

Published in final edited form as:

J Chem Theory Comput. 2012 April 10; 8(4): 1409–1414. doi:10.1021/ct2007814.

Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements

 Kyle A. Beauchamp[†], Yu-Shan Lin[‡], Rhiju Das^{†,¶}, and Vijay S. Pande^{†,*,‡}
[†]Biophysics Program, Stanford University, Stanford, CA

[‡]Chemistry Department, Stanford University, Stanford, CA

[¶]Biochemistry Department, Stanford University, Stanford, CA

Abstract

Recent hardware and software advances have enabled simulation studies of protein systems on biophysically-relevant timescales, often revealing the need for improved force fields. Although early force field development was limited by the lack of direct comparisons between simulation and experiment, recent work from several labs has demonstrated direct calculation of NMR observables from protein simulations. Here we quantitatively evaluate recent molecular dynamics force fields against a suite of 524 chemical shift and J coupling (${}^3J_{NH_\alpha}$, ${}^3J_{NC_\beta}$, ${}^3J_{\alpha C'}$, ${}^3J_{NC'}$, and ${}^3J_{\alpha N}$) measurements on dipeptides, tripeptides, tetra-alanine, and ubiquitin. Of the force fields examined (ff96, ff99, ff03, ff03*, ff03w, ff99sb*, ff99sb-ildn, ff99sb-ildn-phi, ff99sb-ildn-nmr, CHARMM27, OPLS-AA), two force fields (ff99sb-ildn-phi, ff99sb-ildn-nmr) combining recent side chain and backbone torsion modifications achieve high accuracy in our benchmark. For the two optimal force fields, the calculation error is comparable to the uncertainty in the experimental comparison. This observation suggests that extracting additional force field improvements from NMR data may require increased accuracy in J coupling and chemical shift prediction. To further investigate the limitations of current force fields, we also consider conformational populations of dipeptides, which were recently estimated using vibrational spectroscopy.

1 Introduction

Molecular dynamics (MD) simulation is a versatile computational tool that allows investigation of condensed phase systems including neat alkanes,¹ the many phases of water,² the solvation and binding of small molecules,^{3,4} and the folding dynamics of full protein systems.^{5,6} Recent gains in computer performance, parallelized MD codes,⁷ and optimizations for GPU⁸ and other specialized hardware⁹ have enabled simulations of aqueous macromolecules over times exceeding 100 ns in single-day calculations. These accelerations, however, have begun to reveal inaccuracies in current MD force fields.

By and large, molecular dynamics force fields are parameterized to reproduce quantum mechanical calculations on small model systems,^{10–13} then adjusted to provide improved agreement with higher-quality ab initio data,¹⁴ crystallographic structures,¹⁵ or experimental data.^{16–18} Because of the many design choices inherent in parameterization, force fields

*To whom correspondence should be addressed pande@stanford.edu.

Supporting Information

Additional error analysis is available in the supporting information. The Karplus relations used in this work are tabulated along with estimates of systematic uncertainty. Gromacs parameter files and PDBs for the present simulations are also included. This information is available free of charge via the Internet at <http://pubs.acs.org>

yield considerable differences in predicted biophysical properties. For example, studies of protein folding have revealed variation in folding rates between different force fields.¹⁹ Similarly, simulated proteins often have folding midpoint temperatures that err by 25 K or more.⁶

Here, we systematically evaluate eleven recent force fields combined with each of five widely-used water models (55 combinations) against a benchmark set of 524 NMR measurements. The evaluated force fields include recent AMBER, CHARMM, and OPLS-AA variants; the solvent models include recent implicit and explicit models. The 524 NMR measurements include J coupling and chemical shift data of 32 model systems. Measurements span all 19 non-proline amino acids and include dipeptide,^{20,21} tripeptide,^{22,23} tetrapeptide,²² and full protein systems;²⁴ importantly, this systematic benchmark contains model systems not previously considered in the parameterization of the tested force fields. These comparisons, which comprise over 25 μ s of aggregate simulation, suggest that explicit solvent simulations with either the ff99sb-ildn-phi or ff99sb-ildn-nmr force field recover NMR observables with an accuracy close to the uncertainty inherent in current scalar coupling and chemical shift models. In addition to quantitative comparisons to NMR experiments, we also compare conformational populations of the 19 dipeptides to recent estimates made using vibrational spectroscopy.^{20,21}

2 Methods

2.1 Benchmark Systems

For our benchmark, we selected 32 protein systems including capped dipeptides (Ace-X-NME, X = P), tripeptides (XXX, GYG, X \in {A, G, V}, Y \in {A, V, F, L, S, E, K, M}), alanine tetrapeptide, and ubiquitin. Each of these systems has NMR data available in the form of chemical shifts, J couplings, or both. Small peptides provide minimal model systems that can sample the (ϕ , ψ) torsions that are a key component of secondary structure formation. On the other hand, a different balance of forces is at play in larger systems, which led us to include ubiquitin, a key model system in protein folding²⁵ and NMR studies.²⁴

