Inferring Structural Ensembles from Noisy Experiments: Application to Trialanine

Kyle Beauchamp,
Biophysics Program,
Stanford University, Stanford, CA

Rhiju Das
Biochemistry and Applied Physics Departments,
Stanford University, Stanford, CA

Vijay Pande,
Chemistry Department and Structural Biology Program,
Stanford University, Stanford, CA

July 14, 2013

Abstract

Inferring biomolecular conformation from experiment is the fundamental task of structural biology. Due to limited experimental resolution, structure determination often requires a combination of modeling and experiment. The vast majority of algorithms for structure determination, however, are limited to modeling a single conformation and provide only limited uncertainty information. Here we describe Bayesian Energy Landscape Tilting (BELT), a scheme that combines simulation and experiment to infer solution ensembles. Given simple, well-defined assumptions, BELT leverages the machinery of Bayesian statistics to give rigorous uncertainty estimates for structural features and equilibrium properties. Using trialanine as a test system, we show that BELT corrects force field bias, as measured by χ^2 agreement with trialanine measurements *not used* to fit the models. BELT modeling with just six NMR measurements estimates α, β, and PP_{II} conformational populations that are independent of force field.

Introduction

Over the past forty years, structural biologists have solved "ground-state" structures of countless biological macromolecules(1). Modern biology, however, presents many systems that do not fit the single-structure paradigm. Excited states of nucleic acids(2), natively disordered proteins(3), and protein folding intermediates(4) alike are poorly described by single conformation models. For such systems, models of conformational ensembles are required to understand and predict structural and equilibrium properties.

A growing body of research has sought to understand conformational ensembles. Much of it has focused on incorporating dynamical information during NMR structure determination (5, 6) or the extraction of multiple conformers from X-ray diffraction data (7, 8). While these techniques are powerful, they share difficulties in data collection and the unified treatment of heterogeneous experimental data. At the same time, molecular dynamics studies have demonstrated heterogeneity in the atomic-scale structure of macromolecules (9). However, simulation studies suffer from possible inaccuracies in molecular dynamics force fields, as revealed by benchmarks and optimizations of molecular dynamics force fields (10–12). In the present work, we connect these two extremes by combining equilibrium experimental measurements with molecular simulation.

Here we introduce a statistical approach to modeling solution ensembles of biological macromolecules. The algorithm, Bayesian Energy Landscape Tilting (BELT), uses solution experiments to reweight a collection of atomistic models. BELT extends a recent maximum entropy method for restraining simulations (13) to *reweight* existing simulations. Furthermore, BELT leverages Markov Chain Monte Carlo to transform experimental ambiguity into error bars on arbitrary structural features.

To validate BELT, we investigate the conformational propensities of trialanine using NMR measurements (14) and MD simulations performed in five different force fields. Although the raw simulations show wide variations in their conformational preferences, BELT corrects force field errors to provide self-consistent estimates of the α, β, and PP_{II} populations. The ability to correct the biases of all tested forcefields suggests that BELT is a powerful technique for connecting simulation and equilibrium measurements.

Theory: Bayesian Energy Landscape Tilting

Model Inputs

To model an ensemble using BELT requires three components (Fig. 1). First, we need conformations $(x_j)_{j=1}^m$ sampled from the approximate equilibrium distribution of our system. In the present work, such conformations will be generated from molecular dynamics simulations. Second, we require equilibrium experimental measurements $(F_i)_{i=1}^n$ and their associated uncertainties $(\sigma_i)_{i=1}^n$. Third, it is necessary to have a direct connection between simulation and experiment. This connection is achieved by predicting each experimental observable at each conformation: $f_i(x_j)$ is the predicted value of experiment i at conformation x_j.

Reweighting

The next step in constructing an ensemble is to calculate the population of each conformation. For convenience, we work with the log populations (i.e. free energies). Inspired by a previous method for restraining simulations (13) (see Appx. S1), we reweight individual conformations by a biasing potential that is a linear combination of the predicted observables:

