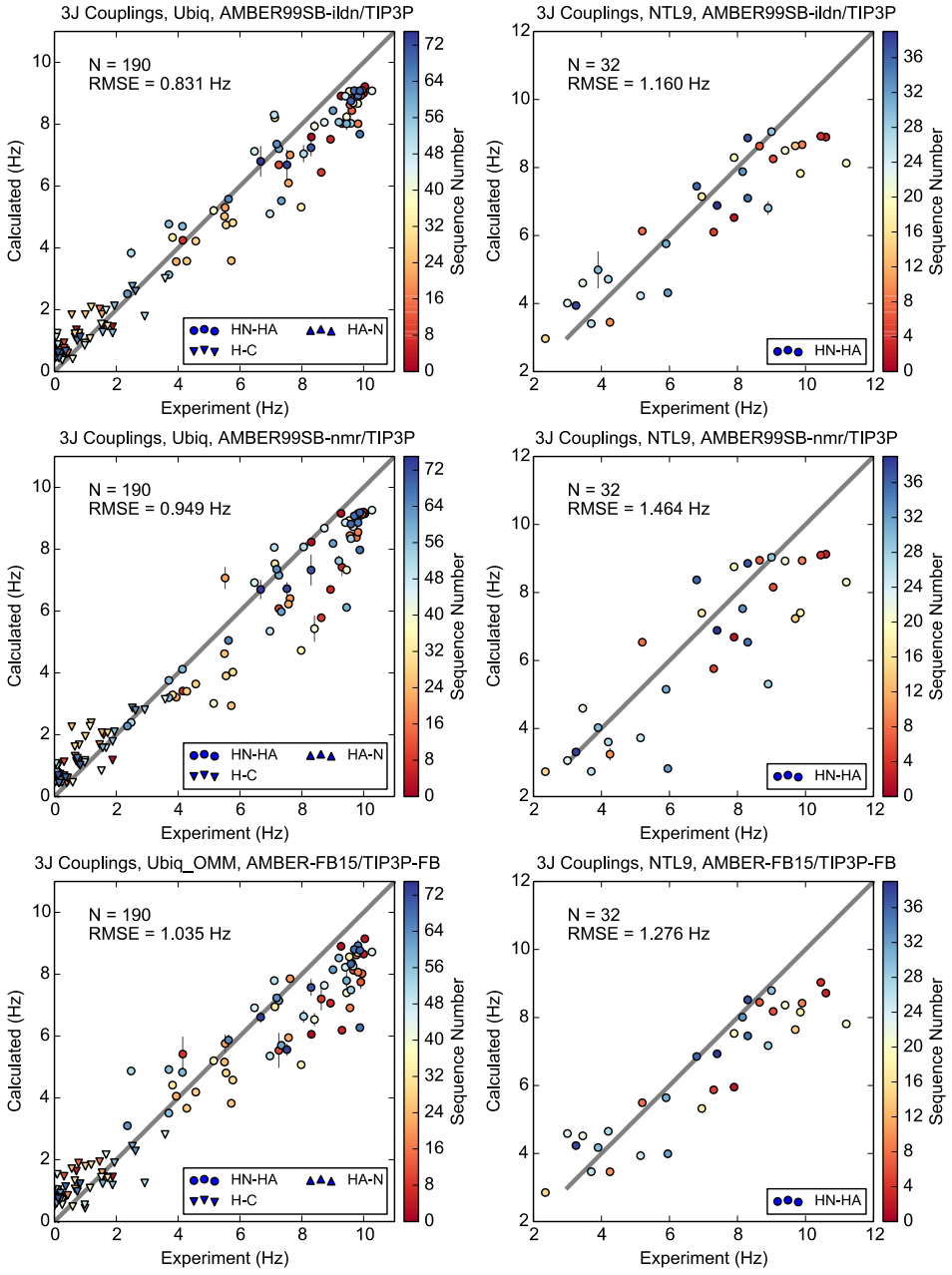


Figure S6. Scatter plots of experimental vs. calculated NMR three-bond scalar couplings. Two proteins are shown (left: ubiquitin, right: NTL9) and three models (top, AMBER99SB-ildn/TIP3P; middle, AMBER99SB-nmr/TIP3P; bottom, AMBER-FB15/TIP3P-FB from this work.) Symbols represent the atom pair involved in the coupling, and colors represent the position of the residue in the protein sequence.