2.2 Force Field Benchmark

We aggregated 524 measurements of chemical shifts and scalar couplings, summarized in Table S1. Data was taken from the BioMagRes Database²⁶ or several recent papers.^{20,22–24} Each of the 32 systems was simulated using Gromacs 4.5.4⁷ using all combinations of the GBSA,²⁷ TIP3P, SPC/E, TIP4P-EW,²⁸ and TIP4P/2005² water models with the ff96,¹⁰ ff99,¹¹ ff03,¹² ff03w,¹⁶ ff03*,¹⁷ ff99sb*,¹⁷ ff99sb-ildn,¹⁴ ff99sb-ildn-phi,¹⁸ and ff99sb-ildn-nmr,²⁹ CHARMM27,^{30,31} and OPLS-AA³² force fields. Here, ff99sb-ildn-phi refers to a force field combining the ff99sb-ildn side chain optimizations¹⁴ with a recently modified ϕ' potential.¹⁸ Compared to ff99sb-ildn, ff99sb-ildn-phi differs by a modification of the periodicity two ($n = 2$) ϕ' torsional potential; this term has strength 2.00 kcal / mol and 1.80 kcal / mol, respectively, for the ff99sb-ildn and ff99sb-ildn-phi force fields. The ff99sb-ildn-nmr force field refers to the combination of the ff99sbildn side chain optimizations with the NMR-optimized backbone torsions of the ff99sb-nmr force field.²⁹

Each production simulation was 25 ns (20 ns for dipeptides) in length and held at constant temperature and pressure; simulations for peptides were started from conformations generated by PyMol. Ubiquitin simulations were started from the crystal structure (PDB: 1UBQ).³³ For error analysis, simulations were repeated with independent starting velocities. For the peptide systems, the independent runs also used different starting conformations; for ubiquitin, the second set of runs began from the crystal structure, had independent starting velocities, and were 50 ns in length. J couplings were estimated using empirical Karplus

relations parameterized by the Bax group;³⁴ chemical shifts were estimated using the Sparta+ program³⁵ (in Figure S1, we consider alternative models for J couplings and chemical shifts).

All simulations were performed with Gromacs⁷ 4.5.4. Starting conformations were solvated, neutralized with Na⁺ or Cl⁻, minimized, and equilibrated before production runs. For explicit solvent, electrostatics were treated using particle mesh ewald³⁶ with a real-space cutoff of 1.0 nm. Van der Waals interactions were switched off between 0.7 and 0.9 nm. Temperature control was achieved using the velocity rescaling thermostat.³⁷ Pressure control (1 atm) was achieved using either the Berendsen barostat (for equilibration) or the Parrinello-Rahman³⁸ barostat (for production). For implicit solvent simulations (GBSA), temperature control was achieved using a Langevin integrator. The temperatures were chosen to match experimental conditions; dipeptides were held at 303K, GXG tripeptides at 298K, homotripeptides at 300K, and ubiquitin at 303K. For ubiquitin, the protein was held fixed during the minimization and equilibration steps.

3 Results

3.1 Optimal performance from ff99sb-ildn-nmr and ff99sb-ildn-phi

Converting all 524 measurements into an uncertainty-weighted objective function

$$\chi^2 = \sum_i \frac{(x_i^{Exp} - x_i)^2}{\sigma_i^2}$$

allows a force field evaluation based on all available data (Figure 1).

We estimate the errors in each comparison as the uncertainty in the relationship between conformation and NMR observable; these errors (Table S2, Table S3) were estimated during the parameterizations of the various Karplus relations and the Sparta+ chemical shift model. Based on this comprehensive analysis, the early (ff96, ff99) force fields are easily rejected in favor of more recent modifications. Furthermore, ff99sb-ildn-nmr and ff99sb-ildn-phi most accurately recapitulate the chosen NMR experiments. Raw data for TIP4P-EW with several force fields are shown in Figure S4–Figure S13.

3.2 Accuracy for Dipeptides, Tripeptides, Tetrapeptides, and Ubiquitin

The model systems in this work consist of ubiquitin, alanine tetrapeptide, tripeptides, and the 19 capped dipeptides (e.g. Acetyl-Ala-N-methylamide). Because these systems differ considerably in size, it is important to ask whether force fields perform well for all classes of model system. In particular, it is important to avoid overfitting to experiments probing only small systems, as that could compromise accuracy on larger systems where less protein is solvent-exposed. We find that ff99sb-ildn-phi and ff99sb-ildn-nmr provide good performance for all three classes of systems (Figure 2).

3.3 Performance By Experiment

For five out of ten experiments, ff99sb-ildn-phi and ff99sb-ildn-nmr achieve accuracy

comparable to the uncertainty of the comparison (e.g. $\frac{\chi^2}{n} \approx 1$; see Table S4, Figure S14–Figure S23). Moderate deviations are found for two experiments (${}^3J(H^N C')$, ${}^3J(H^{\alpha} N)$), but large deviations are found for carbonyl chemical shifts ($CS-C$, ${}^3J(H^{\alpha} C')$, and ${}^3J(H^N H^{\alpha})$). First, large errors in carbon chemical shifts are found for all choices of force field and water model. This error may indicate systematic error in either the chemical shifts estimated by Sparta+ or the conformational propensities of *all* force fields evaluated here. Because Sparta+ is parameterized empirically using a neural network approach, it is challenging to further dissect errors in chemical shift predictions. Second, moderate errors in ${}^3J(H^{\alpha} C')$ are observed; inspection of individual errors (Figure S24) identifies GGG as the largest error

contributor. The failure for GGG is consistent with previous observations¹⁸ that improved Karplus parameterizations may be required for glycine residues. Finally, moderate errors in $^3J(H^N H^{\alpha})$ are observed. However, ff99sb-ildn-phi and ff99sb-ildn-nmr perform significantly better than their predecessors ff99sb-ildn and ff99 (Figure 3). The $^3J(H^N H^{\alpha})$ analysis suggests that correcting backbone torsions^{16–18,29} has led to real improvements in force field accuracy.