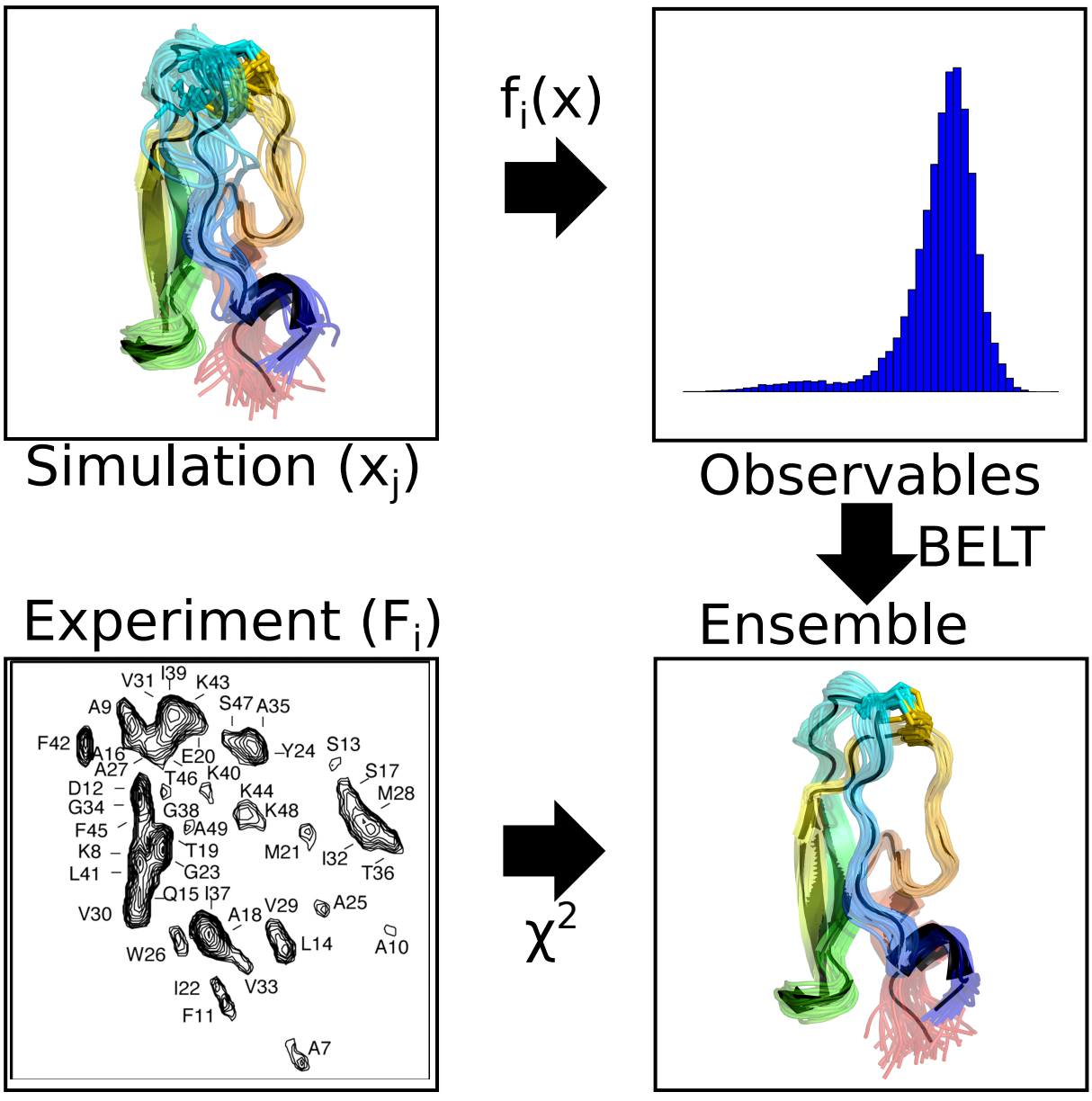
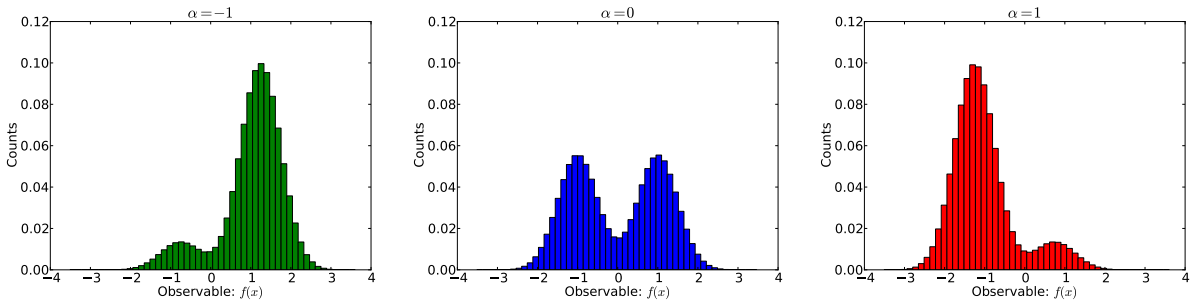


Figure 1: General scheme for BELT modeling.



(a) (b) (c)

Figure 2: Raw ($\alpha = 0$) and reweighted (e.g. tilted) histograms of a one dimensional observable.

$$\Delta U(x;\alpha) = \sum_{i}^{n} \alpha_i f_i(x)$$

In $\Delta U(x;\alpha)$, the parameters α_i determine how strongly each experiment contributes to the biasing potential. One way to think about α is via "tilting" the energy landscape along the order parameters $f_i(x)$. Given the biasing potential, the population of each conformation can be calculated using exponential averaging (see Appx. S2):

$$\pi_j(\alpha) = \frac{1}{\sum_k \exp[-\Delta U(x_k;\alpha)]} \exp[-\Delta U(x_j;\alpha)]$$

It is informative to consider the case of a single observable $f(x)$. Suppose the molecule of interest shows a bimodal observable. If we let $\alpha = 0$, then the biasing potential is 0 everywhere and our reweighted ensemble simply returns the results of the MD simulation (Fig. 2b). If we let $\alpha = -1$, conformations with large values of $f(x)$ are upweighted, while conformations with lower values of $f(x)$ are downweighted (Fig. 2a). Finally, if $\alpha = 1$, the ensemble shifts in the opposite direction (Fig. 2c).