5. Summaries of RMS errors in predicted chemical shifts

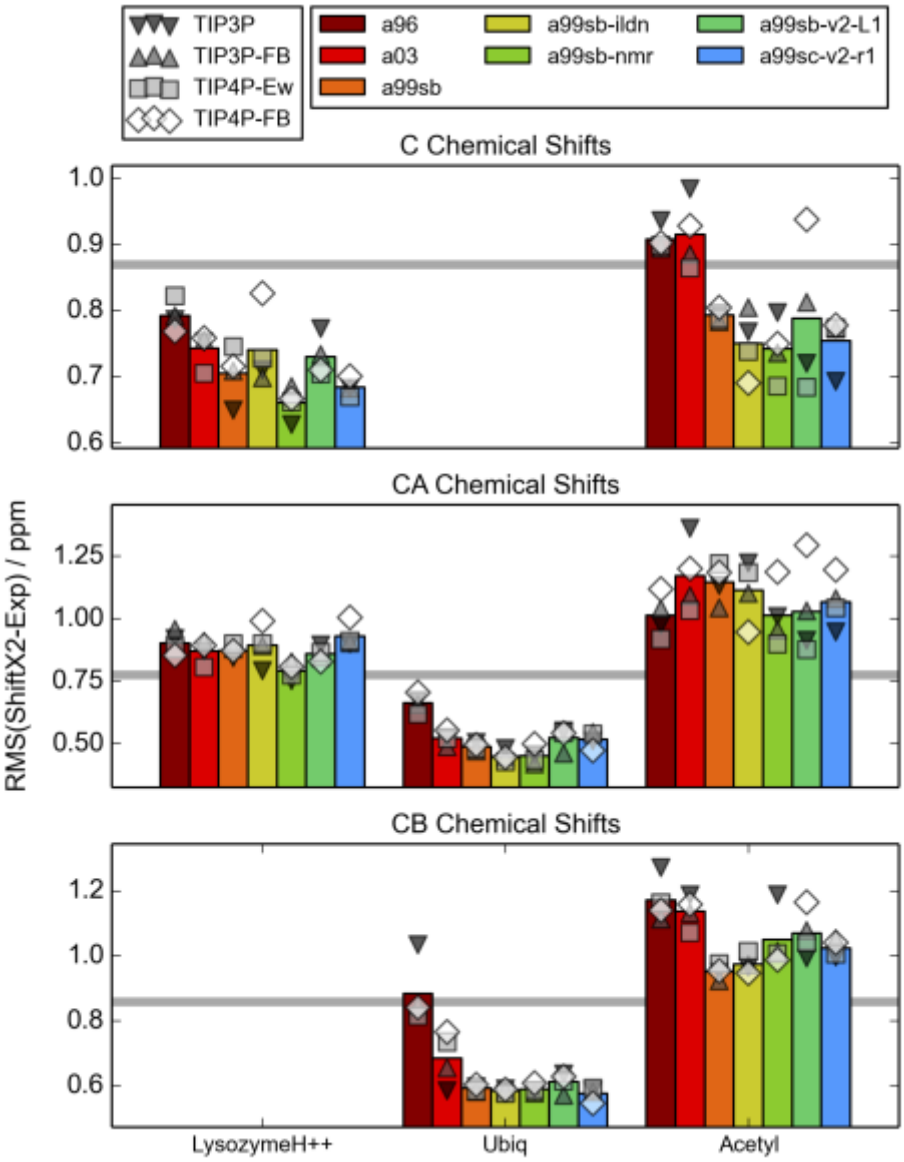


Figure S7: RMS errors in predicted ^{13}C chemical shifts calculated from MD trajectories and the SHIFTX2 model. Top, middle, and bottom panels correspond to chemical shifts measured for carbons in the C (amide bond), CA (alpha), and CB (beta) positions. Left, middle, and right groups of bars correspond to bacteriophage lysozyme, ubiquitin, and acetyltransferase. Each group of seven bars shows results for protein force fields (A96, A03, A99SB, A99SB-ILDN, A99SB-NMR, A99SB-V, AMBER-FB15) averaged over four simulations that employ different water models (TIP3P, TIP3P-FB, TIP4P-Ew, TIP4P-FB). The gray line represents the intrinsic error of the SHIFTX2 model.