3.4 Performance By Amino Acid

Another key question is whether force field performance is consistent among the different amino acids. Large variations in quality between different amino acids might indicate an easy way to improve force fields. To formalize the agreement with experiment, we calculate

values of reduced χ^2 . Values of $\frac{\chi^2}{n}$ under 1 suggest that errors are within the measurement uncertainty. This analysis leaves 6/19 amino acids open to further optimization, including ALA, GLY, SER, VAL, ILE, and ASP. Force field improvements for these residues could lead to reduced errors in the present benchmark. For the remaining amino acids, improved accuracy in chemical shift and J coupling estimation may eventually reveal force field inadequacies that are within the uncertainty of the present comparison.

3.5 Populations of 19 Dipeptides

A recent analysis²¹ used NMR and vibrational spectroscopy to estimate the α_R , β , and P_{II} populations of 19 capped dipeptides. Here, we use (ϕ, ψ) state definitions³⁹ to estimate conformational populations from simulation (Figure 5, Figure S25). Because of uncertainties in state definitions and the fitting procedure used in the experimental analysis,²¹ this comparison is less direct than the NMR comparisons above; despite this limitation, population analysis provides several key insights.

First, with the exception of ff96, the MD force fields uniformly *over-emphasize* α_R in dipeptide simulations. Second, the GBSA implicit solvent model aggravates this error, leading to even further bias towards helical conformations. Third, recent force fields (ff03*, ff99sb-ildn, ff99sb-ildn-phi, ff03w, ff99sb-ildn-nmr) combined with explicit solvent show close agreement with the conformational populations estimated using a PDB-derived coil library.³⁹ This may suggest that current force fields perform better in the context of a full protein system or that direct comparison of simulation to IR-based population estimates is hindered by systematic error.

Although the over-estimation of α_R in dipeptides suggests that further decreasing α_R may improve current force fields, such changes are not always feasible. As an example, we point out that ff99sb is known to *under-emphasize*¹⁷ α_R in the helix-forming Ac-(AAQAA)₃-NH₂.^{40,41} Thus, it may be challenging for fixed-charge force fields to simultaneously achieve accurate helical propensities in both dipeptide and longer protein systems. To further investigate, we simulated Ac-(AAQAA)₃-NH₂ using both ff99sb-ildn-phi and ff99sb-ildn-nmr (Figure S26). Based on this test, ff99sb-ildn-phi is somewhat under-helical, while ff99sb-ildn-nmr is somewhat over-helical. This suggests that a future modification of ff99sb-ildn-nmr with slightly reduced helical content could lead to modest improvements in force field quality.

4 Discussion

4.1 Optimal choice of force field and water model

The overall χ^2 analysis (Figure 1) suggests that the ff99sb-ildn-phi and ff99sb-ildn-nmr force fields (with explicit water) are best able to recapitulate the 524 NMR measurements in

the present benchmark. These results are robust to different models for J couplings and chemical shifts (Figure S1), to sampling uncertainty in the simulations (Figure S2), and to the relative importance placed on J coupling versus chemical shift experiments (Figure S3).

Explicit water models (TIP3P, SPC/E, TIP4P-EW, and TIP4P/2005) outperform GBSA. However, choosing between TIP3P, SPC/E, TIP4P-EW, and TIP4P/2005 is difficult, as the results are force field dependent. However, our current analysis does suggest that variants of ff99sb give reasonable performance with TIP4P-EW (and, to a lesser extent, TIP4P/2005). Recent four-point water models have vastly improved thermodynamic properties;^{2,28} our results recommend their broader use in protein simulations, although further validation of their nonbonded interactions may be necessary.

4.2 Future Force Field Development

With the simple functional forms used, how much can current force fields be improved? Given the many small but measurable improvements recently published,^{14,16–18} it is likely that the current functional forms may allow further increases in accuracy. The best path for improvement, however, is complex. The non-bonded terms, including partial charges, are based on decade old calculations; increasing computational resources allow increasingly accurate QM data to be used for parameterization. More accurate bonded terms might also be possible. Another possibility is the use of amino-acid specific torsional potentials, rather than identical potentials for all (non-glycine) residues. Such a procedure would allow researchers to refine force fields for amino acids where performance is presently inadequate.

For further developments, a key question is whether to use *ab initio*¹⁴ or experimental data^{16–18,29} when fitting parameters. Fitting to *ab initio* data is hindered by the difficulty of modeling solvent effects during parameterization. Furthermore, parameters such as partial charges may not be transferable between gas and condensed phase environments—or even between hydrophobic and solvent-exposed environments. Fitting to experimental data is currently hindered by two key limitations. First, only limited data is available for model systems. Second, the quantitative connection between simulation and experiment typically relies on parameterized relationships such as the Karplus relationship or the chemical shift predictions of Sparta+. These parameterizations have large uncertainties (typically larger than the statistical errors in either simulation or experiment), may contain systematic errors (such as amino-acid specific biases¹⁸), and rely on protein structural models that are ensemble averaged (possibly blurring important short-timescale dynamics⁴²). We point out that the current analysis focuses on agreement with NMR (chemical shift and J coupling) experiments, which emphasize local bonded interactions. Other experimental data, such as solvation free energies, may be critical for evaluating other aspects of force field performance.