With the equilibrium populations, we can calculate the equilibrium expectations of an arbitrary observable $h(x)$:

$$\langle h(x) \rangle_\alpha = \sum_j h(x_j)\pi_j(\alpha)$$

In the above bracket notation, $\langle h(x) \rangle_\alpha$ is the ensemble average of $h(x)$ in an ensemble that is perturbed by a biasing potential $\Delta U(x;\alpha)$. At this point, the determination of α has not yet been discussed. The key idea, however, is that the α reweighted ensemble $\langle \rangle_\alpha$ should recapitulate the experimental measurements:

$$\langle f_i(x) \rangle_\alpha \approx F_i$$

A likelihood framework

We now derive a likelihood framework for determining the coefficients α used in the biasing potential. Given adequate sampling and self-consistent experiments, there should exist some *true* value of α whose ensemble matches the experimental data. However, experimental uncertainty (σ_i) prevents exact agreement between the measurements and the true ensemble. For the models in the

current work, we model σ_i as the uncertainty associated with predicting chemical shifts and scalar couplings from structures; this dominant error is quantified by the RMS uncertainty estimated during the parameterization of chemical shift and scalar coupling models. We use an independent normal approximation (see Appx. S3) to model the agreement between the α ensemble and the measurements:

$$P(F_i|\alpha) \sim N(\langle f_i(x)\rangle_\alpha, \sigma_i^2)$$

Using Bayes Theorem, we can calculate the posterior distribution of α:

$$P(\alpha|F_1, ..., F_n) \propto P(F_1, ..., F_n|\alpha)P(\alpha)$$

Now we let $LP(\alpha)$ denote the log posterior of α and simplify, dropping terms that are independent of α:

$$LP(\alpha) = \log[P(\alpha|F_1, ..., F_n)] = -\sum_i^n \frac{1}{2\sigma_i^2}(\langle f_i(x)\rangle_\alpha - F_i)^2 + \log P(\alpha)$$

Note the simple form of the log posterior. The first term (i.e. the likelihood) measures the χ^2 agreement between the reweighted ensemble and measurements. The second term is the log of the prior distribution on α.

In the present work, we evaluate three different choices of prior (Appx. S4), finding similar results for each. The first is the maximum entropy prior, which penalizes ensembles as they deviate from the raw simulation results:

$$\log P(\alpha) = \lambda \sum_j^m \pi_j(\alpha) \log \frac{\pi_j(\alpha)}{\pi_j^0}$$

In the previous expression, π_j^0 refers to the populations of an unweighted ensemble, which are typically $\frac{1}{m}$. We also consider using a Dirichlet prior, which is functionally similar to the maxent (Appx. S4):

$$\log P(\alpha) = \lambda \sum_j \pi_j^0 \log \pi_j(\alpha)$$

The third prior we consider is a multivariate normal prior, where $\alpha \sim N(0, \Sigma)$. The value of Σ is given by $\Sigma_{ij} = \lambda Cov(f_i(x), f_j(x))$, as derived in Appx. S4.

These priors can be used to achieve regularization, which is a powerful technique to reduce overfitting (15). Large values of λ favor the raw simulation results (i.e. uniform conformational populations): $\pi_j \approx \pi_j^0 = \frac{1}{n}$. The value of λ can be chosen via cross-validation or other methods (see Appx. S5).

Bayesian Modeling of Structural Ensembles

Because ensemble inference often presents many plausible solutions (16, 17), we avoid statistical methods that return a single solution (e.g. maximum likelihood). We therefore use Markov chain Monte Carlo (MCMC), as implemented in PyMC (18), to sample the distribution of structural ensembles consistent with experiment. The result is an ensemble of ensembles–a statistical ensemble of conformational ensembles. Averaging all MCMC samples provides posterior mean estimates of arbitrary structural features. Similarly, examining the MCMC variances provides statistical

uncertainties of equilibrium or structural features. A Bayesian bootstrapping procedure (19) can also be used to model the statistical uncertainty of the MD simulations (see Appx. S6).

Methods

Molecular Dynamics Simulations

Trialanine was simulated in the amber96(20), amber99(21), amber99sbnmr-ildn(22), charmm27(23, 24), and oplsaa(25) force fields, as previously reported (26). Simulations were performed using Gromacs 4.5 (27) and run at constant temperature (300 K) and pressure (1.01 atm). Each simulation was at least 225 ns long. Conformations were stored every 1 ps.

Chemical Shifts and Scalar Couplings

All NMR measurements in this work refer to experiments (14) probing the central residue of trialanine. We chose to focus on this residue because it will be most robust to pH dependent effects, which may be difficult to model with current force fields.

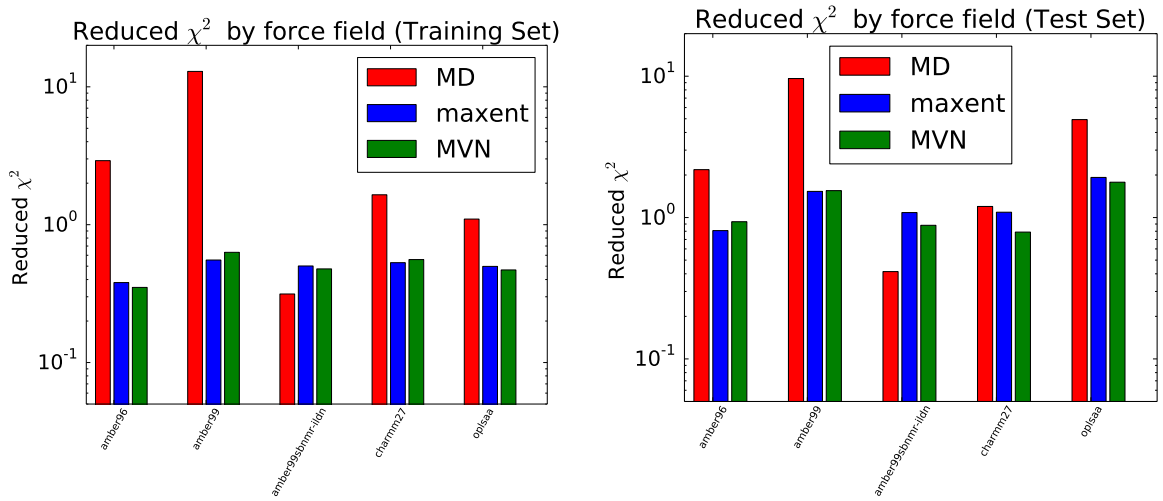
Chemical shifts (H, HA, CA, CB) for each frame were calculated using a weighted average of ShiftX2(28), SPARTA+ (29), and PPM(30) predictions; uncertainties for each model were estimated using their reported RMS prediction errors. Overall uncertainties were estimated as $\sqrt{\sum w_i \sigma_i^2}$, where $w_i \propto \frac{1}{\sigma_i^2}$ is the weight (e.g $\sum_i w_i = 1$) of each chemical shift model and σ_i is the uncertainty of each chemical shift model. The J couplings were calculated using the following Karplus relations: ${}^3J(H^N C')$ (31), ${}^3J(H^N H^\alpha)$ (32), ${}^2J(NC^\alpha)$ (14), ${}^3J(H^\alpha C')$ (31), ${}^1J(NC^\alpha)$ (14), ${}^3J(H^N C^\beta)$ (32). J coupling uncertainties were approximated as the RMS errors reported when fitting the Karplus coefficients.

