

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Development and Application of a Novel Enhanced Sampling Method and Bayesian Analysis for Characterizing Intrinsically Disordered Proteins

Permalink

<https://escholarship.org/uc/item/3101k43k>

Author

Lincoff, James A

Publication Date

2019

Peer reviewed|Thesis/dissertation

Development and Application of a Novel Enhanced Sampling Method and Bayesian Analysis
for Characterizing Intrinsically Disordered Proteins

By

James A. Lincoff

A dissertation submitted in partial satisfaction of the

requirements of the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Teresa Head-Gordon, Chair

Professor Kranthi Mandadapu

Professor David Wemmer

Summer 2019

Development and Application of a Novel Enhanced Sampling Method and Bayesian Analysis
for Characterizing Intrinsically Disordered Proteins

Copyright 2019

by

James A. Lincoff

Abstract

Development and Application of a Novel Enhanced Sampling Method and Bayesian Analysis for Characterizing Intrinsically Disordered Proteins

by

James A. Lincoff

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Teresa Head-Gordon, Chair

Intrinsically disordered proteins (IDPs) are a class of proteins with wide-ranging significance in signaling and disease that do not adopt a dominant folded structure as monomers. Rather, the structures of IDPs in solution are best described as ensembles of conformational states that may range from being fully random coil to partially ordered. This structural plasticity of IDPs is theorized to facilitate regulation of their interaction with other species, as in signal transduction or aggregation of IDPs into ordered fibrils. Characterizing the structural ensembles of IDPs in the free, solvated state is key to understanding the mechanisms of these interactions, and correspondingly the role an IDP species plays in signaling or disease.

The rapid interconversion between conformational states, however, complicates the experimental study of IDPs because most experimental signals report highly averaged information. Computational modeling with validation through comparison to experiment has therefore been a main approach to characterizing IDP structure and dynamics. The focus of my dissertation is on the development of new methods for computational study of IDPs, facilitating better and less expensive *de novo* generation of IDP structural ensembles and improving the metrics used to evaluate the degree of agreement between a simulated ensemble and a set of experimental data.

Despite vast improvements in computational power and efficiency, molecular dynamics (MD) simulations of IDPs for generating conformational ensembles are still limited by the expense of calculations. In Chapter 2 I present the development of a new enhanced sampling method – temperature cool walking (TCW) – and comparison of its performance against a standard method – temperature replica exchange (TREx). The TCW method accelerates the rate of convergence to the equilibrium conformational ensemble with increased sampling acceleration relative to TREx at greatly reduced computational cost.

The second major limitation in MD is the accuracy of the force field. Most classical fixed charge force fields were parameterized using data from folded proteins, and have been thought to be biased to overly collapsed and structured conformations. This has motivated the development of IDP-tailored force fields that sample greater disorder, at the potential expense of the ability to model stabilizing interactions between an IDP and its binding partners. In Chapter 3, I assess to what degree the shortcomings assigned to

standard force fields may be due to insufficient sampling by characterizing the performance of standard and newly modified force fields on the Alzheimer's peptide amyloid- β using both TReX and TCW. We find that with improved sampling, standard and modified force fields produce similar structural ensembles, suggesting that both are appropriate for simulation of the disordered state. In Chapter 4 I present preliminary results building off of this work by characterizing the performance of a polarizable force field modeling a synthetic peptide that demonstrates complete loss of helical content with increasing temperature. Inclusion of polarization effects has been thought to be key for accurate modeling of such multicomponent systems, especially when there is a shift in the electrostatic environment as is the case for the unfolding peptide. Our early results, while limited by current lack of convergence for tests using the polarizable force field and needing further confirmation, match that expectation by finding early evidence of greater response to temperature by the polarizable force field than fixed charge comparators.

The last work presented here is in the development of new methods for calculating the degree of agreement between a simulated IDP ensemble and experimental data. Back-calculation of experimental data from structure can be very imprecise, motivating the development in Chapter 5 of scoring formalisms that account for variable uncertainties in both back-calculation and experiment for diverse experimental data types. In summary, the methods described in this dissertation seek to improve computational study of IDPs by facilitating better, less expensive generation of IDP ensembles and producing more informative metrics for evaluating their agreement with experiment.

To my family

Contents

1. Introduction	1
1.1 Intrinsically Disordered Proteins (IDPs)	1
1.2 Structure Generation Methods for IDP Ensembles	2
1.3 The Combined Force Field-Sampling Problem	3
1.4 Bayesian Comparison of IDP Ensembles to Experiment	4
1.5 References	5
2. Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom	9
2.1 Introduction	9
2.2 Theory	10
2.3 Materials and Methods	13
2.4 Results	14
2.5 Discussion and Conclusions	23
2.6 Acknowledgments	25
2.7 References	25
2.8 Appendix	26
3. The Combined Force Field-Sampling Problem in Simulations of Disordered Amyloid-β Peptides	34
3.1 Introduction	34
3.2 Materials and Methods	37
3.3 Results	39
3.4 Experimental Validation of Simulated Ensembles	46
3.5 Discussion	50
3.6 Conclusions	52
3.7 Acknowledgments	53
3.8 References	54
3.9 Appendix	60
4. Modeling of α-Helix Formation Across a Range of Temperatures by a Polarizable Force Field in Comparison with Diverse Fixed Charge Force Fields	63
4.1 Introduction	63
4.2 Materials and Methods	64
4.3 Results	65
4.4 Discussion and Conclusions	67
4.5 References	67
4.6 Appendix	70

5. A Bayesian Method for Validating Conformational Ensembles with Experimental Data: Application to Ensemble Refinement for Intrinsically Disordered Proteins	71
5.1 Introduction	71
5.2 Materials and Methods	73
5.3 Results	79
5.4 Discussion and Conclusions	90
5.5 References	91
5.6 Appendix	94

Acknowledgments

Perhaps even more difficult than the process of conducting this work over the past years, and describing it in the pages to follow, is the task of expressing the depth of gratitude felt for everybody who has contributed in this last page of writing. There are many people I would like to thank who have supported, motivated, and taught me in far more ways than I could ever fully describe. Thank you. May this be but one small way among many that thanks continues to be given, and through which support is offered in kind.

To my parents Andy and Meiling and to my sisters Nina and Julia—thank you for your guidance and for your encouragement. You all played such a part in getting me ready for the start of this journey, and I have been so lucky to have you close by throughout. Thank you for listening, for keeping me grounded, and for keeping me going.

I would like to thank my advisor Professor Teresa Head-Gordon for guiding me through graduate school and helping me develop as a researcher and educator. Thank you for introducing me to this field, and challenging me to do my best work. I would also like to thank the members of the Head-Gordon lab, for fostering a welcoming and inspiring environment. Thank you to Dr. Sukanya Sasmal, Dr. Saurabh Belsare, and Dr. Sara Cheng, with whom I had the pleasure of collaborating, and to Dr. Alex Albaugh and Dr. Lisa Felberg for conversations in lab, and walks outside of lab. I would also like to thank Professor Julie Forman-Kay, Dr. Mickael Krzeminski, and Dr. Robert Vernon, who have been excellent collaborators in my last year.

There are many other people I have met while at Berkeley who bring so much joy to my life here. Thank you to Amadeus, for always being there for me. Aaron, Bob, Eric, Favio, Mark, and Patrick—thank you for your friendship. I would also like to acknowledge my colleagues and classmates from the 2014 cohort in the Chemical Engineering department. It's been a pleasure to share this experience with you all, and you continue to inspire in and out of the laboratory. Thank you as well to the extraordinary people I had the pleasure of working with in the Graduate Assembly.

To my partner Alex, your love and support have changed my life. Thank you for every moment of peace and comfort at the end of every long day.

Last, I would like to acknowledge the new cohorts of students at Berkeley. You amaze with your passion and brilliance in your work. Here, most of all, I want to pay tribute to your energy and drive in pushing for our department and campus to become more inclusive and more supportive. None of the research we do here is easy on its own, and for so many of us that work is made even more difficult by some circumstance, obstacle, or experience of isolation. To take up the additional burden of naming those obstacles and working to change them for the better of everybody here requires true strength of character and compassion. I will forever remember, be indebted to, and hope to match that strength that is so present and focused in my colleagues and friends here. Thank you.

Chapter 1

Introduction

1.1 Intrinsically Disordered Proteins (IDPs)

Intrinsically disordered proteins (IDPs) are proteins that do not form a stable folded structure in the solvated monomer state¹. Contrary to folded proteins, which tend toward a single or small set of very closely related conformations, the structure of an unfolded IDP must be described as a structural ensemble of conformational states^{2,3} that interconvert on the ns to ms timescale^{4,5}. The rapid dynamics and conformational flexibility of IDPs are theorized to play functional roles modulating the interaction of IDPs with their numerous binding partners⁶, such as folded signal receptors⁷, by driving the IDP to adopt more or less favorable conformations for binding with particular on-off rates⁸. The identification of these functionally relevant conformations, and their mechanisms of interaction, is of high interest due to the role of IDPs when function goes awry as in diseases such as Parkinson's⁹ and Alzheimer's¹⁰⁻¹² amongst others^{13,14}.

The conformational dynamics of IDPs, however, complicate their structural characterization. Whereas numerous techniques such as X-ray crystallography¹⁵ or cryogenic electron microscopy¹⁶ are able to fully resolve the dominant structure for folded proteins—key to our understanding of the mechanisms of their function—there is no analogous single experiment for IDPs that is able to resolve the full ensemble of conformational states. This is because the interconversion between states of an IDP is such that experiments, depending on their timescale, generally only report on highly averaged properties of IDPs^{4,17}. The characterization of an IDP conformational ensemble therefore requires application of multiple solution experiments to gather as much structural information as is possible – using techniques such as nuclear magnetic resonance spectroscopy (NMR)^{18,19}, small-angle X-ray scattering (SAXS)²⁰, and single molecule and solution Förster resonance energy transfer (FRET)^{21,22} – in combination with molecular simulation^{4,17,23}.

The focus of this thesis is on two key aspects of the computational side of IDP ensemble characterization: efficient generation of accurate structural ensembles, and objective metrics for differentiating between simulation ensembles based on the quality and quantity of experimental data. In Chapter 2 I describe the theory, development, and validation of a novel enhanced sampling algorithm for the simulation of the IDP monomer state^{24,25}. I apply the method in Chapter 3 to the simulation of the Alzheimer's peptide amyloid- β ^{11,12}, studying the interplay of sampling efficiency and force field accuracy among standard and newly developed fixed charge models¹⁰. In Chapter 4 I study these fixed charge models alongside a polarizable force field²⁶, evaluating their ability to capture the strong changes in conformational propensity for a synthetic peptide²⁷ in response to temperature. Lastly, in Chapter 5 I describe the extension of a Bayesian scoring metric²⁸ that accounts for multiple sources of uncertainty in evaluating the agreement between a simulated ensemble and a diverse array of experimental data types, and its application to refinement of IDP ensembles. Overall, this thesis describes the development and

application of improved methods for both the generation and validation of IDP ensembles to facilitate better study of IDP structure and behavior.