5 Conclusions

Molecular simulation promises atomic-detail modeling of key processes in chemistry and biophysics, but only if the underlying force field has demonstrated accuracy. Here, we have shown that recent force field enhancements lead to increased accuracy in recapitulating a benchmark set of 524 NMR measurements. Simulations performed in explicit water with either the ff99sb-ildn-phi or ff99sb-ildn-nmr force field achieve RMS errors that are comparable to the uncertainty in calculating the experimental observables. Future work may require advances in both force field development and accurate calculation of NMR observables.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank NSF CNS-0619926 for computing resources and NIH R01-GM062868, NSF-DMS-0900700, and NSF-MCB-0954714 for funding. KAB was supported by a Stanford Graduate Fellowship. R.D. was supported by a Burroughs-Wellcome Foundation Career Award at the Scientific Interface. We acknowledge the following award for providing computing resources that have contributed to the research results reported within this paper: MRI-R2: Acquisition of a Hybrid CPU/GPU and Visualization Cluster for Multidisciplinary Studies in Transport Physics with Uncertainty Quantification <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0960306>. This award is funded under the American Recovery and Reinvestment Act of 2009 (Public Law 111-5).

References

1. Jorgensen W, Maxwell D, Tirado-Rives J. *J. Am. Chem. Soc.* 1996; 118:11225–11236.
2. Abascal J, Vega C. *J. Chem. Phys.* 2005; 123:234505. [PubMed: 16392929]
3. Mobley D, Graves A, Chodera J, McReynolds A, Shoichet B, Dill K. *J. Mol. Biol.* 2007; 371:1118–1134. [PubMed: 17599350]
4. Shirts M, Pande V. *J. Chem. Phys.* 2005; 122:134508. [PubMed: 15847482]
5. Beauchamp K, Ensign D, Das R, Pande V. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:12734. [PubMed: 21768345]
6. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W. *Science*. 2010; 330:341–346. [PubMed: 20947758]
7. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. *J. Chem. Theory Comput.* 2008; 4:435–447.
8. Eastman P, Pande V. *Comput. Sci. Eng.* 2010; 12:34–39.
9. Shaw D, Deneroff M, Dror R, Kuskin J, Larson R, Salmon J, Young C, Batson B, Bowers K, Chao J. *Commun. ACM.* 2008; 51:91–97.
10. Kollman P. *Acc. Chem. Res.* 1996; 29:461–469.
11. Wang J, Cieplak P, Kollman P. *J. Comput. Chem.* 2000; 21:1049–1074.
12. Duan Y, Wu C, Chowdhury S, Lee M, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T. *J. Comput. Chem.* 2003; 24:1999–2012. [PubMed: 14531054]
13. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. *Proteins: Struct., Funct., Bioinf.* 2006; 65:712–725.
14. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J, Dror R, Shaw D. *Proteins: Struct., Funct., Bioinf.* 2010; 78:1950–1958.
15. MacKerell A Jr, Feig M, Brooks C III. *J. Am. Chem. Soc.* 2004; 126:698–699. [PubMed: 14733527]
16. Best RB, Mittal J. *J. Phys. Chem. B.* 2010; 114:14916–14923. [PubMed: 21038907]
17. Best R, Hummer G. *J. Phys. Chem. B.* 2009; 113:9004–9015. [PubMed: 19514729]
18. Nerenberg P, Head-Gordon T. *J. Chem. Theory Comput.* 2011; 7:1220–1230.
19. Piana S, Lindorff-Larsen K, Shaw D. *Biophys. J.* 2011; 100:L47–L49. [PubMed: 21539772]
20. Avbelj F, Grdadolnik S, Grdadolnik J, Baldwin R. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:1272. [PubMed: 16423894]
21. Grdadolnik J, Mohacek-Grosec V, Baldwin R, Avbelj F. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:1794. [PubMed: 21205907]
22. Graf J, Nguyen P, Stock G, Schwalbe H. *J. Am. Chem. Soc.* 2007; 129:1179–1189. [PubMed: 17263399]
23. Hagarman A, Measey T, Mathieu D, Schwalbe H, Schweitzer-Stenner R. *J. Am. Chem. Soc.* 2009; 132:540–551. [PubMed: 20014772]
24. Wang A, Bax A. *J. Am. Chem. Soc.* 1996; 118:2483–2494.
25. Krantz B, Sosnick T. *Biochemistry.* 2000; 39:11696–11701. [PubMed: 10995237]