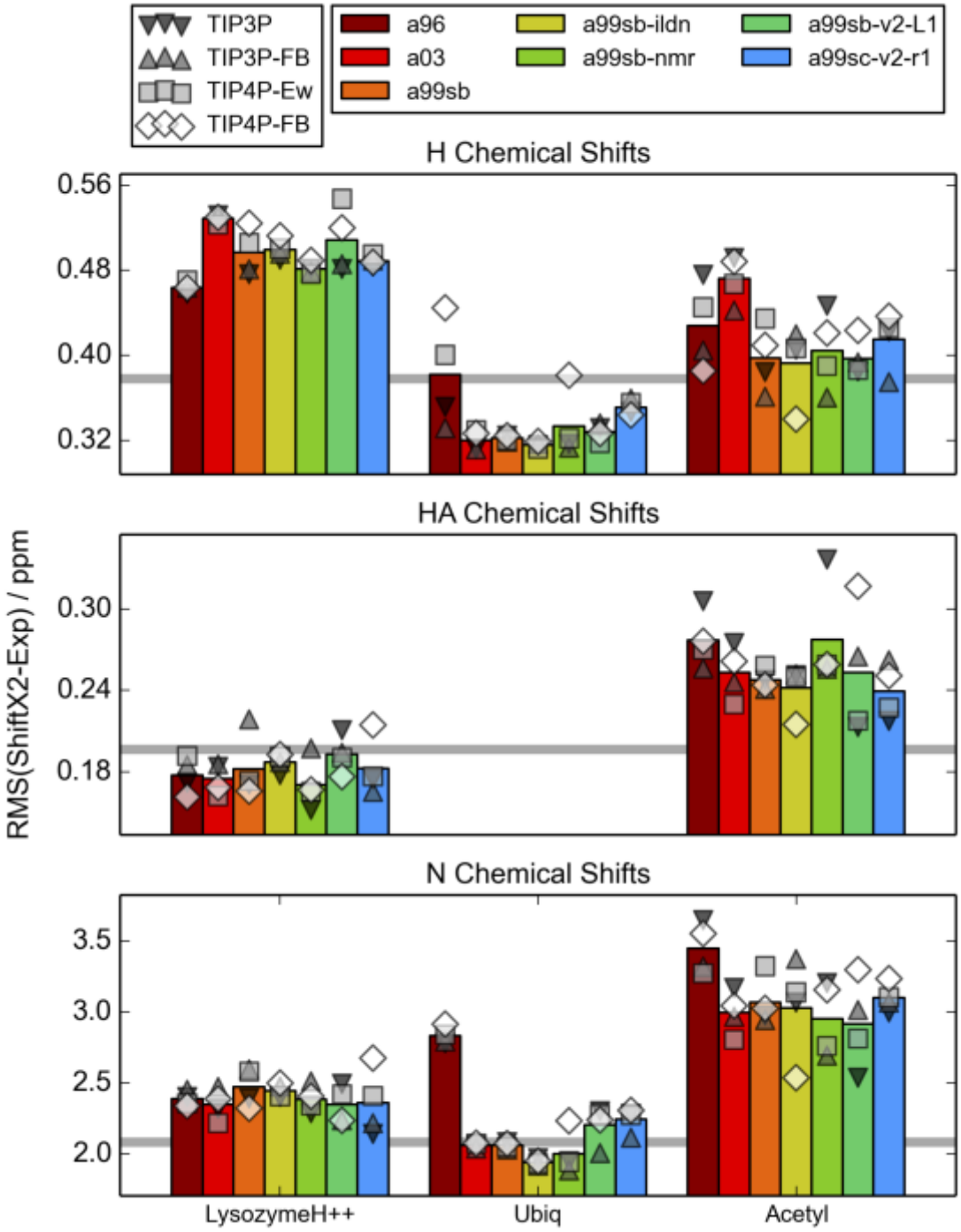


Figure S8: RMS errors in predicted ^{1}H and ^{15}N chemical shifts calculated from MD trajectories and the SHIFTX2 model. Top and middle panels correspond to ^{1}H chemical shifts for protons in the amide (H) and alpha (HA) positions. Left, middle, and right groups of bars correspond to lysozyme, ubiquitin, and acetyltransferase. Each group of seven bars shows results for protein force fields (A96, A03, A99SB, A99SB-ILDN, A99SB-NMR, A99SB-V, AMBER-FB15) averaged over four simulations that employ different water models (TIP3P, TIP3P-FB, TIP4P-Ew, TIP4P-FB). The gray line represents the intrinsic error of the SHIFTX2 model.