We have divided the available experimental measurements into a training and test set, with the training set consisting of ${}^3J(H^N C')$, ${}^2J(NC^\alpha)$, ${}^3J(H^N C^\beta)$ scalar couplings and the C_α, the H, and the C_β chemical shifts. The test set consists of ${}^3J(H^N H^\alpha)$, ${}^3J(H^\alpha C')$, and ${}^1J(NC^\alpha)$. The division into training and test sets serves three purposes. First, it provides a test of overfitting. Second, it allows us to reduce the computational cost of BELT calculations. Third, it allows us to train on data that is approximately uncorrelated; BELT is best suited for working with uncorrelated data. Additional suggestions for data curation are provided in Appx. S7.

BELT

All BELT calculations were performed using the FitEnsemble package (`https://github.com/kyleabeauchamp/FitEnsemble`). The online FitEnsemble tutorial demonstrates the use of BELT with a single experimental measurement (${}^3J(H^N H^\alpha)$). Source code for calculations in this work will be made available at `https://github.com/kyleabeauchamp/EnsemblePaper`.

Regularization strength was determined via cross validation as described in Appx. S5. MCMC sampling for each of the current models the current work took approximately 12 hours. These datasets used included six measurements and approximately 250,000 conformations. For each model, we used PyMC to sample at least 5,000,000 values of α; sampled values of α were thinned 100-fold to reduce correlation. The first 5,000 samples (before thinning) are discarded as burn-in. MCMC traces are shown in Fig. S2 and discussed in Appx. S8. To incorporate simulation uncertainty, we used Bayesian Bootstraping (Appx. S6). Two Bayesian bootstrap replicates were performed.



(a) (b)

Figure 3: The reduced χ^2 error (e.g. $\frac{\chi^2}{n}$) for MD and BELT (maxent, Dirichlet, and MVN priors) models. The BELT reduced χ^2 is estimated as the mean reduced χ^2 over all MCMC samples. (a). Calculated using the six measurements used to fit the BELT model. (b). Calculated using four measurements *not* used to fit the BELT model. See Methods for the definition of training and test sets.

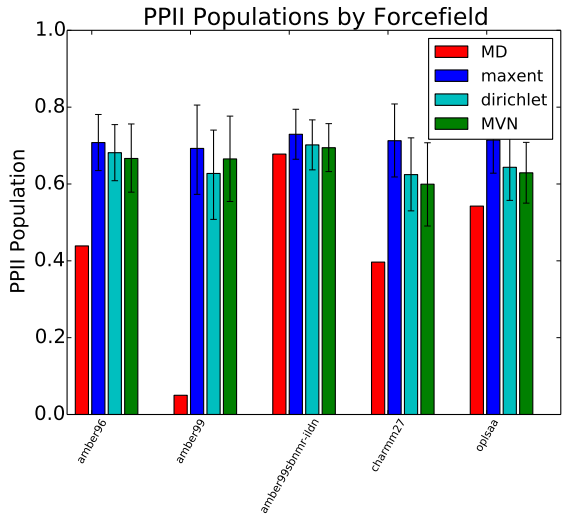
Results

Conformational Propensities of Trialanine

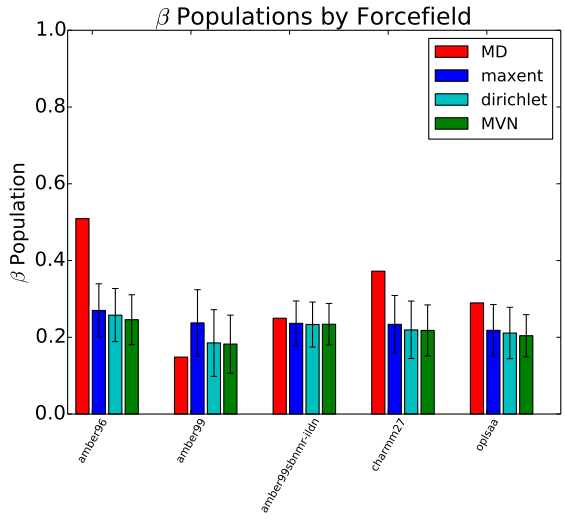
Short peptides provide crucial tests for evaluating and optimizing molecular dynamics force fields (14, 26, 33–35). Such peptides offer a window into the intrinsic conformational propensities of amino acids, free from the secondary structure bias found in statistical surveys of protein structures (36). Here, we use BELT to infer the conformational populations of trialanine from chemical shift and scalar coupling measurements (14).