26. Ulrich E, Akutsu H, Doreleijers J, Harano Y, Ioannidis Y, Lin J, Livny M, Mading S, Maziuk D, Miller Z. *Nucleic Acids Res.* 2008; 36:D402. [PubMed: 17984079]
27. Onufriev A, Bashford D, Case D. *Proteins.* 2004; 55:383–394. [PubMed: 15048829]
28. Horn H, Swope W, Pitera J, Madura J, Dick T, Hura G, Head-Gordon T. *J. Chem. Phys.* 2004; 120:9665. [PubMed: 15267980]
29. Li D, Bruschweiler R. *Angew. Chem.* 2010; 122:6930–6932.
30. Mackerell A Jr, Feig M, Brooks C III. *J. Comput. Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
31. Bjelkmar P, Larsson P, Cuendet M, Hess B, Lindahl E. *J. Chem. Theory Comput.* 2010; 6:459–466.
32. Kaminski G, Friesner R, Tirado-Rives J, Jorgensen W. *J. Phys. Chem. B.* 2001; 105:6474–6487.
33. Vijay-kumar S, Bugg C, Cook W. *J. Mol. Biol.* 1987; 194:531–544. [PubMed: 3041007]
34. Hu J-S, Bax A. *J. Am. Chem. Soc.* 1997; 119:6360–6368.
35. Shen Y, Bax A. *J. Biomol. NMR.* 2010; 48:13–22. [PubMed: 20628786]
36. Darden T, York D, Pedersen L. *J. Chem. Phys.* 1993; 98:10089.
37. Bussi G, Donadio D, Parrinello M. *J. Chem. Phys.* 2007; 126:014101.
38. Parrinello M, Rahman AJ. *Appl. Phys.* 1981; 52:7182–7190.
39. Jha A, Colubri A, Zaman M, Koide S, Sosnick T, Freed K. *Biochemistry.* 2005; 44:9691–9702. [PubMed: 16008354]
40. Scholtz J, York E, Stewart J, Baldwin R. *J. Am. Chem. Soc.* 1991; 113:5102–5104.
41. Shalongo W, Dugad L, Stellwagen E. *J. Am. Chem. Soc.* 1994; 116:8288–8293.
42. David A, Scheurer C, Bruschweiler R. *J. Am. Chem. Soc.* 2000; 122:10390–10397.
43. Schmidt J, Blümel M, Löhr F, Rüterjans H. *J. Biomol. NMR.* 1999; 14:1–12. [PubMed: 21136331]
44. Neal S, Nip AM, Zhang H, Wishart DS. *J. Biomol. NMR.* 2003; 26:215–240. [PubMed: 12766419]
45. Frishman D, Argos P. *Proteins.* 1995; 23:566–579. [PubMed: 8749853]
46. Case D, Darden T, Cheatham Iii T, Simmerling C, Wang J, Duke R, Luo R, Crowley M, Walker R, Zhang W, Case DA, Darden TA, Cheatham Iii TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, et al. University of California, San Francisco. 2008; 32 year.
47. Wang AC, Bax A. *J. Am. Chem. Soc.* 1995; 117:1810–1813.

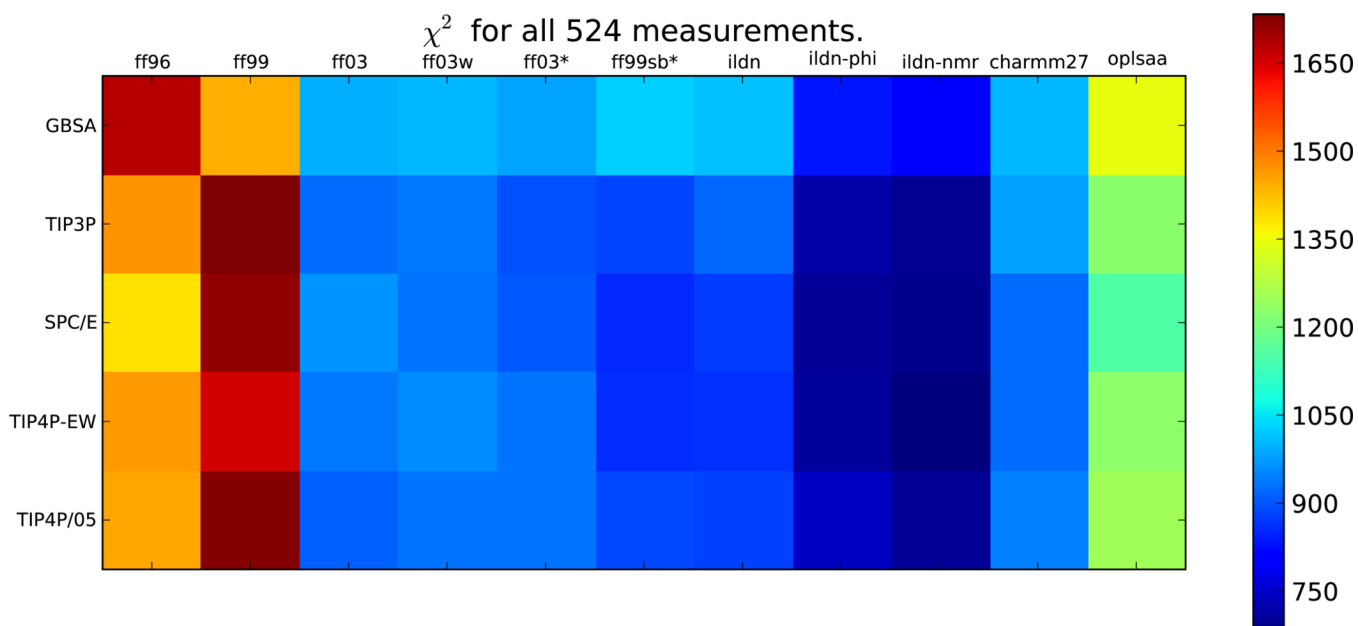
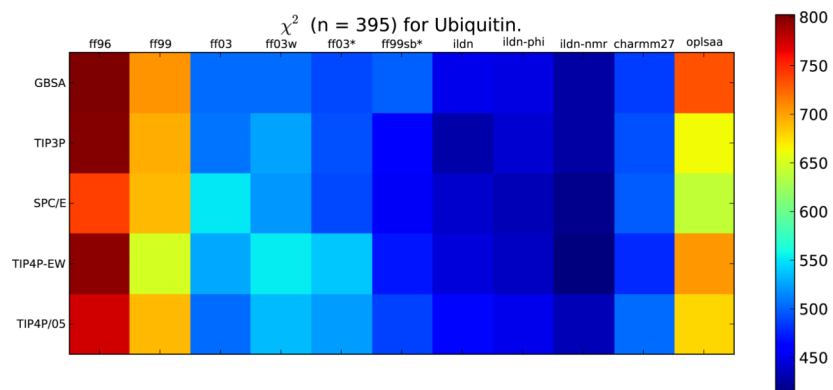
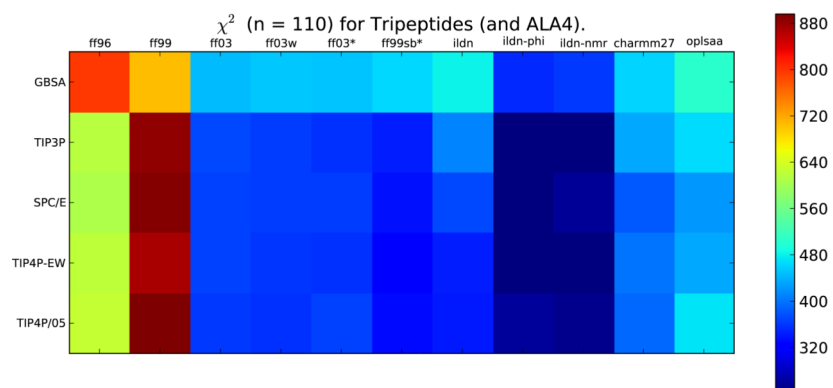