1.2 Structure Generation Methods for IDP Ensembles

Several approaches, spanning different levels of structural accuracy and computational speed, exist for generating conformational ensembles of IDPs. Toward the lower end of accuracy but with extremely high computational efficiency, statistical random coil ensembles can be constructed with programs such as TraDES²⁹ and Flexible-Meccano³⁰, where the resultant ensemble can be constructed to match expected secondary structure populations as derived from experiment or sequence analysis. Such methods, however, do not naturally incorporate Boltzmann statistics, and may not be able to predict certain local tertiary folding motifs that an IDP may adopt⁴. Given that, these methods are commonly used to generate large trial ensembles of conformers followed by heavy refinement by comparison to experimental data³¹⁻³³.

Fully atomistic molecular dynamics (MD) is generally able to guarantee the highest accuracy of an ensemble, subject to the quality of the potential energy surface used³⁴ and the level of sampling used to characterize it¹⁰. At the most brute force level, ensembles of IDP conformations can be collected over long independent standard MD trajectories³⁵, now commonly run to μs ³⁶⁻³⁹ timescales thanks to advances in computational efficiency such as more widespread use of GPU acceleration⁴⁰ and numerous techniques that reduce the amount of calculation required to propagate the system forward in time⁴¹. Especially when using standard MD, it is important to conduct multiple replicates of simulations⁴² in order to ensure adequate exploration of conformational space, as individual trajectories can become trapped in local energy minima. Unlike random coil ensemble generation, physical simulations have the advantage of generating a Boltzmann-weighted population (with reweighting of the ensemble needed by some methods^{43, 44}) of conformers that matches the energetics of the true equilibrium distribution, subject to the accuracy of the force field³⁴ and extent of sampling¹⁰. Similarly, coarse-grained simulations generate trajectories of Boltzmann-weighted structures at lower cost due to integrating out of unimportant degrees of freedom, though at potentially lower structural accuracy depending on the coarse-graining method and solvent model used¹². This thesis therefore focuses primarily on atomistic MD-based methods for ensemble generation.

Development of a Novel Enhanced Sampling Method

As opposed to proteins that stably fold, where the potential energy surface is dominated by one global minimum that drives the system to that preferred fold, the energy landscape of an IDP in the monomer state is relatively flat^{2, 45}. The main problem encountered during simulation is to ensure that the system transits quickly enough between different potential energy minima to sample the entire conformational space at the appropriate levels. As a result, though it is possible to generate a properly converged ensemble with just standard MD provided enough simulation time and a diverse set of starting conditions³⁵, the use of enhanced sampling methods in IDP simulation is itself fairly standard^{22, 46}.

Enhanced sampling methods refer to the ability to increase the sampling efficiency beyond accessible MD time scales⁴⁷. Depending on the enhanced sampling method, it can require greater computational effort than a single standard MD trajectory of equivalent

length, but the key is to validate that the gain in rate of convergence exceeds any potential additional cost²⁴. Thermodynamically based enhanced sampling methods preclude the direct extraction of dynamical information from the trajectories, but do generate Boltzmann weighted ensembles that are critical to getting IDP sub-population weighting correct. Such methods include Replica Exchange (REx)^{22, 47}, steered MD that incorporates experimentally-derived structural restraints⁴⁸, umbrella sampling⁴⁹, and accelerated MD^{43, 50, 51}.

Temperature replica exchange (TREx)⁴⁷ is among the most commonly used enhanced sampling methods for IDPs, given its generally reliable performance and simple parallelizability. Multiple copies of the simulation are run in parallel each at a different temperature. This temperature ladder usually spans a roughly 200 K range from a low temperature of interest, commonly set to match the temperature condition of an experiment to which data will be compared¹¹, and a high temperature of 400-500 K⁴. At high temperatures, the system exhibits little to no trapping in any conformational minima, rapidly traversing phase space. At regular intervals, exchanges to swap conformations up and down the ladder are attempted such that over time, the low temperature replicas are able to bounce between distant conformational minima more rapidly than would be possible with standard MD. There are two main drawbacks, however, to TREx. In order to maintain efficient passing of conformations up and down the temperature ladder, many replicas at tight temperature spacing are needed, making the method highly expensive⁵². Second, and stemming partially as a result of this need for many closely spaced replicas, the increase in convergence rate relative to standard MD can be somewhat limited²⁴.

In Chapter 2, I present the development and validation of Temperature Cool Walking (TCW)^{10, 11, 24, 25}, a new enhanced sampling method that is designed to overcome these limitations of TREx. TCW is formulated such that, rather than requiring massive amounts of parallel computational power as for TREx, it can be run on single commercial GPUs, using a simulated annealing-like protocol to enhance the rate of convergence of a low temperature walker by introducing conformations that are cooled out of a high temperature ensemble. We validate TCW on a model 1-D potential surface and two small peptides, comparing in each case to optimally run TREx simulations, thereby demonstrating the more rapid convergence and overall lower cost of TCW.

1.3 The Combined Force Field-Sampling Problem

Classical fixed charge force fields have been primarily parameterized to reproduce behavior of folded proteins⁵³, and much attention has been paid to the possible limitations that these force fields have when applied to IDPs^{34, 36, 37}. More specifically, many groups have noted that standard MD and TREx simulations of IDPs using these classical force fields produce conformational ensembles that are compact or exhibit too much folded structure³⁷. As a result, many reparameterized versions of classical force fields have been recently developed to bias them to more expanded, IDP-like conformations^{36, 53-55}.

In Chapter 3 I present a previously published study¹⁰ of the interplay of sampling method and force field, seeking to elucidate to what extent the shortcomings commonly attributed to only the force field may have been due to insufficient sampling as well. We find that upon improving the rate of sampling by shifting from TREx to TCW, a standard force field^{56, 57} was as capable as newly modified force fields^{36, 54, 58} of matching most of the

structural features predicted by NMR scalar coupling experiments^{18, 22} and exhaustive (~200 μ s) standard MD³⁵. The newer IDP-targeted force fields still produce somewhat more extended conformations, which may point to some overall improvement if they are instead compared to FRET and SAXS data.

However, some groups have noted that the modifications made to disorder-inducing force fields render them incapable of maintaining stable folds for other proteins³⁷, which may make them incapable of capturing the stabilization that some IDPs are predicted to undergo when undergoing folding-upon-binding^{7, 12}. An optimal force field for modeling the full process of an IDP moving from a disordered, solvated state to a partially or fully structured bound state must be able to accurately model each of these distinct states.

One possible avenue for this is the use of a polarizable force field²⁶. While significantly more computationally expensive than classical fixed charge force fields, the much greater physical fidelity in the treatment of electrostatics by polarizable force fields has been shown to be important in accurately capturing many physical processes^{59, 60}. In Chapter 4 I present initial results evaluating the applicability of a polarizable force field to a disordered peptide at high temperature that undergoes a sharp transition to occupying helical states at low temperatures²⁷, to test whether a polarizable model²⁶ is better suited to the modeling of such systems that transiently sample a wide variety of degrees of folding.

1.4 Bayesian Comparison of IDP Ensembles to Experimental Data

IDP conformational ensembles are evaluated based on their expected ability to reproduce a range of experimental data⁴. The structural, and in some cases dynamical, information from the simulation is used to back-calculate experimental signals, which are then compared to the actual experimental results. This process is complicated by the uncertainties involved in both experiment and back-calculation²⁸. While an experiment may produce very precise results for the ensemble-averaged signals⁶¹, if the back-calculation process is significantly less precise⁶², multiple structurally diverse ensembles could produce the same back-calculated experimental data to the point that it is impossible, based on that experiment, to validate or refine the ensemble²⁸. Correspondingly many experimental methods lose information on sub-population structure of an IDP due to fast averaging.

In Chapter 5 I describe the development of improved scoring methods for IDP ensembles, using a Bayesian analysis to calculate the log-likelihood of a structural ensemble corresponding to a given set of experimental results, accounting for the uncertainties in both experiment and back-calculation. Having been previously formulated in the group²⁸ for chemical shifts and J couplings from NMR, I expand upon the method by developing scoring algorithms for a wider variety of NMR data including nuclear Overhauser effects (NOEs) and residual dipolar couplings (RDCs) as well as FRET efficiencies²². Such methods that are able to test wide varieties of experimental data that report on various different aspects of IDP structure, with the ability to differentiate between the varying levels of certainty that are possible for each experiment and back-calculation pair, are key for improved characterization of IDP behavior.

1.5 References

1. Dyson, H. J.; Wright, P. E., Intrinsically Unstructured Proteins and Their Functions. *Nat Rev Mol Cell Biol* **2005**, *6* (3), 197-208.
2. Dunker, A. K.; Gough, J., Sequences and Topology: Intrinsic Disorder in the Evolving Universe of Protein Structure. *Curr Opin Struct Biol* **2011**, *21* (3), 379-81.
3. Fisher, C. K.; Stultz, C. M., Protein Structure Along the Order-Disorder Continuum. *J Am Chem Soc* **2011**, *133* (26), 10022-5.
4. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-Beta Peptides. *J Phys Chem B* **2014**, *118* (24), 6405-16.
5. Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T., Finding Our Way in the Dark Proteome. *J Am Chem Soc* **2016**, *138* (31), 9730-42.
6. Wright, P. E.; Dyson, H. J., Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat Rev Mol Cell Biol* **2015**, *16* (1), 18-29.
7. Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D., Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an Scf Ubiquitin Ligase. *Structure* **2010**, *18* (4), 494-506.
8. Liu, Y.; Wu, J.; Sun, N.; Tu, C.; Shi, X.; Cheng, H.; Liu, S.; Li, S.; Wang, Y.; Zheng, Y.; Uversky, V. N., Intrinsically Disordered Proteins as Important Players During Desiccation Stress of Soybean Radicles. *J Proteome Res* **2017**, *16* (7), 2393-2409.
9. Toth, G.; Gardai, S. J.; Zago, W.; Bertoncini, C. W.; Cremades, N.; Roy, S. L.; Tambe, M. A.; Rochet, J. C.; Galvagnion, C.; Skibinski, G.; Finkbeiner, S.; Bova, M.; Regnstrom, K.; Chiou, S. S.; Johnston, J.; Callaway, K.; Anderson, J. P.; Jobling, M. F.; Buell, A. K.; Yednock, T. A.; Knowles, T. P.; Vendruscolo, M.; Christodoulou, J.; Dobson, C. M.; Schenk, D.; McConlogue, L., Targeting the Intrinsically Disordered Structural Ensemble of Alpha-Synuclein by Small Molecules as a Potential Therapeutic Strategy for Parkinson's Disease. *PLoS One* **2014**, *9* (2), e87133.
10. Lincoff, J.; Sasmal, S.; Head-Gordon, T., The Combined Force Field-Sampling Problem in Simulations of Disordered Amyloid-Beta Peptides. *J Chem Phys* **2019**, *150* (10), 104108.
11. Sasmal, S.; Lincoff, J.; Head-Gordon, T., Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-Beta. *Biophys J* **2017**, *113* (5), 1002-1011.
12. Sasmal, S.; Schwierz, N.; Head-Gordon, T., Mechanism of Nucleation and Growth of Abeta40 Fibrils from All-Atom and Coarse-Grained Simulations. *J Phys Chem B* **2016**, *120* (47), 12088-12097.
13. Uversky, V. N., Intrinsic Disorder in Proteins Associated with Neurodegenerative Diseases. *Front Biosci (Landmark Ed)* **2009**, *14*, 5188-238.
14. Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K., Unfoldomics of Human Diseases: Linking Protein Intrinsic Disorder with Diseases. *BMC Genomics* **2009**, *10 Suppl 1*, S7.
15. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181* (4610), 662-6.

16. Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H., Model for the Structure of Bacteriorhodopsin Based on High-Resolution Electron Cryo-Microscopy. *J Mol Biol* **1990**, *213* (4), 899-929.
17. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-Beta Peptides. *Biochemistry* **2011**, *50* (35), 7612-28.
18. Roche, J.; Shen, Y.; Lee, J. H.; Ying, J.; Bax, A., Monomeric Abeta(1-40) and Abeta(1-42) Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* **2016**, *55* (5), 762-75.
19. Roche, J.; Ying, J.; Bax, A., Accurate Measurement of $(^3J)(H\alpha)$ Couplings in Small or Disordered Proteins from Watergate-Optimized Trosy Spectra. *J Biomol NMR* **2016**, *64* (1), 1-7.
20. Kikhney, A. G.; Svergun, D. I., A Practical Guide to Small Angle X-Ray Scattering (Saxs) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett* **2015**, *589* (19 Pt A), 2570-7.
21. Liu, B.; Chia, D.; Csizmok, V.; Farber, P.; Forman-Kay, J. D.; Gradinaru, C. C., The Effect of Intrachain Electrostatic Repulsion on Conformational Disorder and Dynamics of the Sic1 Protein. *J Phys Chem B* **2014**, *118* (15), 4088-97.
22. Meng, F.; Bellaiche, M. M. J.; Kim, J. Y.; Zerze, G. H.; Best, R. B.; Chung, H. S., Highly Disordered Amyloid-Beta Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys J* **2018**, *114* (4), 870-884.
23. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in Beta-Strand Populations of Monomeric Abeta40 and Abeta42. *Biophys J* **2013**, *104* (12), 2714-24.
24. Brown, S.; Head-Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J Comput Chem* **2003**, *24* (1), 68-76.
25. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J Chem Phys* **2016**, *145* (17), 174107.
26. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-Based Amoeba Force Field for Proteins. *J Chem Theory Comput* **2013**, *9* (9), 4046-4063.
27. Merutka, G.; Shalongo, W.; Stellwagen, E., A Model Peptide with Enhanced Helicity. *Biochemistry* **1991**, *30* (17), 4245-8.
28. Brookes, D. H.; Head-Gordon, T., Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. *J Am Chem Soc* **2016**, *138* (13), 4530-8.
29. Feldman, H. J.; Hogue, C. W., A Fast Method to Sample Real Protein Conformational Space. *Proteins* **2000**, *39* (2), 112-31.
30. Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M., Flexible-Meccano: A Tool for the Generation of Explicit Ensemble Descriptions of Intrinsically Disordered Proteins and Their Associated Experimental Observables. *Bioinformatics* **2012**, *28* (11), 1463-70.
31. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W. Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinformatics* **2013**, *29* (3), 398-9.
32. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins* **2012**, *80* (2), 556-72.

33. Marsh, J. A.; Forman-Kay, J. D., Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J Mol Biol* **2009**, *391* (2), 359-74.
34. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmuller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J Chem Theory Comput* **2015**, *11* (11), 5513-24.
35. Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S., Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys J* **2012**, *102* (2), 315-24.
36. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71-73.
37. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc Natl Acad Sci U S A* **2018**, *115* (21), E4758-E4766.
38. Henriques, J.; Cragnell, C.; Skepo, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J Chem Theory Comput* **2015**, *11* (7), 3420-31.
39. Henriques, J.; Skepo, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the Tip4p-D Water Model and the Representativeness of Protein Disorder Models. *J Chem Theory Comput* **2016**, *12* (7), 3407-15.
40. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., Openmm 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659.
41. Margul, D. T.; Tuckerman, M. E., A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J Chem Theory Comput* **2016**, *12* (5), 2170-80.
42. Amini, Z.; Fatemi, M. H.; Rauk, A., Molecular Dynamics Studies of a B-Sheet Blocking Peptide with the Full-Length Amyloid Beta Peptide of Alzheimer's Disease. *Canadian Journal of Chemistry* **2016**, *94* (10), 833-841.
43. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J Chem Theory Comput* **2014**, *10* (7), 2677-2689.
44. Zhu, F.; Hummer, G., Convergence and Error Estimation in Free Energy Calculations Using the Weighted Histogram Analysis Method. *J Comput Chem* **2012**, *33* (4), 453-65.
45. Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M., The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments. *Sci Rep* **2015**, *5*, 15449.
46. Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N. V.; Cote, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; Laio, A.; Li, M. S.; Melchionna, S.; Mousseau, N.; Mu, Y.; Paravastu, A.; Pasquali, S.; Rosenman, D. J.; Strodel, B.; Tarus, B.; Viles, J. H.; Zhang, T.; Wang, C.; Derreumaux, P., Amyloid Beta Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chem Rev* **2015**, *115* (9), 3518-63.

47. Mitsutake, A.; Sugita, Y.; Okamoto, Y., Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60* (2), 96-123.
48. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping Long-Range Interactions in Alpha-Synuclein Using Spin-Label Nmr and Ensemble Molecular Dynamics Simulations. *J Am Chem Soc* **2005**, *127* (2), 476-7.
49. Chen, J., Intrinsically Disordered P53 Extreme C-Terminus Binds to S100b(Betabeta) through "Fly-Casting". *J Am Chem Soc* **2009**, *131* (6), 2088-9.
50. Markwick, P. R.; Bouvignies, G.; Blackledge, M., Exploring Multiple Timescale Motions in Protein Gb3 Using Accelerated Molecular Dynamics and Nmr Spectroscopy. *J Am Chem Soc* **2007**, *129* (15), 4724-30.
51. Pierce, L. C.; Salomon-Ferrer, R.; Augusto, F. d. O. C.; McCammon, J. A.; Walker, R. C., Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* **2012**, *8* (9), 2997-3002.
52. Denschlag, R.; Lingenheil, M.; Tavan, P., Optimal Temperature Ladders in Replica Exchange Simulations. *Chemical Physics Letters* **2009**, *471* (1-3), 193-195.
53. Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. F., New Force Field on Modeling Intrinsically Disordered Proteins. *Chem Biol Drug Des* **2014**, *84* (3), 253-69.
54. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J Phys Chem B* **2015**, *119* (16), 5113-23.
55. Best, R. B.; Zheng, W.; Mittal, J., Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput* **2014**, *10* (11), 5113-5124.
56. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712-25.
57. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J Chem Phys* **2004**, *120* (20), 9665-78.
58. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins* **2010**, *78* (8), 1950-8.
59. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P., Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-up Strategy. *J Chem Theory Comput* **2007**, *3* (6), 1960-1986.
60. Bhowmick, A.; Sharma, S. C.; Head-Gordon, T., The Importance of the Scaffold for De Novo Enzymes: A Case Study with Kemp Eliminase. *J Am Chem Soc* **2017**, *139* (16), 5793-5800.
61. Conicella, A. E.; Fawzi, N. L., The C-Terminal Threonine of Abeta43 Nucleates Toxic Aggregation Via Structural and Dynamical Changes in Monomers and Protofibrils. *Biochemistry* **2014**, *53* (19), 3095-105.
62. Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S., Shiftx2: Significantly Improved Protein Chemical Shift Prediction. *J Biomol NMR* **2011**, *50* (1), 43-57.

Chapter 2

Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom*

We compare two standard replica exchange (REx) methods using temperature and dielectric constant as the scaling variables for independent replicas, against two new corresponding enhanced sampling methods based on non-equilibrium statistical cooling (temperature) or descreening (dielectric). We test the four methods on a rough 1D potential as well as for alanine dipeptide in water, for which the relatively small phase space allows for the ability to define quantitative convergence metrics. We show that both dielectric methods are inferior to the temperature enhanced sampling methods, and in turn that temperature cool walking (TCW) systematically outperforms the standard temperature replica exchange (TREx) method. We extend our comparisons of the TCW and TREx methods to the 5-residue met-enkephalin peptide, in which we evaluate the Kullback-Leibler divergence metric to show that the rate of convergence between pairs of independent trajectories is faster for TCW compared to TREx. Finally we apply the temperature methods to the 42-residue amyloid $\alpha\beta$ peptide in which we find non-negligible differences in the disordered ensemble using TCW compared to standard TREx for simulations of the same length, highlighting the difference in sampling capability for disordered systems of interest.

2.1 Introduction

Enhanced sampling refers to simulation methods that generate configurations that are not easily accessed from standard molecular dynamics (MD) trajectories, but which are needed in order to generate a meaningful statistical average¹. The popular temperature replica exchange method (TREx) utilizes multiple replicas to generate configurations at higher temperatures to facilitate jumping between minima on the potential energy surface of a lower target temperature replica of interest²⁻⁴. Hamiltonian replica exchange (HREx) operates under similar principles and protocols to TREx, but using judicious modification of the Hamiltonian to increase access of the target replica to important local minima^{5, 6}. Overall, the REx approaches have been shown to improve sampling efficiency by increasing the rate of convergence of properties compared to averages accumulated over a standard MD trajectory at the target temperature or Hamiltonian⁷. In addition, they have the attractive feature that tens to hundreds of replicas can be run in parallel to improve computational efficiency on CPUs, although REx is somewhat more cumbersome on GPUs.

Although REx approaches are in practice an improvement over straight MD, we have shown previously that in fact their sampling efficiency can be improved even further⁸. Since REx methods rely on multiple intermediate replicas to more efficiently pass trial

* Reproduced with permission from: Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *The Journal of Chemical Physics*. **2016**, *145* (17).

moves to the target replica, that in turn can lead to diffusiveness in the exploration of configuration space; i.e. in the limit of an infinite number of replicas the swap acceptance probability would be 1, but exploration through parameter (temperature or forces) space would become a random walk. To address this problem, we have previously introduced a non-equilibrium alternative to TReX known as Temperature Cool Walking (TCW)⁸. TCW uses only one high temperature replica to generate non-local trial moves to ensure better ergodic sampling, while also being formulated to satisfy detailed balance to reach the correct limiting distribution appropriate to the lower target temperature. Furthermore, TCW correctly weights the configuration exchange from high temperature at any stage of cooling, thereby finding the sweet spot for optimal overlap with the target temperature distribution to increase the acceptance rates. The TCW algorithm was shown to have superior sampling capabilities compared to TReX on a model problem of a rough 1D potential energy surface for which ergodicity metrics can be analytically quantified⁸.