Trialanine was simulated (see Methods) in five different force fields; six experimental measurements (three scalar couplings, three chemical shifts) probing the central alanine residue were used to construct a BELT ensemble. The five force fields show considerable variation in their agreement with experiment (Fig. 3). The amber96, amber99, charmm27, and oplsaa force fields, for example, initially show significant deviation from the experimental measurements. Upon reweighting, however, all five force fields agree with experiment–including experiments that were *not* used to fit the model (Fig. 3b). Additionally, these results are robust to differences in the prior placed on α.

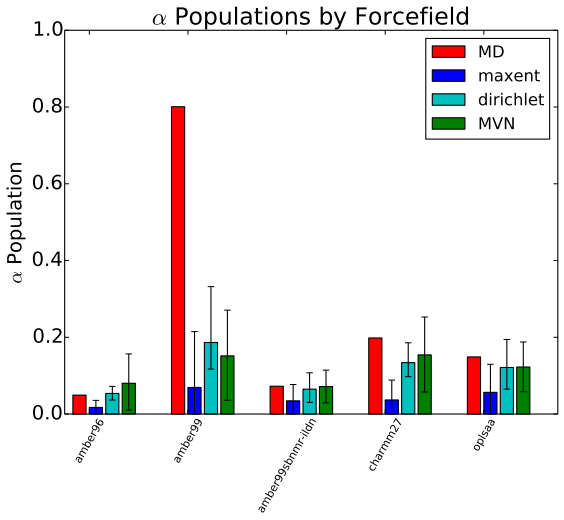
Several independent experimental studies have shown that short alanine peptides prefer the polyproline type helix (PPII) in solution (14, 35, 37). Most molecular dynamics force fields, however, are known to underpopulate the PPII state (14, 26, 33, 34). Our trialanine simulations recapitulate this known deficiency (Fig. 4; red), with amber96 showing a strong β bias (Fig. 4b) and amber99 showing a strong α bias (Fig. 4c). However, combining simulation and experiment leads to conformational ensembles that are robust to differences in force field and prior (Fig. 4; green and blue). Full Ramachandran plots of the (ϕ, ψ) preferences (Fig. 5) reveal similar features in all five BELT models.



(a)



(b)



(c)

Figure 4: MD and BELT (maxent, Dirichlet, and MVN priors) conformational propensities for each force field.

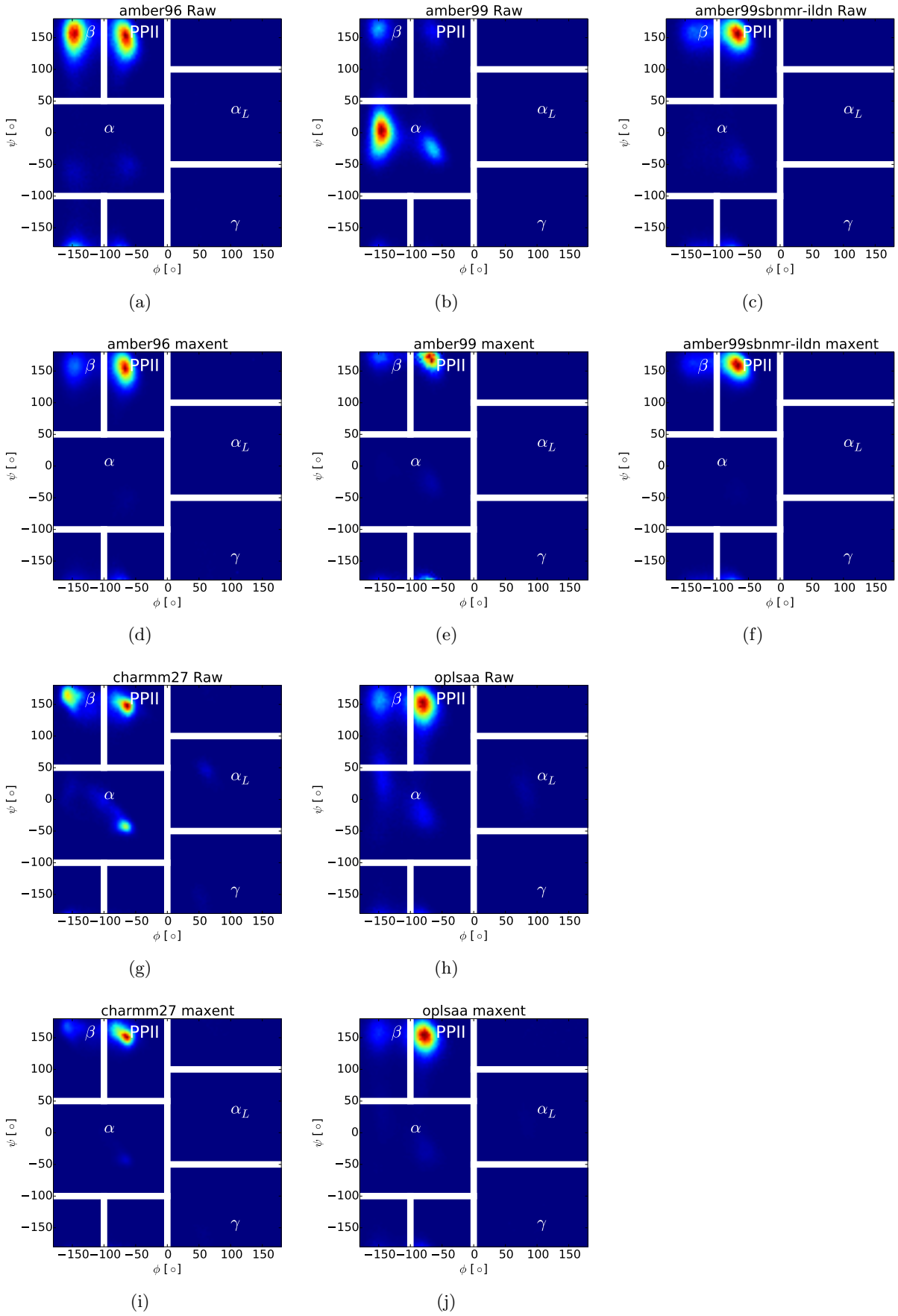
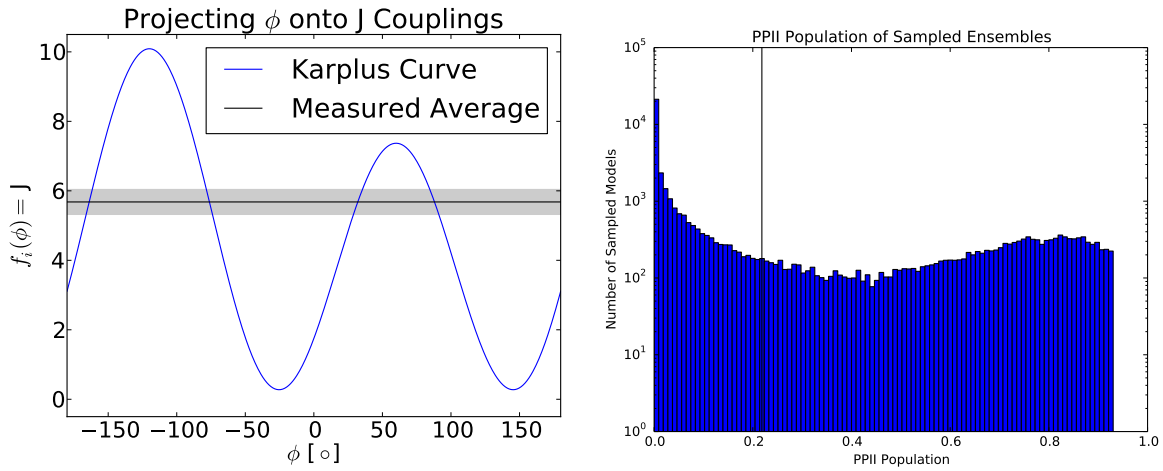