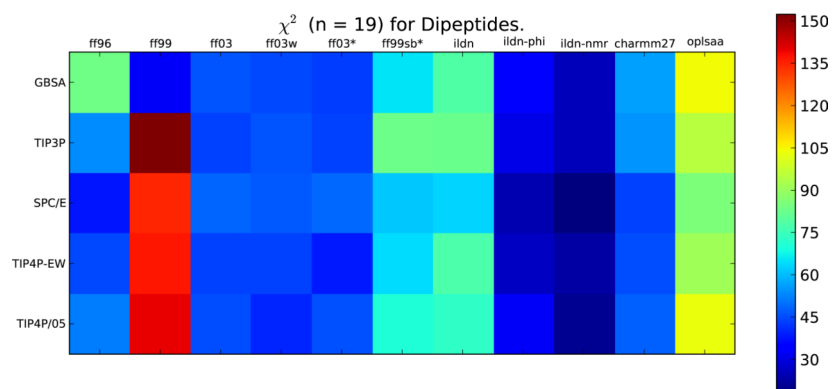
Figure 1.
The overall χ^2 quantifies the agreement with all 524 experimental measurements.



(a)



(b)



(c)

Figure 2. For each class of model system, χ^2 quantifies the agreement between simulation and experiment.

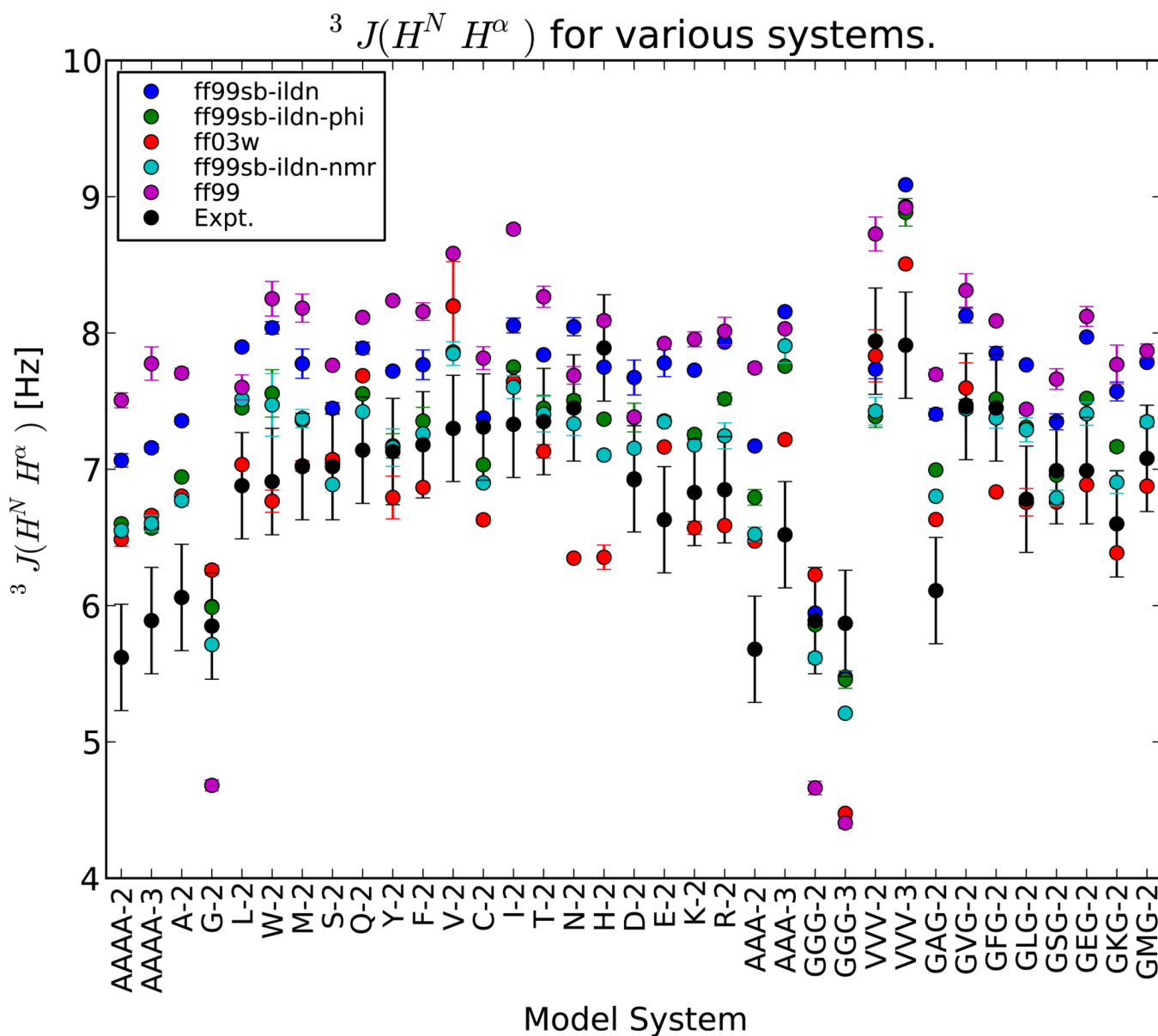