6. Protein temperature dependence with TIP4P-Ew and TIP4P-FB water models

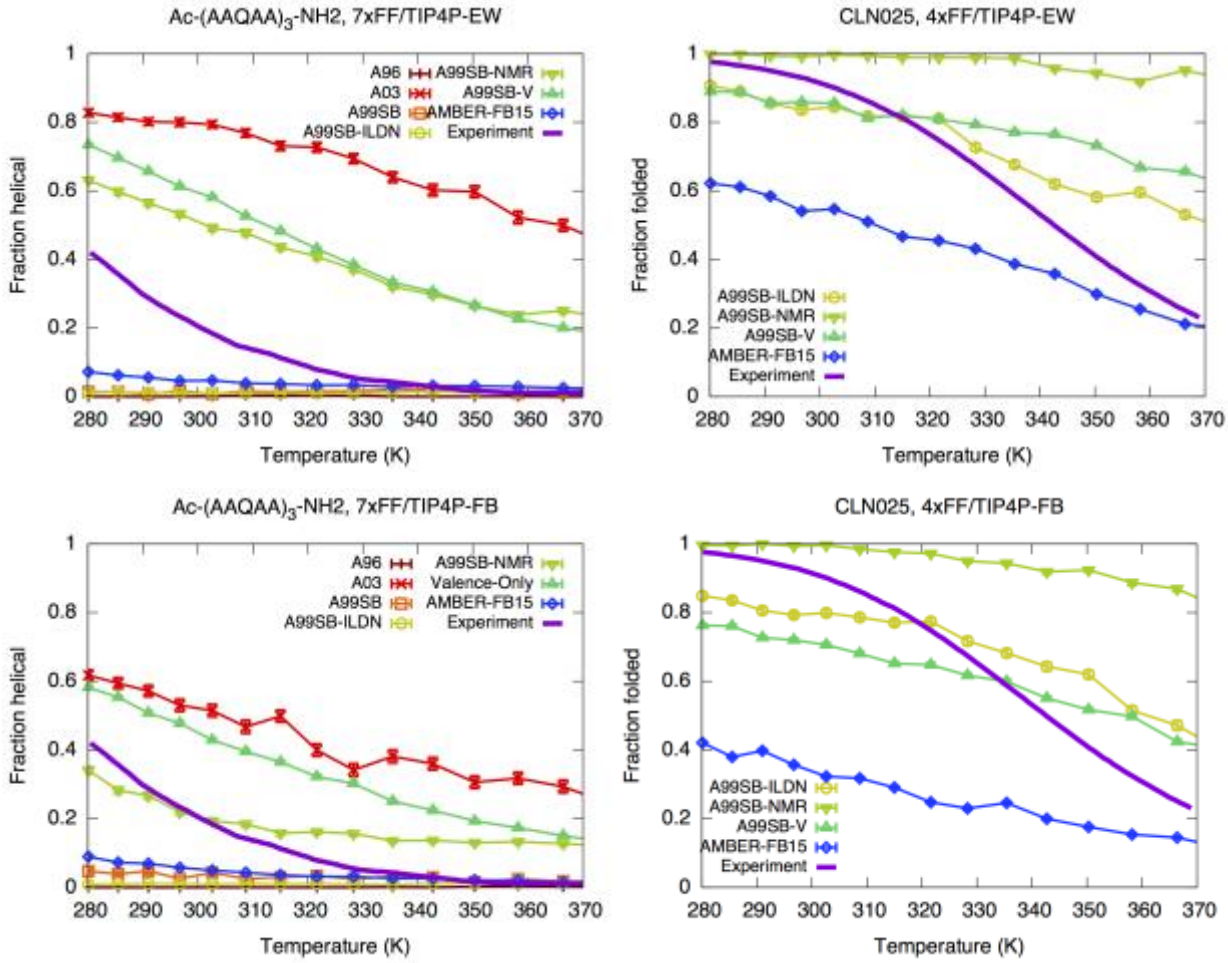


Figure S9. Temperature dependence of secondary structure for two small peptides as a function of temperature and several force field / water model combinations. Left column: The helical fraction of Ac-(AAQAA)$_3$-NH$_2$. Right column: The fraction folded of CLN025. Upper row: Comparison of multiple protein force fields using TIP4P-Ew water model. Lower row: Same comparison using TIP4P-FB water model.

7. Comparison of water-protein interaction energies

AAQAA simulated with A99SB-V		
Water Model	**Water-Protein Interaction (kJ/mol)**	**Fraction Folded**
TIP3P	-965.4 (4.2)	0.761 (0.006)
TIP3P-FB	-1066.1 (5.1)	0.697 (0.005)
TIP4P-Ew	-1146.0 (7.9)	0.592 (0.005)
TIP4P-FB	-1224.0 (8.5)	0.462 (0.006)

Table S2. Comparison of water-protein interaction energies and folded fraction for Ac-$(AAQAA)_3$-NH_2 simulated with the A99SB-V model and four water models. The fraction folded is derived from a linear fit to the temperature dependence plots and evaluated at 300 K. Quantities in parentheses are one standard error.