Figure 5: Ramachandran plot of MD and BELT (maxent prior) ensembles for each force field. Comparison of BELT models with different priors are given in Fig. S1.



(a) (b)

Figure 6: (a). The Karplus equation connecting the backbone torsion ϕ to ${}^3J(H_N H_\alpha)$ is ambiguous; observed values of ${}^3J(H_N H_\alpha)$ are consistent multiple conformations. (b). BELT modeling (oplsaa force field) with two measurements (${}^3J(H_N C_\beta)$, C_β chemical shift) shows a bimodal posterior. The mean is shown as a black line; due to the bimodality, the mean value lies near a population *minimum*.

Bayesian Inference allows Ambiguous Experiments

BELT allows the full characterization of posterior distributions using MCMC. Most previous approaches, however, have focused on obtaining point estimates of conformational ensembles (14, 38). In many cases, however, ambiguous experimental data precludes a point-estimate of the conformational ensemble. To illustrate this point, we plot the measured (14) value of ${}^3J(H_N H_\alpha)$ in the context of the Karplus equation relating ϕ to ${}^3J(H_N H_\alpha)$. The measured coupling corresponds to four *different* values of ϕ (Fig. 6a), suggesting that a point estimate is inappropriate for modeling conformations or ensembles. As a concrete example of bimodality, we calculated the posterior distribution of PP_{II} for the oplsaa force field reweighted with two NMR measurements (Fig. 6b). The observed bimodal distribution indicates the need to fully characterize the posterior distribution.

Discussion

Structural (Ensemble?) Biology

Why model structural ensembles, rather than just structures? At least three compelling reasons favor ensembles. First, biological molecules are multi-state machines that fold, unfold, bind ligands, aggregate, and change conformation. Biology is controlled by the relative populations of these states. Ensembles capture aspects of these phenomena by encoding equilibrium populations with structures. A second argument for ensemble modeling is fidelity to experiment. Most solution experiments measure ensemble average equilibrium properties–chemical shifts, scalar couplings, NOEs, SAXS, and FRET are reasonably well-described as equilibrium properties. A truly quantitative connection to these measurements requires modeling the equilibrium ensemble. Finally, recent advances in atomistic simulation (27, 39–41), special-purpose hardware (42), and distributed computing analysis (43, 44) have enabled atomistic simulations to reach the millisecond timescale (9, 45–47); the computational cost of ensemble modeling is quickly becoming manageable.

One might argue that structural ensembles are unnecessary because many proteins occupy a single state under physiological conditions. For such proteins, it is probably safe to enforce single state behavior, as is done in current modeling approaches. However, we suggest that the number of states be *inferred*–not *assumed*.

Resolving Fine Structure may Require Improved Observables

The five BELT ensembles show quantitative agreement in α, β, and PP_{II} populations (Fig. 4). However, the finer details of each Ramachandran plot (Fig. 5) suggest subtle differences between the five models. Because all five BELT ensembles show excellent agreement with experiment (Fig. 3), we conclude that current predictors of chemical shifts and scalar couplings are insufficiently precise to resolve (and falsify) subtle force field differences. The most obvious such difference is the width, shape, and orientation of the PP_{II} basin. Most strikingly, amber96 and oplsaa have PP_{II} basins that are vertically oriented, while amber99, amber99sbnmr-ildn, and charmm27 show diagonally oriented PP_{II} basins. This highlights the need for more sensitive connections between simulation and experiment, such as chemical shift and scalar coupling models.