Figure 3. The errors in ${}^3 J(H^N H^\alpha)$ suggest that the ff99sb-ildn-phi and ff99sb-ildn-nmr force fields correct a significant bias in the ϕ potential of the ff99sb-ildn force field. Values are shown for TIP4P-EW.

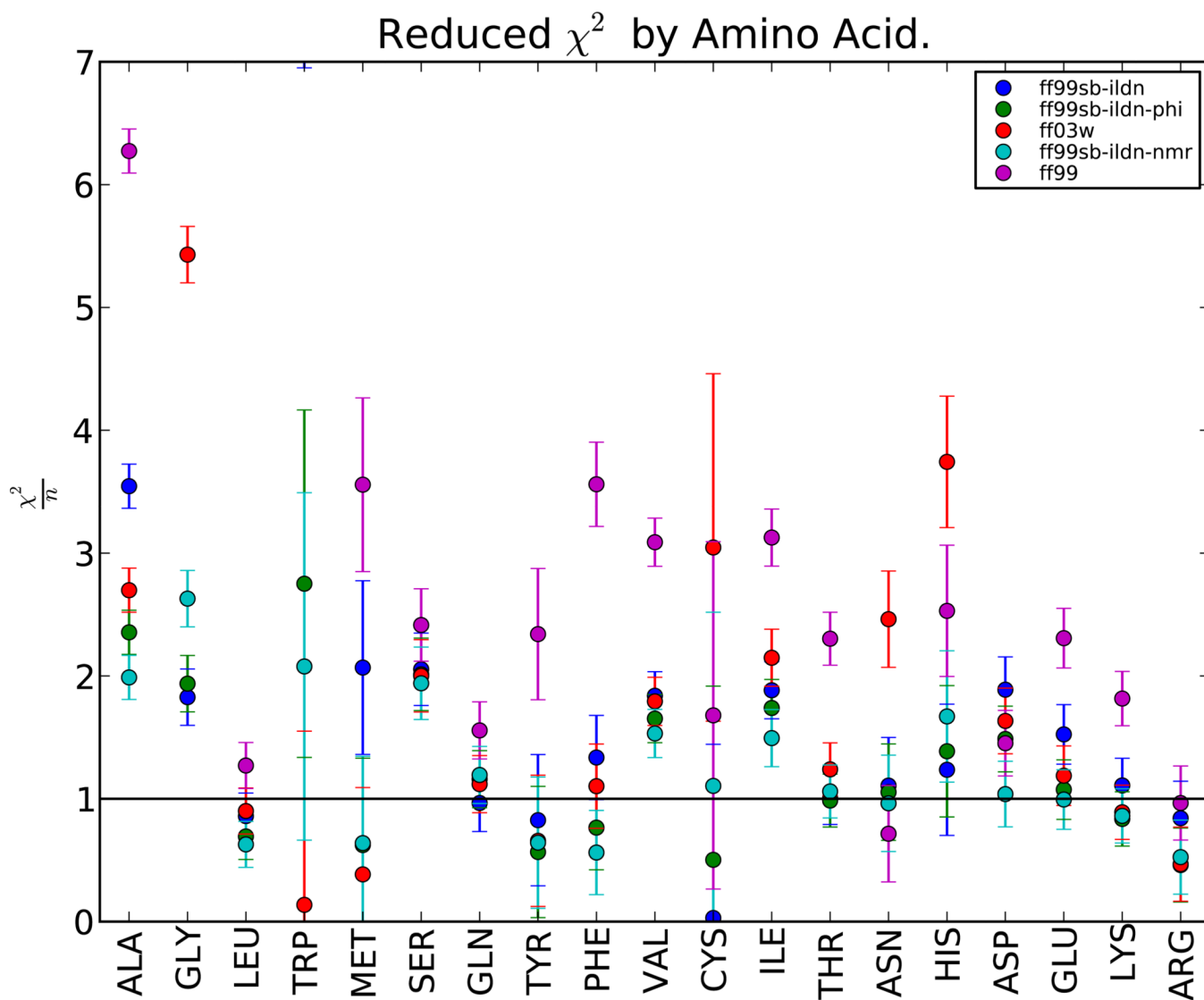


Figure 4.