In this work we apply TCW to a more challenging set of applications of all atom simulations of peptides of increasing size in explicit water. We demonstrate its efficiency and ability to more rapidly converge to the proper limiting distribution when compared to standard TReX as well as compared to two altered Hamiltonian methods⁹, Coulomb REx (CREx) introduced by Itoh and Okumura¹⁰, and Dielectric Walking (DW) which is formulated in the same spirit as the original TCW approach but using the dielectric constant instead of temperature as the annealing variable. In Section 2.2 we describe the non-equilibrium TCW and DW methods in some detail, and computational details of each simulation in Section 2.3. In Section 2.4 we present the results of the various methods applied to a simple 1D potential, as well as the all-atom systems alanine dipeptide^{10, 11}, met-enkephalin¹², and amyloid α -peptides¹³⁻¹⁵, all simulated with explicit water, where TCW is shown to be superior to the other methods. We conclude in Section 2.5 with a discussion and summary of results.

2.2 Theory

Here we review the TCW method⁸ that in turn helps describe the new DW method introduced for the first time here, in which variations in temperature are replaced by variations in the dielectric constant. For TCW, only two explicit replicas are defined, corresponding to simultaneously propagating molecular dynamics (or Monte Carlo) at a high temperature T_H in which better ergodic sampling is ensured, in order to benefit the sampling convergence needed at the target state of low temperature T_L . The TCW algorithm uses a simulated annealing protocol to bring a trial configuration from the high temperature replica close enough to the distribution of the low temperature space to more readily accept the trial move. To formulate an acceptance criterion that satisfies detailed balance, we must perform an equal and opposite simulated annealing *heating* protocol on a low temperature conformation.

Algorithmically, we define a jump probability p_j that defines the frequency at which high temperature configurations will launch a statistical cooling run to serve as trial moves for the low temperature replica. When a random deviate ξ between 0 and 1 satisfies the condition that $\xi < p_j$, the current configuration for the high temperature replica, x_H , and low temperature replica, x_L , are stored. We continue to propagate the high temperature replica during the cooling of the configuration x_H to ensure that, in between successive statistical

cooling attempts, the high temperature replica has sufficient time to fully decorrelate. We then perform simulated annealing on the configuration x_H , propagating it over a schedule of intermediate temperatures T_{H-j} , with $j = 1$ through $H-1$.

We begin by capturing the probability for transitioning x_H from T_H to T_{H-1} , i.e. $P_{cw}(x_H \rightarrow x_{H-1})$, by propagating x_H at this new temperature T_{H-1} for a number of MD steps to generate a configuration x_{H-1} . If we desired to provide a trial move for the low temperature replica from the T_{H-1} replica, the transition process for the first step of cooling, $P_{cw}(x_H \rightarrow x_{H-1})$ is seen to simply be the ratio of the Boltzmann factors $e^{-\beta U(x)}$ for $x=x_H$ evaluated at T_{H-1} and T_H . We then capture the next transition probability $P_{cw}(x_{H-1} \rightarrow x_{H-2})$ for cooling between T_{H-1} to T_{H-2} , by evolving x_{H-1} at the lower temperature T_{H-2} to generate x_{H-2} , etc. Thus at any point in the cooling process, we can define a total transition probability, $P_{cw}(x_H \rightarrow x_{H-j})$ which in general has the form:

$$P_{cw}(x_H \rightarrow x_{H-j}) \propto \frac{P_{H-j}(x_{H-j+1})}{P_{H-j+1}(x_{H-j+1})} \frac{P_{H-j+1}(x_{H-j+2})}{P_{H-j+2}(x_{H-j+2})} \dots \frac{P_{H-2}(x_{H-1})}{P_{H-1}(x_{H-1})} \frac{P_{H-1}(x_H)}{P_H(x_H)} \quad (1)$$

where each $P_q(x_r)$ is the canonical Boltzmann probability of the configuration x_r in the ensemble at temperature T_q .

After the evaluation of every transition probability to the temperature T_{H-j} , we generate another random deviate ξ and compare it to the exchange probability p_E at the temperature T_{H-j} . If the random deviate ξ is less than p_E , we must similarly generate a transition probability for performing the reverse process, i.e. heating from T_{H-j} to T_H in order to satisfy detailed balance. To do so, we take the initial low replica configuration x_L , and evolve it using MD with a temperature of T_{H-j} to generate a configuration x'_{H-j} . We use the notation x' to describe the configurations whose temperature is increased from their original replica. The temperature of the x'_{H-j} configuration is then increased to T_H using the same schedule and process as used for cooling, ultimately yielding configuration x'_H . This allows us to define a general transition probability for heating $P_{cw}(x'_{H-j} \rightarrow x'_H)$:

$$P_{cw}(x'_{H-j} \rightarrow x'_H) \propto \frac{P_{H-j}(x'_{H-j+1})}{P_{H-j+1}(x'_{H-j+1})} \frac{P_{H-j+1}(x'_{H-j+2})}{P_{H-j+2}(x'_{H-j+2})} \dots \frac{P_{H-2}(x'_{H-1})}{P_{H-1}(x'_{H-1})} \frac{P_{H-1}(x'_H)}{P_H(x'_H)} \quad (2)$$

that allows us to formulate an acceptance criteria that satisfies detailed balance.

Before we define the acceptance criteria, we note that we can further increase the acceptance rate by defining a window of states protocol that evaluates transition probabilities between volumes of phase space as opposed to points in phase space to minimize "bad" fluctuations that diminish acceptance rates. To describe the window of states component, we consider an intermediate temperature T_{H-j} from which we attempt to impose the configuration x_{H-j} onto the low temperature replica. From the cooling trajectory we accumulate the following window of weights

$$W(x_{H-j}) = \sum_{k=1}^m e^{-\beta U_{H-j}(x_{H-j}^{(m)})} \quad (3)$$

and simultaneously for the heating trajectory

$$W(x'_{H-j}) = e^{-\beta U_{H-j}(x_0)} + \sum_{k=2}^m e^{-\beta U_{H-j}(x'_{H-j}(k))} \quad (4)$$

where the indices k through m in each summation refer to the individual conformations sampled. The two transition probabilities in Eqs. (1) and (2), as well as the two weighting factors in Eqs. (3) and (4), are combined in a standard Metropolis acceptance rule for imposing the cooled configuration x_{H-j} onto the low temperature replica, thus replacing the original configuration x_L

$$acc(x_L \rightarrow x_{H-j}) = \min \left[1, \frac{P_{H-j}(x_L)P_L(x_{H-j})P_{cw}(x_H \rightarrow x_{H-j})W(x_{H-j})}{P_L(x_L)P_{H-j}(x_{H-j})P_{cw}(x'_{H-j} \rightarrow x'_H)W(x'_{H-j})} \right] \quad (5)$$

In summary, the acceptance rule has three main parts: first, the simple Metropolis criterion for exchanging x_L with x_{H-j} ; second, the ratio of transition probabilities for moving from temperature T_H to T_{H-j} and the reverse process; and finally the ratio of the window of states data.

If the move is accepted based on Eq. (5), we exit the cooling cycle and continue propagation of the two replicas, after updating the configuration of the low temperature replica to x_{H-j} . If the move is rejected, we continue by cooling x_{H-j} to the next temperature T_{H-j-1} and continue through the cooling cycle until either an exchange is accepted or the cooling schedule completes without an exchange, at which point we return to propagation of the two replicas until the next CW cycle is attempted. We find it optimal to set p_E such that at least one exchange attempt is performed per CW cycle, on the order of 3 – 10%. Setting p_E too high increases computational expense, as a new heating cycle must be performed for every exchange attempt during a cooling cycle, whereas setting p_E too low can result in a waste of a cooling trajectory if insufficient exchange attempts are made. Note that the window of states weighting is not necessary for the maintenance of detailed balance and does not have to be used, although we have found it to improve acceptance rates; addition of the window of states protocol was found to decrease the amount of annealing required by 25 % in the original TCW paper on a simple 1D potential.⁸

Acceptance ratios for TCW moves tend to increase with more stringent schedules of the simulated annealing protocol used, and can readily achieve acceptance rates of 20 – 30% or greater. Depending on the size of the system, we find a range of 25 – 50 fs of annealing at each temperature in the schedule to be sufficient for adequately equilibrating configurations to the lower temperatures as they are annealed. Given this, and the fact that every exchange results in imposition of a configuration that originated from the high temperature replica on the low temperature replica (as compared to REx, which requires many successive exchanges between intermediate replicas to “pass” a configuration all the way from the highest to the lowest temperature), p_j can be much lower in TCW than TREx, typically around 0.1 – 0.5 ps⁻¹. We have found that a ratio of 8 fs of additional high temperature replica propagation for decorrelation per 1 fs of cooling gives us the best increase in rate of convergence, though lower ratios are also workable.

We emphasize that TCW does not perform a complete configuration “exchange” as in standard REx protocols; instead we impose the annealed configuration from the high temperature replica onto the target replica, without imposing the configuration of the low temperature replica onto the high temperature replica. This assumes that the high

temperature replica is an infinite reservoir of configurational states that decorrelate rapidly on the timescale between swap attempts. As we will see this is a good approximation for TCW, provided the maximum temperature is adequately high, but needs to be reevaluated when the CW procedure is applied to other system parameters, such as modifying the Hamiltonian in our DW method.

In particular, what we have outlined in Eqs. (1)-(5) for TCW applies to a modification where temperature is fixed, but instead the Hamiltonian changes to define a series of dielectric constants that scales the permanent electrostatics (DW), allowing the protein to exit minima and pass through maxima caused primarily by electrostatic forces. In this case, the “ergodic sampling” replica could correspond to a higher dielectric than the target replica, which is decreased as we move x_H through annealing to x_{H-j} . Finally we can in principle also combine the TCW and DW approach (TCW-DW). In this method, the non-physical replica has both increased temperature and altered protein dielectric, in which both decrease when moving the configuration x_H to x_{H-j} . If modification of the protein dielectric is indeed a viable way to increase sampling, when it is combined with temperature we might anticipate an increase in rate of convergence beyond that of standard TCW.

2.3 Materials and Methods

We have implemented TCW, DW, and TCW-DW into the OpenMM software package so that others can access the methodology presented here; we have also implemented the TReX and CREx methods in OpenMM in order to facilitate direct comparison. The methods have been tested on a number of systems presented here including a 1D potential^{8,16}, as well as atomistic simulations of alanine dipeptide^{10,11}, the 5-residue met-enkephalin peptide¹⁷, and the 42-residue A β 42 peptide¹³⁻¹⁵, all in explicit water.