Comparison to Previous Ensemble Methods

Most previous ensemble modeling efforts involve a protocol with three key ingredients: state decomposition, a χ^2 objective function, and population inference on the clusters. For example, this general recipe describes the approach used in previous analyses of homopeptides (14), the EROS technique for SAXS modeling (38), and the Bayesian Weighting (BW) formalism (16). Note that of these three techniques, BW fully characterizes the posterior distribution via MCMC; below we therefore focus our attention on BW as it is most directly comparable to BELT in scope and purpose.

In our opinion, the primary disadvantage of previous techniques is the need for a state decomposition, which can be defined either by hand or by clustering. Working with a given state decomposition can introduce two different errors, depending on the number and quality of states. In the limit of few states, clustering can overly coarsen the system of interest, possible preventing the model from reproducing multiple experimental observables. At the other extreme, too many states could lead to over-fitting when estimating the many populations required to parameterize the model. This will lead to poor generalization performance and large errors when predicting experiments *not* used to train the model. One symptom of this regime is discontinuity in conformational populations. For example, imagine two nearby conformations at the boundary between two BW states–one conformation on each side of the boundary. In BW, the populations of each conformation could fluctuate dramatically with the corresponding state populations. In BELT, however, the two conformations will have nearly identical populations if the predicted observables vary smoothly.

In contrast, BELT avoids arbitrary state decompositions by projecting simulations onto a basis defined by the predicted experimental observables. The advantage of working in this basis are threefold. First, in BELT, one estimates a single parameter (α_i) for each experimental observable. If the number of experiments is small, as is often the case, the inference problem involves only a few parameters. Second, the predicted observables are a *natural* basis for biophysical calculations. This basis allows direct connection to experiment and often provides insight into the molecular interactions driving biophysical phenomena. For example, the projection onto observables could be used to rationally infer force field parameters, similar to the ForceBalance method (48, 49). Third, in the limit of exact measurements, BELT reduces to a previous (13) maximum entropy approach

(see Appx. S1).

We also point out some surprising differences between BELT and BW-like methods. BW-like methods have the interesting property that the in-state means of features are preserved. More precisely, suppose that $\chi_s(x)$ is the indicator function of a conformational state s. Then in-state averages of the form $< h(x)\chi_s(x) >$ do *not* depend on the reweighted populations. BELT, however, does not preserve the in-state averages; in fact, this property is the direct result of BELT's connection to maximum entropy modeling (see Appx. S1 and ref. (13)). The effect of this property is that the "peaks" of reweighted histograms are slightly shifted relative to the raw MD results, as observed in Fig. 2.

From this discussion, we hope the reader is informed about some of the key differences between BW-like methods and BELT. In some ways, these differences are analogous to the differences between (regularized) linear regression and nearest-neighbors regression (15). The best method for a given system will likely depend on factors including the presence or absence of an obvious few-state decomposition, the number of experimental measurements, and the properties of of the experimental observables.

Future Work

The BELT method can be extended in several ways. We have already worked out some of these extensions. In App. S3, we derive an approximate correction for working with dependent data. Another obvious extension is the use of non-Normal error models. These models can be directly inserted into the current framework by replacing the χ^2 term in the likelihood with some other loss function. More sophisticated models could separately treat the uncertainties associated with predicting observables and the uncertainties of conformations. This would replace the regularization and Bayesian Bootstrapping (App. S6) approaches used herein. Another promising avenue is to combine Bayesian modeling of structural ensembles with Bayesian models for experimental observables. A Bayesian formalism for NMR experiments has previously been developed (17, 50, 51); connecting BELT to these methods may be straightforward.

Conclusion

Bayesian Energy Landscape Tilting allows the simultaneous characterization of structural and equilibrium properties. Through its use of MCMC, BELT is robust to ambiguous experiments and provides rigorous uncertainty estimates. BELT models constructed with a handful of NMR measurements correct significant force field bias and provide generalizable, *force field independent* trialanine ensembles.

Acknowledgements

We thank John Chodera, TJ Lane, Frank Cochran, Pehr Harbury, Xuesong Shi, and Dan Herschlag for helpful discussions.

References

[1] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

[2] Dethoff, E. A., K. Petzold, J. Chugh, A. Casiano-Negroni, and H. M. Al-Hashimi, 2012. Visualizing transient low-populated structures of RNA. *Nature* .

[3] Fink, A. L., 2005. Natively unfolded proteins. *Current opinion in structural biology* 15:35–41.

[4] Korzhnev, D. M., X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson, and L. E. Kay, 2004. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430:586–590.

[5] Lindorff-Larsen, K., R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo, 2005. Simultaneous determination of protein structure and dynamics. *Nature* 433:128–132.

[6] Lange, O. F., N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot, 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471–1475.

[7] DePristo, M. A., P. I. de Bakker, and T. L. Blundell, 2004. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* 12:831–838.

[8] Lang, P. T., H.-L. Ng, J. S. Fraser, J. E. Corn, N. Echols, M. Sales, J. M. Holton, and T. Alber, 2010. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science* 19:1420–1431.

[9] Shaw, D. E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, 2010. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* 330:341–346.