Reduced χ^2 is shown for all 19 amino acids, indicating force field quality as a function of individual amino acid. Values are shown for TIP4P-EW with five well-performing force fields. Reduced χ^2 values near 1 indicate that force field error is comparable to the experimental uncertainty, while values much larger than 1 indicate possible room for force

field improvements. Errors for reduced χ^2 are given by $\sqrt{\frac{2}{n}}$, where n is number of measurements available for that amino acid. Plotted error bars underestimate the true error, as error estimates include only the contribution of the Karplus and chemical shift prediction. This contribution tends to be the dominant source of error in the present benchmark.

Conformational Populations

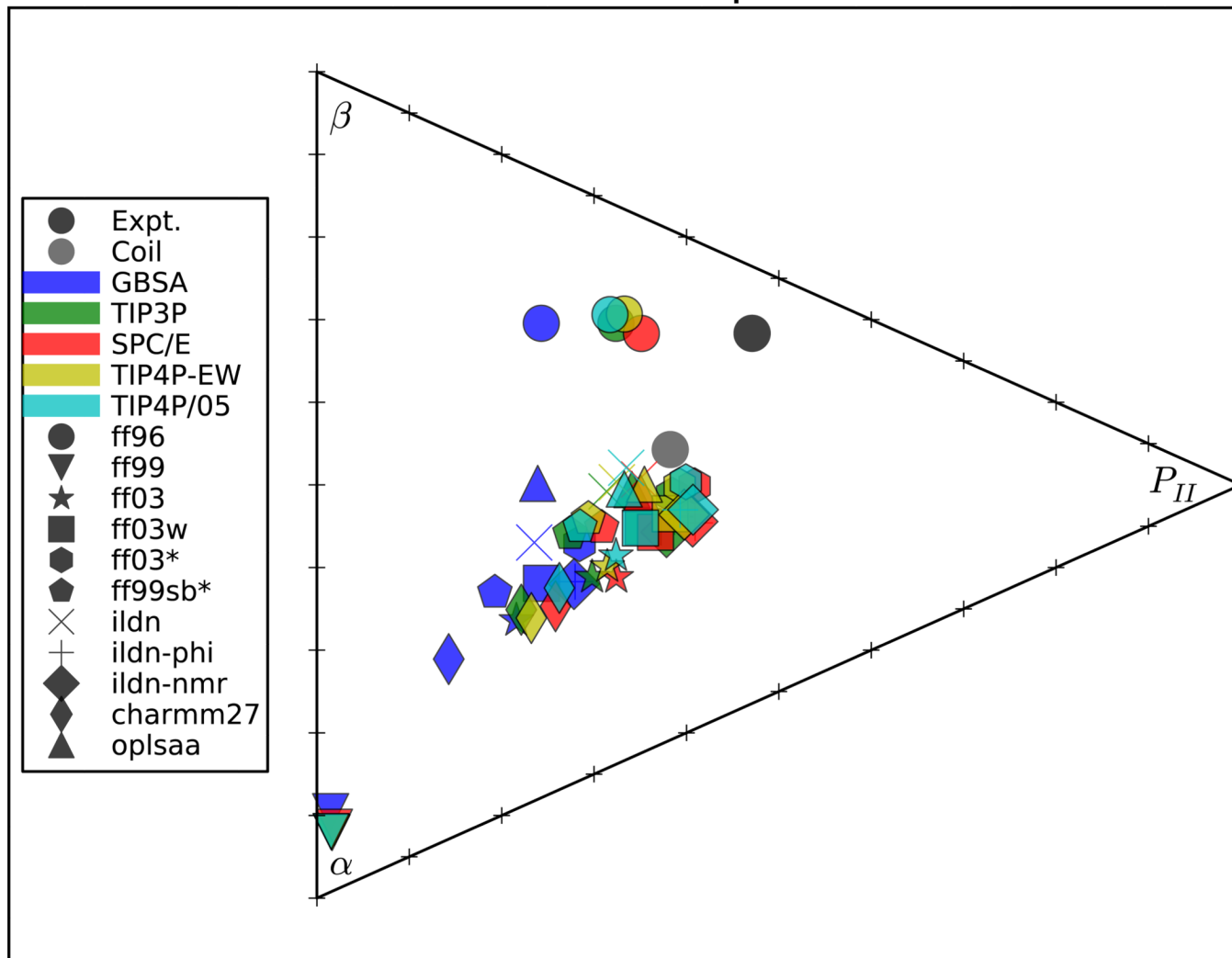


Figure 5. The conformational populations for the 19 dipeptides (averaged over all 19) are shown for various force fields. Individual amino acid predictions are given in Figure S25. Grid ticks represent population increments of 0.1; the corners of the triangle represent the distributions with all β , α_R , and P_{II} , respectively. Also shown are experimental estimates²¹ and statistics from a PDB-derived coil library.³⁹