Our first model system is a 1-dimensional potential energy function defined as

$$V(x) = \sum_{n=1}^{20} C_n \sin\left(\frac{2n\pi x}{L}\right) \quad (6)$$

where the coefficients are chosen on the interval [-1, 1], and the length of the simulation box in reduced units is $L = 10$. The coefficients are the same as used in the original papers^{8,16}, but are given in Supplementary Table 1 with a corresponding plot of the potential energy (Supplementary Figure 1) provided for the benefit of the reader. For this simple system we have used a hybrid Monte Carlo (HMC) method to propagate sampling on its energy surface, which is described in the original TCW paper. For DW we use only two replicas which are propagated by HMC at values of $\epsilon=30$ and $\epsilon=1$, and then trial moves are generated for the low dielectric replica using a (roughly) geometric schedule for a statistical unscreening process ($\epsilon=30$, $\epsilon=29$, $\epsilon=24$, and $\epsilon=1$) and using a similar geometric schedule for the screening process ($\epsilon=1$, $\epsilon=3$, $\epsilon=10$, $\epsilon=30$) needed for detailed balance. For CREx, we found 4 replicas to be optimal, with a geometric schedule of dielectrics replicas ($\epsilon=1$, $\epsilon=3$, $\epsilon=10$, $\epsilon=30$). All HMC parameters and swap attempt rates that we test for the DW and CREx methods are the same in the comparison, and the DW and CREx methods differ only in the number and the way in which replicas interact.

We also consider three successively larger peptides in explicit water to evaluate the TCW, DW and CREx methods, to see whether they actually perform better than the standard TREx protocol. For alanine dipeptide, we use the Amber ff99sb force field¹⁸ for the protein and TIP3P for the water model¹⁹ to compare to previous studies^{10,11} that use this combination of force fields. For met-enkephalin, we use the same combination of force field and water model, again to compare to previous studies¹⁷. For A β 42 we used the Amber ff99sb force field¹⁸ and TIP4P-Ew water model²⁰ since we have published results¹³⁻¹⁵ using TREx with this force field combination. All simulations used a 1 fs time step with SHAKE constraints to freeze out hydrogen vibrations, while an Andersen thermostat maintained the temperature. Ewald was used for calculating long-range electrostatic forces, with a cutoff of 9.5 Å for the real space electrostatics and Lennard-Jones forces. The replica exchange attempt frequencies were once per 1 ps for alanine dipeptide and once per 500 fs for met-enkephalin and A β 42. For alanine dipeptide and met-enkephalin, we used the LEaP module to prepare the initial extended structure of the peptide, and then OpenMM to solvate it within a periodic cube of water appropriate to the system size (233 and 499 water molecules for alanine dipeptide and met-enkephalin respectively.) Using these starting states we equilibrated the temperature or dielectric replicas for 100 ps, and generated multiple independent production runs of 50-100 ns. For A β 42, we used a pdb file generated from TREx simulations previously run in our lab, and it includes three sodium ions to neutralize charges on the peptide within the box. The initial structure used for A β 42 is shown in Supplementary Figure 2.

2. 4 Results

One-Dimensional Potential We first consider the performance of CREx and DW on a rough one-dimensional energy surface, which we have done previously for TCW and TREx.⁸ This simple model system has proved useful to confirm detailed balance of implemented code and to measure the sampling efficiency, i.e. the time required to reach convergence as measured by the ergodicity factor, $\chi(t)$

$$\chi^2(t) = \int_0^L dx [\rho(x, t) - \rho_{\text{exact}}(x)]^2 \quad (7)$$

where

$$\rho_{\text{exact}}(x) = \exp(-\beta V(x)/\epsilon)/Z \quad (8)$$

and ϵ is the relevant dielectric constant and Z is the configuration integral, and in which the exact probability distribution is analytically solvable. Figure 1 plots $\chi_{1D}(t)$ vs. time for the 1D case, in which it is evident from the much more rapid decrease of $\chi_{1D}(t)$ to zero that the DW algorithm converges more quickly to the correct limiting distribution compared to the CREx method, and once again illustrates that any replica exchange approach (temperature or Hamiltonian) can be plagued by problems of intermediate replicas that retard the movement of configurations generated by the most ergodic replica from reaching the target replica.

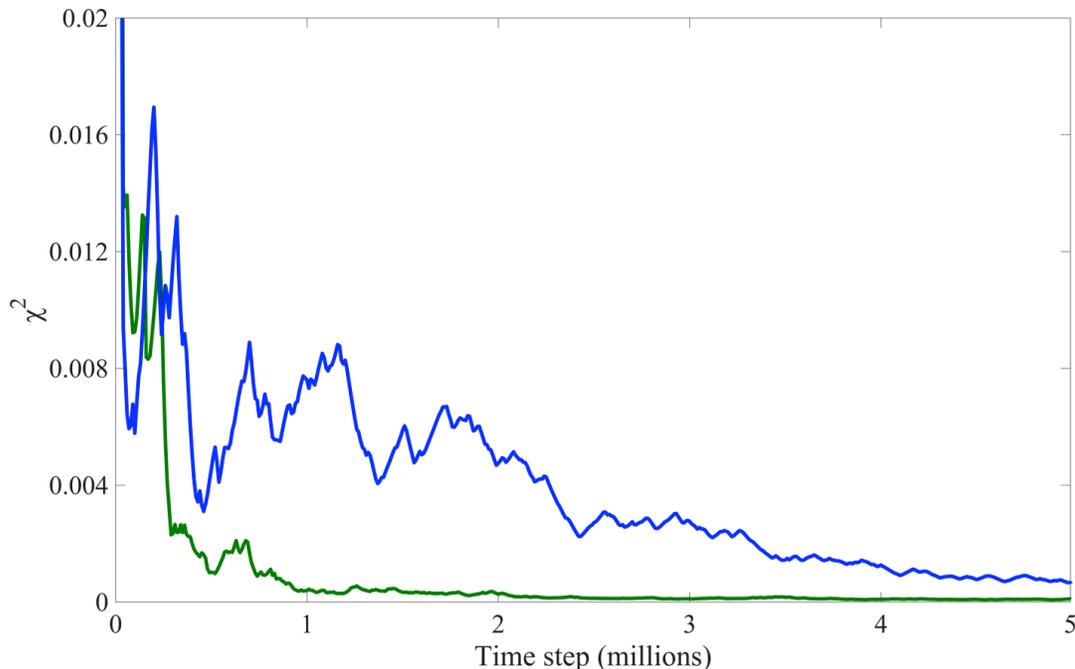


Figure 1: Ergodicity measure $\chi^2(t)$ versus number of MD steps performed by the low-temperature walker for CREx (blue) and DCW (green).

Alanine Dipeptide We next consider the alanine dipeptide in explicit solvent as the first fully atomistic test of our TCW and DW approaches, and compare their performance to TREx and CREx. Given the relatively small conformational space defined by the ϕ and ψ dihedral angles of the dipeptide, we make the assumption that two independent microsecond long standard MD trajectories, using the same force field, will serve as an exact benchmark for testing the rate of convergence of the four sampling methods, since it should be a sufficient amount of time for the peptide to fully cover its conformational space. A good first test of the different enhanced sampling methods is to access the relatively favorable left-handed α -helix ($\phi = 60^\circ$, $\psi = 50^\circ$) that must overcome barriers of ~ 10 - 15 $k_B T$ from the right-handed α -helix ($\phi = 60^\circ$, $\psi = 50^\circ$), extended- β ($\phi = 180^\circ$, $\psi = 180^\circ$), or the polyproline II ($\phi = 150^\circ$, $\psi = 150^\circ$) conformations. This is an interesting test case for the dielectric methods in particular since the barrier is electrostatic in origin, i.e. in which the two peptide oxygens need to come into close contact to execute the transition.

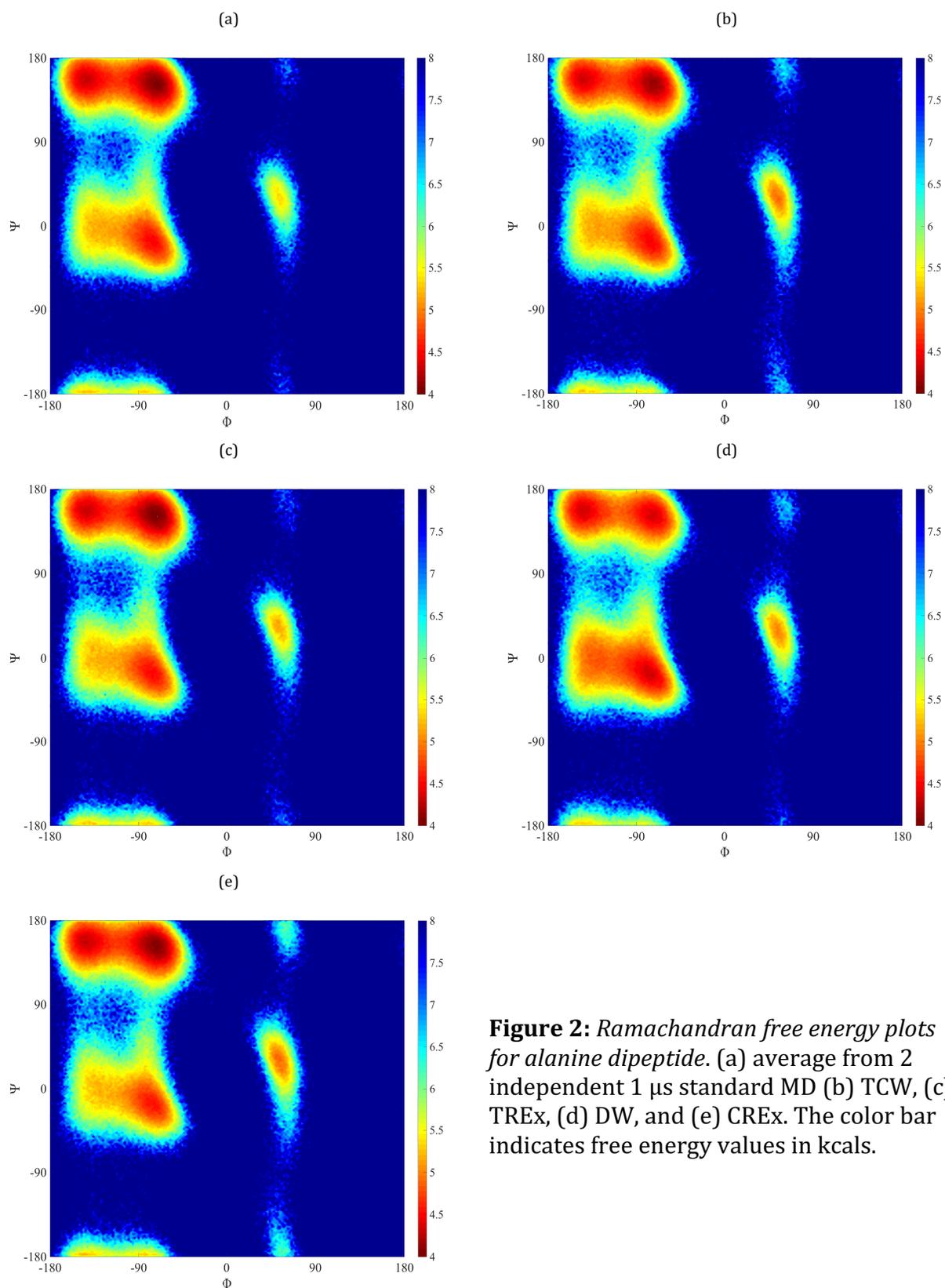
For TREx and CREx, the initial spacing of temperature and dielectric replicas, respectively, were the same as that published in the CREx paper¹⁰. However we found that the temperature replica spacing was non-optimal, and put TREx at a significant disadvantage compared to CREx. Therefore we adjusted the maximum temperature and replica spacings to obtain more standard acceptance rates of $\sim 20\%$ for configurational exchanges between replicas using TREx (Supplementary Table 2). In both cases, replicas were run for 50 ns each. For TCW, the maximum temperature was set to be the same as that for TREx and followed an annealing schedule that was largely the same as the TREx replica spacings. For DW, the maximum peptide dielectric was not the same as that for CREx, which was infinity, but instead it was set to 16; repeated testing of DW did not

demonstrate any significant difference once the maximum dielectric was set at or above this value. The “descreening” schedule used for DW closely followed the same replica spacings used in CREx. The two replicas for TCW and DW were also run for 50 ns each. Replica spacings and all exchange probabilities for TREx and CREx, as well as cooling and heating schedules for TCW and DW, are presented in Supplementary Table 2.