[10] Li, D.-W., and R. Bruschweiler, 2011. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. *J. Chem. Theory Comput.* 7:1773.

[11] Best, R. B., X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *J. Chem. Theory Comput.* .

[12] Lindorff-Larsen, K., P. Maragakis, S. Piana, M. Eastwood, R. Dror, and D. Shaw, 2012. Systematic Validation of Protein Force Fields against Experimental Data. *PloS one* 7:e32131.

[13] Pitera, J., and J. Chodera, 2012. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* .

[14] Graf, J., P. Nguyen, G. Stock, and H. Schwalbe, 2007. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* 129:1179–1189.

[15] Friedman, J., T. Hastie, and R. Tibshirani, 2001. The elements of statistical learning, volume 1. Springer Series in Statistics.

[16] Fisher, C. K., A. Huang, and C. M. Stultz, 2010. Modeling intrinsically disordered proteins with bayesian statistics. *J. Am. Chem. Soc.* 132:14919.

[17] Rieping, W., M. Habeck, and M. Nilges, 2005. Inferential structure determination. *Science* 309:303–306.

[18] Patil, A., D. Huard, and C. J. Fonnesbeck, 2010. PyMC: Bayesian stochastic modelling in Python. *Journal of statistical software* 35:1.

[19] Rubin, D., 1981. The bayesian bootstrap. *The annals of statistics* 9:130–134.

[20] Kollman, P., 1996. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* 29:461–469.

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

[22] Li, D., and R. Bruschweiler, 2010. NMR-Based Protein Potentials. *Angew. Chem.* 122:6930–6932.

[23] Mackerell Jr, A., M. Feig, and C. Brooks III, 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

[24] Bjelkmar, P., P. Larsson, M. Cuendet, B. Hess, and E. Lindahl, 2010. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* 6:459–466.

[25] Kaminski, G., R. Friesner, J. Tirado-Rives, and W. Jorgensen, 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105:6474–6487.

[26] Beauchamp, K., Y. Lin, R. Das, and V. Pande, 2012. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* 8:1409.

[27] Hess, B., C. Kutzner, D. Van Der Spoel, and E. Lindahl, 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

[28] Han, B., Y. Liu, S. Ginzinger, and D. Wishart, 2011. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 1–15.

[29] Shen, Y., and A. Bax, 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* 48:13–22.

[30] Li, D.-W., and R. Brüschweiler, 2012. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *Journal of Biomolecular NMR* 1–9.

[31] Schmidt, J., M. Blümel, F. Löhr, and H. Rüterjans, 1999. Self-consistent 3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles. *J. Biomol. NMR* 14:1–12.

[32] Vögeli, B., J. Ying, A. Grishaev, and A. Bax, 2007. Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *J. Am. Chem. Soc.* 129:9377–9385.

[33] Nerenberg, P., and T. Head-Gordon, 2011. Optimizing Protein- Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* 7:1220–1230. `http://pubs.acs.org/doi/abs/10.1021/ct2000183`.

[34] Best, R., N. Buchete, and G. Hummer, 2008. Are current molecular dynamics force fields too helical? *Biophys. J.* 95:L07–L09.

[35] Grdadolnik, J., V. Mohacek-Grosev, R. Baldwin, and F. Avbelj, 2011. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. U. S. A.* 108:1794.

[36] Jha, A. K., A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick, and K. F. Freed, 2005. Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44:9691–9702.

[37] Avbelj, F., S. Grdadolnik, J. Grdadolnik, and R. Baldwin, 2006. Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 103:1272.

[38] Rozycki, B., Y. C. Kim, and G. Hummer, 2011. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19:109–116.

[39] Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, and D. van der Spoel, 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* .

[40] Eastman, P., M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, and V. S. Pande, 2012. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* 9:461–469.

[41] Eastman, P., and V. Pande, 2010. OpenMM: a hardware-independent framework for molecular simulations. *Comp. in Sci. Eng.* 12:34–39.

[42] Shaw, D., M. Deneroff, R. Dror, J. Kuskin, R. Larson, J. Salmon, C. Young, B. Batson, K. Bowers, and J. Chao, 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51:91–97.

[43] Senne, M., B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schütte, and F. Noé, 2012. EMMA - A software package for Markov model building and analysis. *J. Chem. Theory Comput.* .

[44] Beauchamp, K., G. Bowman, T. Lane, L. Maibaum, I. Haque, and V. Pande, 2011. MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* 7:3412–3419.

[45] Voelz, V., G. Bowman, K. Beauchamp, and V. Pande, 2010. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9 (1- 39). *J. Am. Chem. Soc.* 132:1526–1528.

[46] Bowman, G. R., V. A. Voelz, and V. S. Pande, 2011. Atomistic folding simulations of the five helix bundle protein λ6–85. *J. Am. Chem. Soc.* 133:664.

[47] Lindorff-Larsen, K., S. Piana, R. Dror, and D. Shaw, 2011. How Fast-Folding Proteins Fold. *Science* 334:517–520.

[48] Wang, L.-P., J. Chen, and T. Van Voorhis, 2012. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* 9:452–460.

[49] Wang, L.-P., T. L. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martínez, and V. S. Pande, 2013. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* .

[50] Habeck, M., W. Rieping, and M. Nilges, 2006. Weighting of experimental evidence in macromolecular structure determination. *Proc. Natl. Acad. Sci.* 103:1756–1761.

[51] Habeck, M., W. Rieping, and M. Nilges, 2005. Bayesian estimation of Karplus parameters and torsion angles from three-bond scalar couplings constants. *Journal of Magnetic Resonance* 177:160–165.