CLN025 simulated with AMBER-FB15		
Water Model	**Water-Protein Interaction (kJ/mol)**	**Fraction Folded**
TIP3P	-1488.2 (2.8)	0.896 (0.007)
TIP3P-FB	-1630.4 (3.3)	0.755 (0.006)
TIP4P-Ew	-1699.4 (4.8)	0.543 (0.004)
TIP4P-FB	-1799.6 (5.5)	0.342 (0.005)

Table S3. Comparison of water-protein interaction energies and folded fraction for CLN025 simulated with the AMBER-FB15 model and four water models. The fraction folded is derived from a linear fit to the temperature dependence plots and evaluated at 300 K. Quantities in parentheses are one standard error.

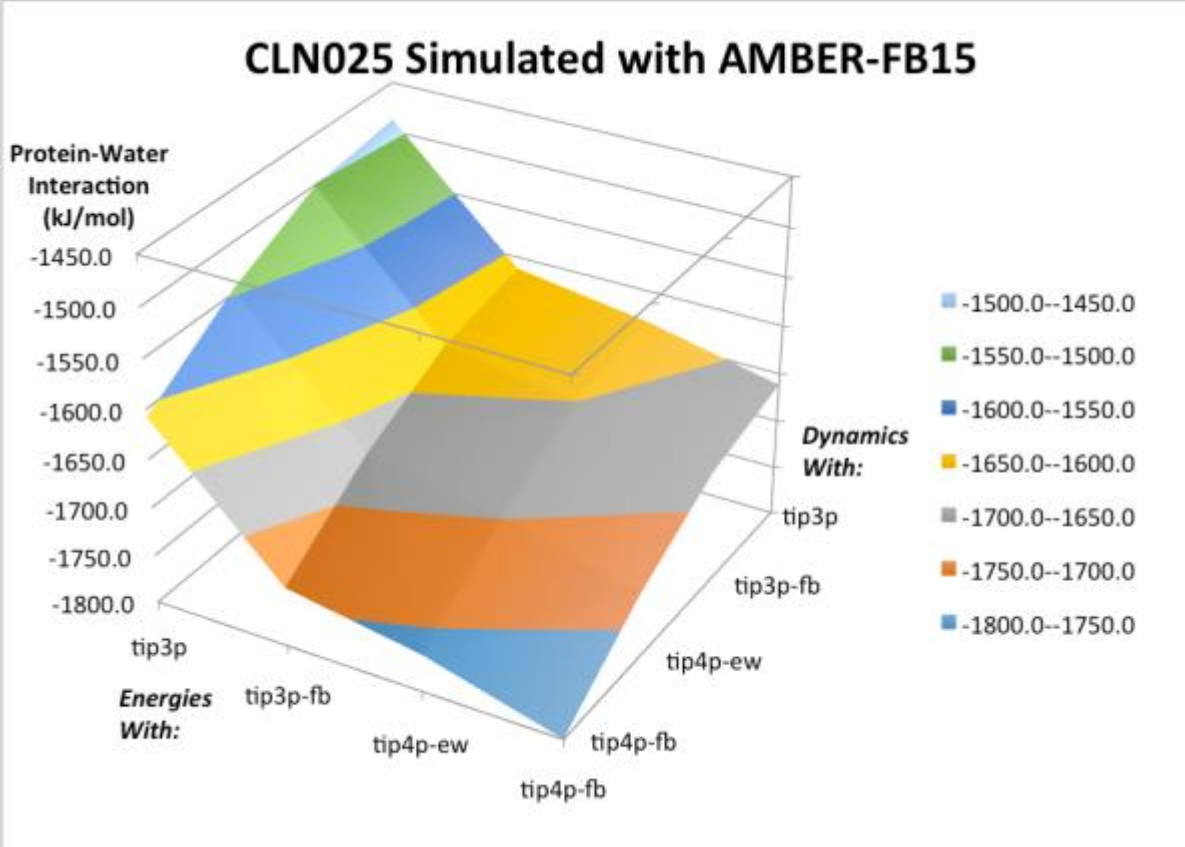


Figure S10. Comparison of water-protein interaction energies and folded fraction for CLN025 simulated with AMBER-FB15 and four water models. To create the off-diagonal entries, short (10 ps) simulations were initiated from the snapshots of one long simulation (e.g. AMBER-FB15/TIP3P) with the water molecules replaced by another model (e.g. TIP3P-FB). The slope indicates that changing the water model from TIP3P→TIP3P-FB→TIP4P-EW→TIP4P-FB not only increases the protein-water interaction strength, but also biases the conformations toward stronger protein-water (unfolded) interactions.