Figure 2 shows the free energy plots for the alanine dipeptide in which it is evident that all 4 methods qualitatively reproduce the result generated from the exhaustive MD benchmark. To measure the rate of convergence, the Ramachandran plots are discretized into a 73 x 73 grid, and the square root of the total sum of squared differences between the Ramachandran free energy values, A , of the enhanced sampling techniques over time relative to our reference MD Ramachandran plot is calculated

$$\chi_{AD}(t) = \sqrt{\frac{1}{N} \sum_i (A_i(t) - A_{i,MD})^2} \quad (9)$$

and normalized such that it varies between 0 and 1. Five independent trajectories of each kind of enhanced sampling method were run to obtain a measure of the average performance for the $\chi_{AD}(t)$ metric for alanine dipeptide. Figure 3 shows the convergence profile over the 50 ns for each method. The TCW method clearly converges the fastest, reproducibly dropping below a $\chi_{AD}(t)$ value of 0.1 within the first 10 ns of simulation, while TREx is noticeably more slowly convergent and variable between trajectories. The DW and CREx methods are roughly competitive with each other, and exhibit significantly greater variance in their rate of convergence depending on the trajectory compared to the temperature methods. We attempted multiple variations of the DW protocol, including shaping the schedule differently, or having both replicas run unperturbed but have the peptide dielectric increase and then decrease back to one through the course of the annealing schedule. Neither resulted in improvement over the published data here. We also combined DW with TCW to see if the additional alteration of the temperature would improve the rate of convergence relative to standard DW, but observed no significant improvement relative to TCW. We determine at this stage that the CREx and DW approaches are not competitive to the temperature enhanced sampling methods, even for problems dominated by electrostatic barriers, and we move forward on only testing TCW and TREx on successively larger all-atom systems. We return to this point in the Discussion.



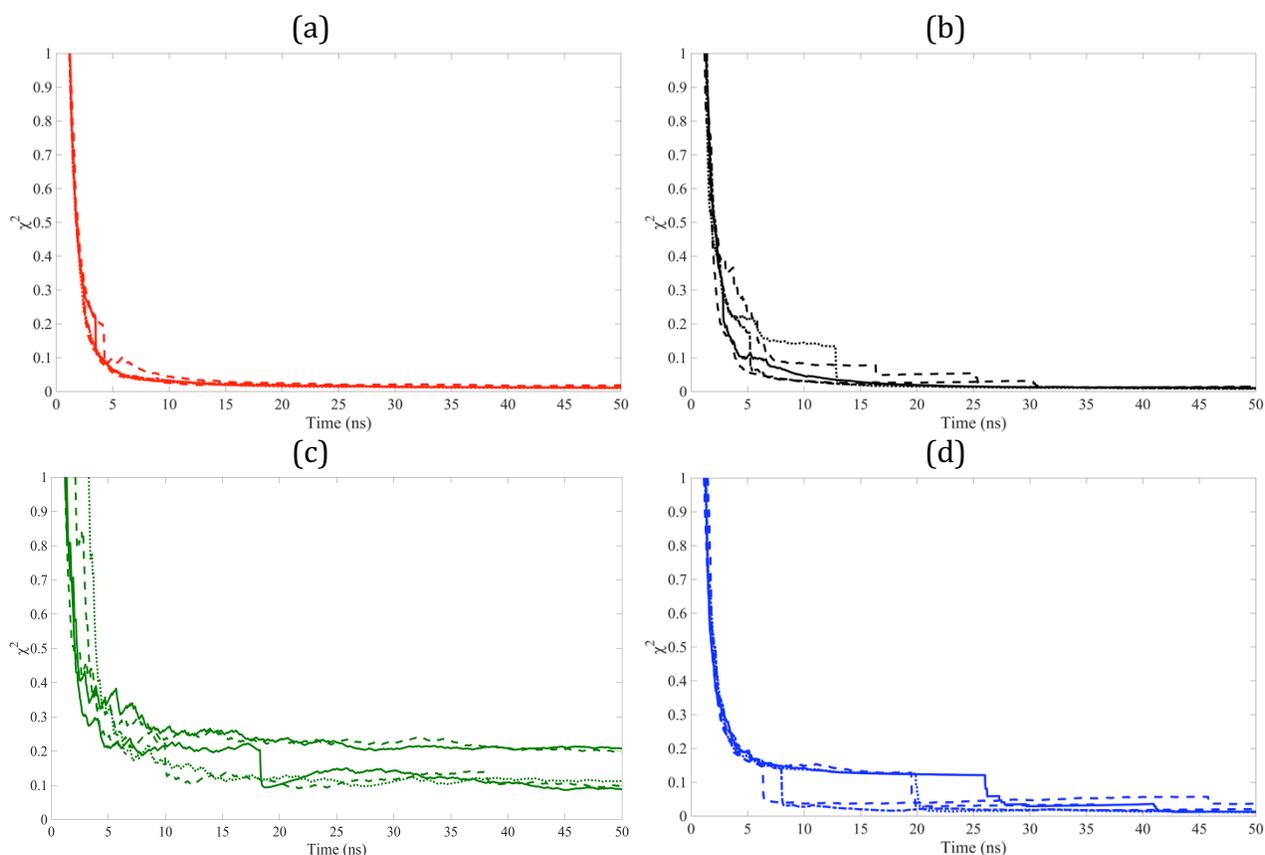


Figure 3: Numerical ergodicity measure $\chi(t)$ versus t (in units of the number of MD steps performed by the low-temperature walker) for convergence of the 4 advanced sampling methods to the 2 μ s MD simulations performed on the alanine dipeptide (a) TCW (red), (b) TReX (black), (c) DW (green) and (d) CRex (blue).

Met-Enkephalin We next compare TReX and TCW on the met-enkephalin peptide. The temperature schedule for TCW and temperatures and acceptance rates for exchanges between replicas for TReX are given in Supplementary Table 2, and we ran 5 independent 50 ns trajectories for each method. Met-enkephalin, having many more degrees of freedom, would be more difficult to fully sample from a standard MD trajectory to serve as a gold standard for convergence. Instead, we measure the rate of convergence of the independent trajectories to each other; if the trajectories are initiated from very different initial conditions, the expectation is that any significant differences would indicate that a given method suffers from inferior sampling efficiency over the 50 ns period.

We begin with the measured variance between different trajectories for free energy values generated across the Ramachandran plots for each of the 5 residues in met-enkephalin. However, since higher deviations are likely to occur in the high-energy regions where sampling is poorer, we have plotted a normalized version where the standard deviation at a point is divided by the average free energy at that point. This will more heavily weight the lower free energy regions of the Ramachandran plot, and thus the degree to which the 5 independent trajectories converge to a stable free energy value for

the most important minima. Figures 4 and 5 show the two normalized standard deviation Ramachandran plots for both methods for representative residues tyrosine-1 and methionine-5, respectively; all additional Ramachandran plots for the other amino acids are available in the supplementary material (Supplementary Figures 5-7). Across all residues for met-enkephalin the TREx methods exhibits higher standard deviations than for TCW. This indicates that the free energy of the low energy basins are not as well converged after 50ns for TREx, whereas for TCW there is clearly convergence to a stable free energy value in each of the basins.

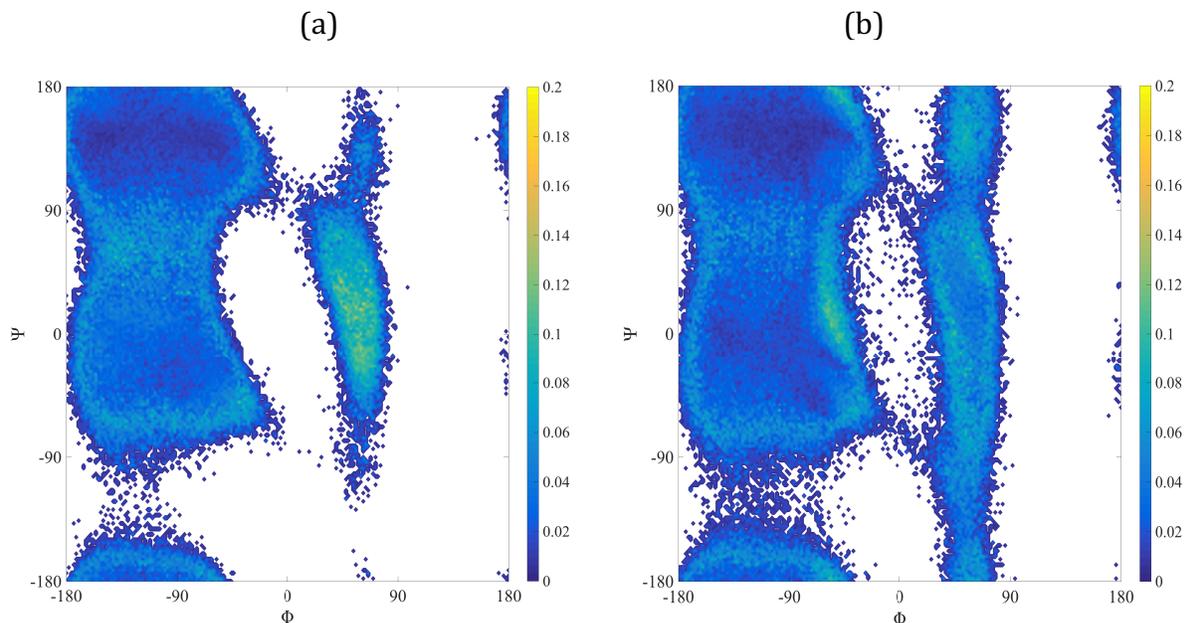


Figure 4: Normalized standard deviation between trajectories quantified for the Ramachandran ϕ , ψ angles for Tyr-1. (a) TREx and (b) TCW.

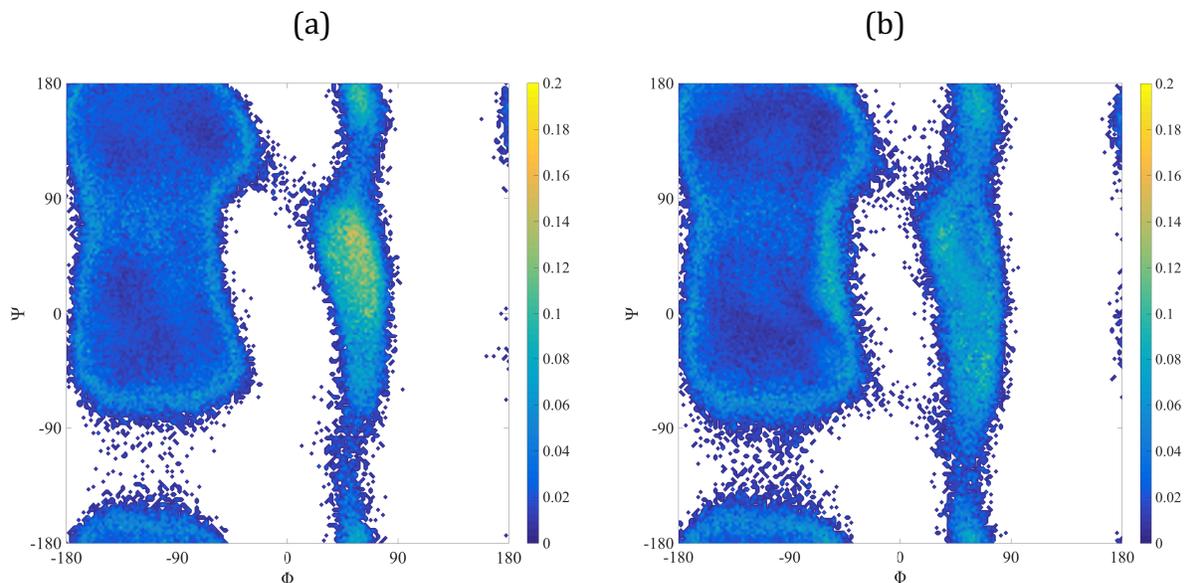


Figure 5: Normalized standard deviation between trajectories quantified for the Ramachandran phi, psi angles for Met-5. (a) TREx and (b) TCW.

We can also get a better sense of global sampling of the met-enkephalin peptide conformations using a principal component analysis. In this case we grouped all five trajectories together to generate an average conformation, center all conformations to this average, redefine all conformations as deviations from the average, and then calculate the eigenvectors that describe the major modes of the collective ensemble. Each individual conformation of a given trajectory is then projected onto the principal modes, and in Figure 6 we plot these projections onto the first principal component against the projections onto the second for two trajectories each of TCW and TREx. All additional plots for the remaining trajectories are available in the supplementary material (Supplementary Figures 8 and 9). While the TREx trajectories exhibit very limited, localized sampling of a small and closely related group of conformations, conformational sampling is evidently more uniform in the TCW tests. Though TREx is likely identifying local minimum states, as evidenced by the high probabilities for certain regions, the strong disagreement over which states are highly probable between independent trajectories indicates that each trajectory spends a disproportionate amount of time trapped in a few minima.

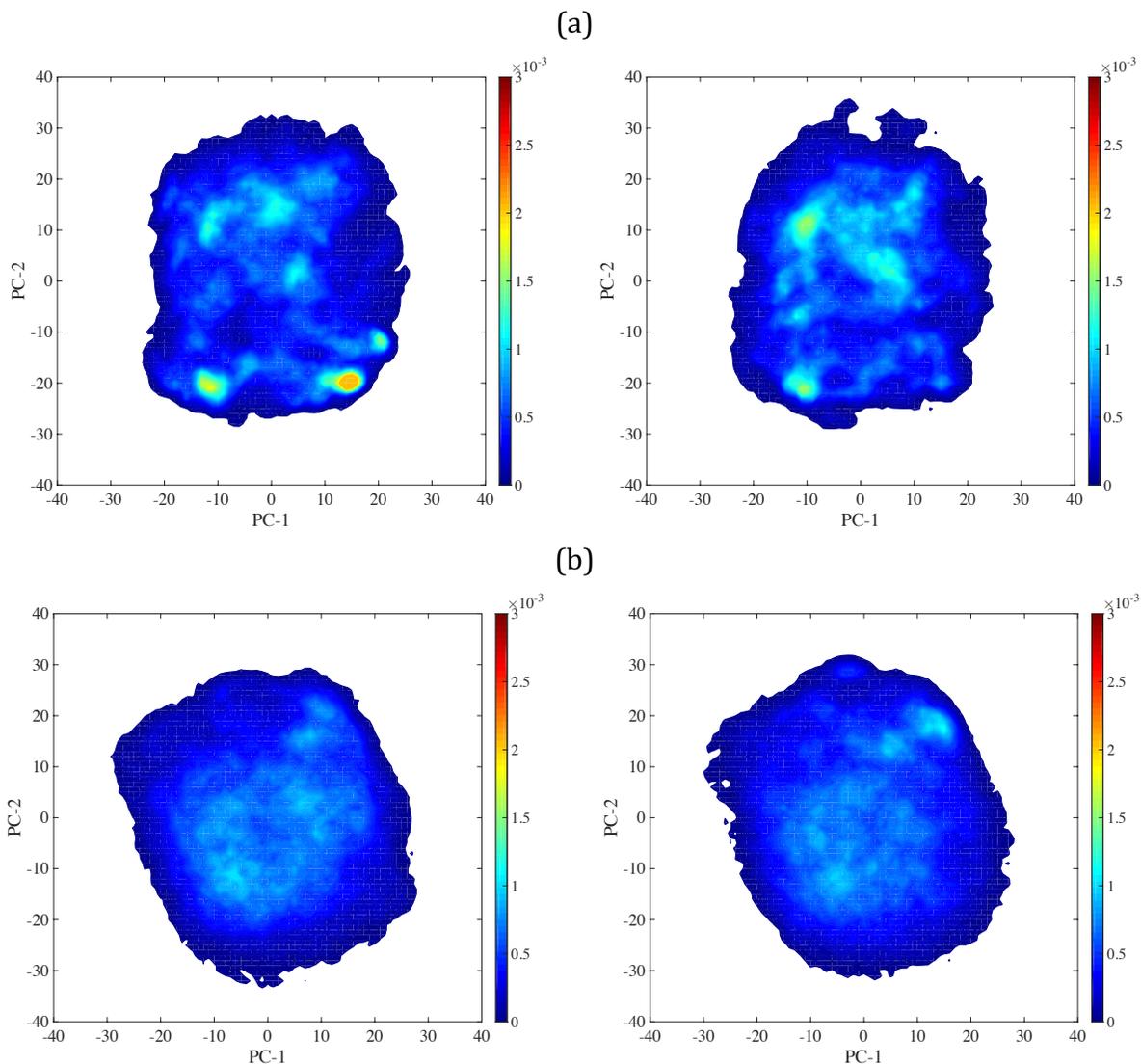


Figure 6. Projection of individual configurations onto the two lowest normal modes from a principal component analysis. Two independent trajectories for each method are shown, with the color bars representing degree of sampling in the two-dimensional space defined by projection of conformations onto the first two principal components. (a) TREx and (b) TCW.

The PCA results can also be used to develop a time-dependent metric for the rate of convergence of two independent trajectories for each method. The Kullback-Leibler divergence (KLD) quantifies the difference between two probability distributions, and we use it here to calculate the degree of difference in the exploration of principal component space between pairs of independent trajectories over time. If two trajectories reproducibly cover the same regions in phase space, the KLD will decrease toward 0. We calculated the KLD for each pair of trajectories, monitoring their overlap in sampling of the first two principal components.

The KLD values for the two principal components are plotted in Figure 7. For the first principal component, PC-1, the greatest change occurs over the first 30 ns, indicating

that much of phase space is covered by then, with steady convergence to the global metric for the remaining 20 ns (Figure 7a). The KLD metric converges faster for the TCW method, decreasing to a final value of about 0.015 over the 50 ns sample, indicating that the independent trajectories are thus very close to converging to each other with respect to PC-1. By contrast, the TREx result converges somewhat more slowly in the first 30 ns, and it remains at a higher value after 50 ns relative to TCW, indicating that the independent trajectories for TREx are farther from convergence for PC-1 than that found for TCW. Similar behavior is evident for PC-2 in which the KLD metric for TCW much more rapidly attains lower values than is observed for TREx (Figure 7b).

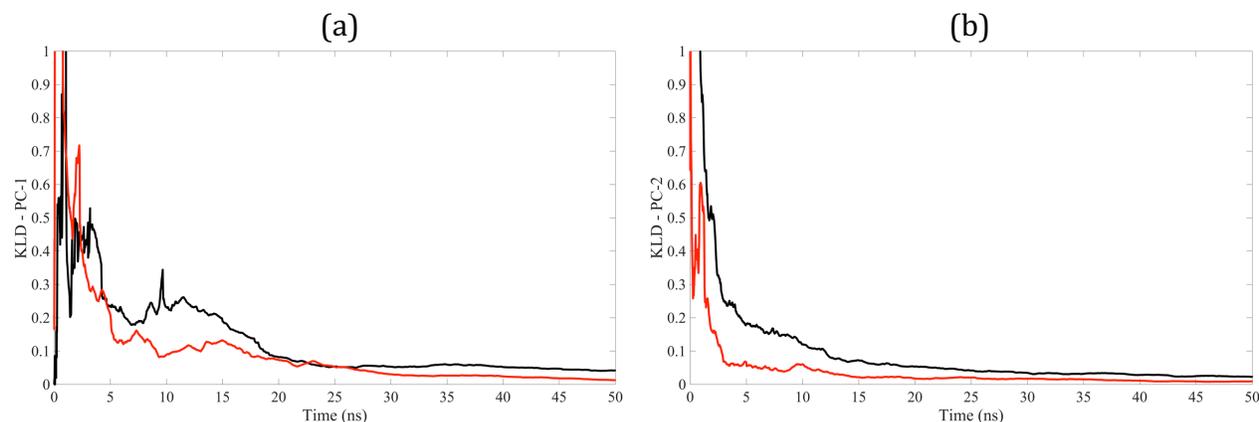


Figure 7. Time to convergence to each principal component for TCW (red) and TREx (black). (a) PCA1 and (b) PCA2.

Amyloid- β 42 To illustrate that the TCW method is computationally tractable for much larger and more difficult systems, we have evaluated the structural ensemble of the A β 42 IDP in explicit water. Because A β 42 is an intrinsically disordered peptide, its structural ensemble requires enhanced sampling techniques. We have previously used TREx using 58 replicas, running each replica out to 50 ns across a large number of CPUs to take advantage of fine-grained parallelization of energy and forces, to create two independent structural ensembles of the A β 42 peptide¹³⁻¹⁵. For the A β 42 peptide sampling reported here, we extended these two independent TREx trajectories to 100ns, to which we compare to two independent 100 ns trajectories with TCW on a single GPU. Details of the replicas and temperature schedule are reported in Supplementary Table 3. Figure 8 compares the radius of gyration distribution, in which we see that the A β 42 monomer R_g is more extended for the TCW ensemble than that observed for the TREx ensemble. Secondary structure propensities for each residue are presented in Figure 9 in which we consider helices (α , 3_{10} , and π in Figure 9a), all β content (bridges, hairpins, and sheets in Figure 9b) as well as all turn types (Figure 9c). Overall there is a qualitative change in the distribution of secondary structure propensities, in which there is more localized regions of helical and turn propensities and much less β content in the TCW ensembles, consistent with the R_g trends. We will report on the comparison to recent NMR data obtained on A β 42 in a future publication.

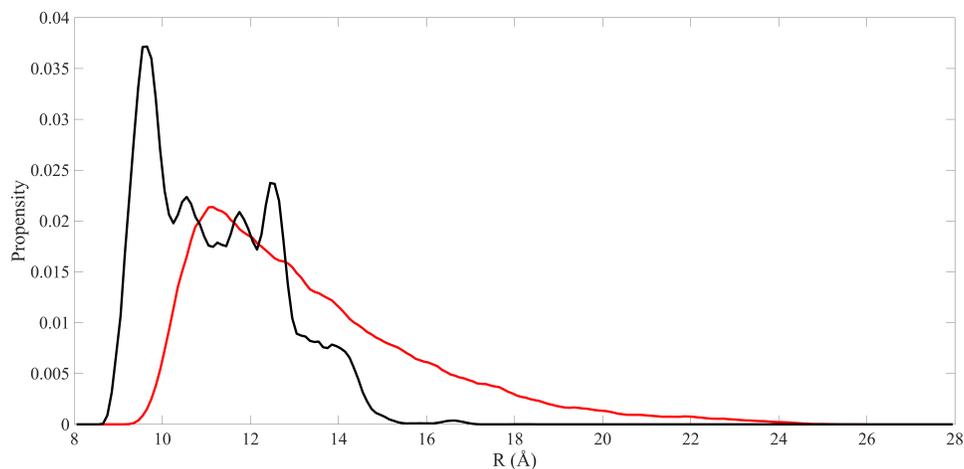


Figure 8. Probability distribution of radius of gyration of A β 42. Using TCW (red) and TREx (black).

3.5 Discussion and Conclusions

We have compared four different methods for their ability to enhance sampling and thereby increase the rate of convergence to the limiting structural ensemble, two which use temperature to overcome conformational barriers^{2, 8} and two that alter the potential energy surface by scaling the electrostatic charges¹⁰. We first tested the methods on two simple systems, an analytic rough 1D energy surface and an all-atom alanine dipeptide in water simulation, of which the relatively small phase space allows for the ability to define quantitative convergence metrics. While the DW method is shown to better than CREx for the 1D potential, the DW and CREx methods perform poorly with respect to the temperature methods for the alanine dipeptide, even though this should be a good case for both dielectric-based methods where the dominate barrier is electrostatic in origin. In essence, screening of electrostatics merely gives rise to new potential energy barriers that inhibit ergodic sampling, and we can only conclude that the dielectric CREx and DW methods are far inferior to standard TREx.

In this quantitative comparison for alanine dipeptide we also show that TCW is superior to TREx, and this outcome was shown to hold for met-enkephalin¹⁷ based on the KLD metric that measures the convergence of independent trajectories to converge onto the two primary principal components of the peptide. Lastly, we compared the results of TCW and TREx on A β 42, the Alzheimer’s peptide, known for its role in Alzheimer’s disease²², in which we find non-negligible differences in the disordered ensemble compared to standard TREx.

Finally, we consider the computational efficiency of the two temperature methods. The TREx approach requires tens of replicas and a CPU cluster with a real communication backbone to perform swaps and can be made highly efficient on CPUs. By contrast, TCW always requires only two replicas and is better suited to GPU systems. In order to disseminate our new enhanced sampling approaches applied to complex systems such as peptides and proteins in water, OpenMM²³ codes for TCW and DW have been made available; for comparison, the TREx and CREx methods have been made available as well.

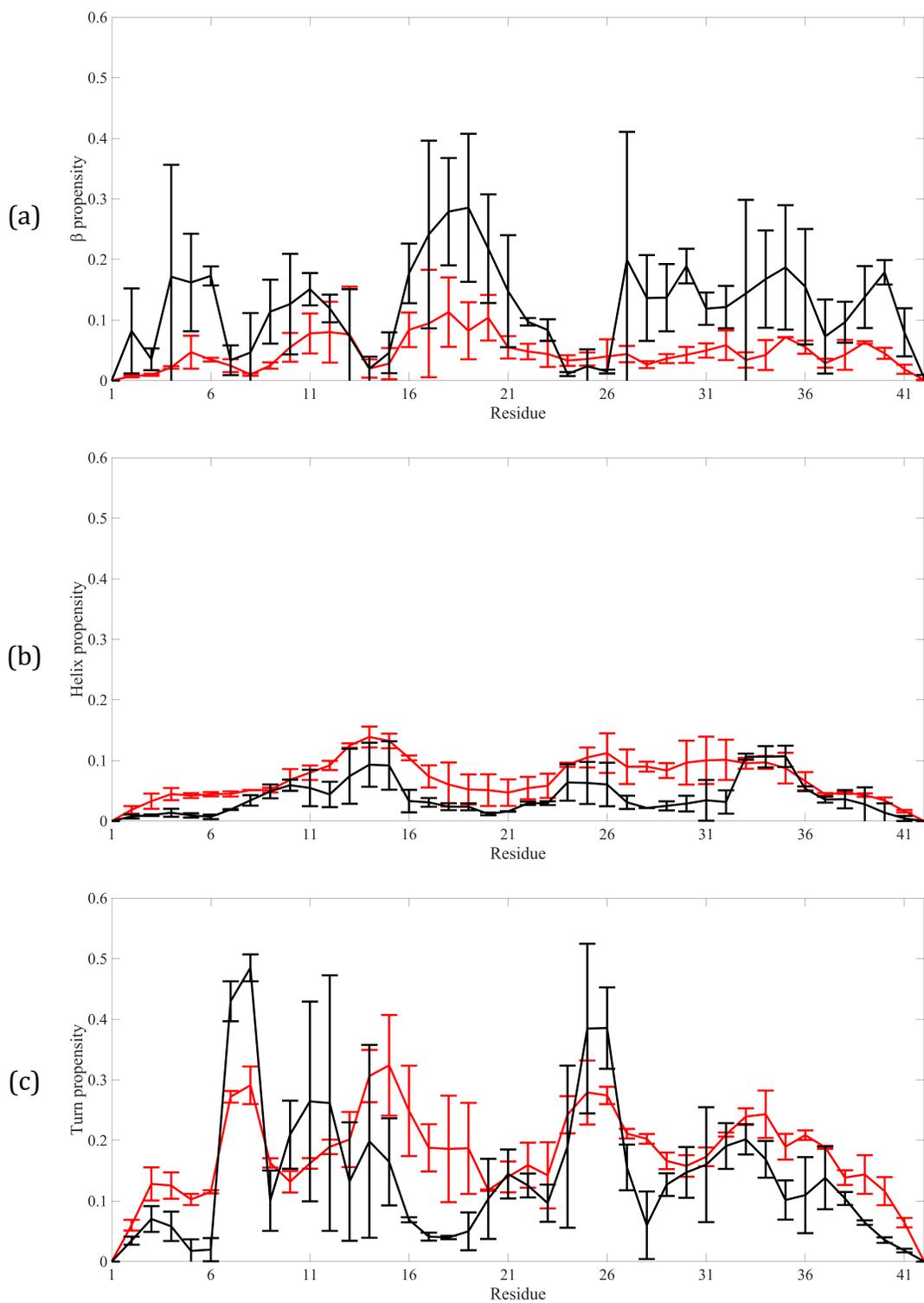


Figure 9: Propensity of the different types of secondary structure for Aβ42. β-bridges or β-strands (a), helix (b), and turns (c), using TCW (red) and TReX (black).

3.6 Acknowledgments

THG acknowledges support from the NSF under grant CHE-1363320. JL acknowledges partial support under the NIH Molecular Biophysics TG, T32-GM008295. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

3.7 References

1. Mitsutake, A.; Sugita, Y.; Okamoto, Y., Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Peptide Science* **2001**, *60*, 96-123.
2. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314*, 141-151.
3. Zuckerman, D. M.; Lyman, E., A Second Look at Canonical Sampling of Biomolecules Using Replica Exchange Simulation. *Journal of Chemical Theory and Computation* **2006**, *2*, 1200-1202.
4. Nadler, W.; Hansmann, U. H. E., Optimized Explicit-Solvent Replica-Exchange Molecular Dynamics from Scratch. *J. Phys. Chem. B* **2008**, *112*, 10386.
5. Fukunishi, H.; Watanabe, O.; Takada, S., On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *The Journal of Chemical Physics* **2002**, *116*, 9058-9067.
6. Roe, D. R.; Bergonzo, C.; Cheatham, T. E., Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *The Journal of Physical Chemistry B* **2014**, *118*, 3543-3552.
7. Periole, X.; Mark, A. E., Convergence and Sampling Efficiency in Replica Exchange Simulations of Peptide Folding in Explicit Solvent. *The Journal of Chemical Physics* **2007**, *126*, 014903.
8. Brown, S.; Head-Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *Journal of Computational Chemistry* **2003**, *24*, 68-76.
9. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J Chem Phys* **2000**, *113* (15), 6042-6051.
10. Itoh, S. G.; Okumura, H., Coulomb Replica-Exchange Method: Handling Electrostatic Attractive and Repulsive Forces for Biomolecules. *Journal of Computational Chemistry* **2013**, *34*, 622-639.
11. Flores-Canales, J. C.; Kurnikova, M., Targeting Electrostatic Interactions in Accelerated Molecular Dynamics with Application to Protein Partial Unfolding. *Journal of Chemical Theory and Computation* **2015**, *11*, 2550-2559.
12. Hansmann, U., Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140-150.
13. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-B Peptides. *Biochemistry* **2011**, *50*, 7612-7628.
14. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in B-Strand Populations of Monomeric Amyloid-B 40 and Amyloid-B 42. *Biophys. J.* **2013**, *104* (12), 2714-2724.

15. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-B Peptides. *The Journal of Physical Chemistry B* **2014**, *118*, 6405-6416.
16. Opps, S. B.; Schofield, J., Extended State-Space Monte Carlo Methods. *Phys. Rev. E* **2001**, *63*, 056701.
17. Sanbonmatsu, K. y.; García, A. e., Structure of Met-Enkephalin in Explicit Aqueous Solution Using Replica Exchange Molecular Dynamics. *Proteins: Structure, Function, and Bioinformatics* **2002**, *46*, 225-234.
18. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712-25.
19. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79* (2), 926-935.
20. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *The Journal of Chemical Physics* **2004**, *120*, 9665-9678.
21. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J Mol Graph Model* **1996**, *14* (1), 33-38.
22. Haass, C.; Selkoe, D. J., Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid B-Peptide. *Nature Reviews Molecular Cell Biology* **2007**, *8*, 101-112.
23. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S., Openmm 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of chemical theory and computation* **2013**, *9*, 461-469.

3.8 Appendix

Supplementary Table 1: *Coefficients for the one-dimensional potential energy function.*

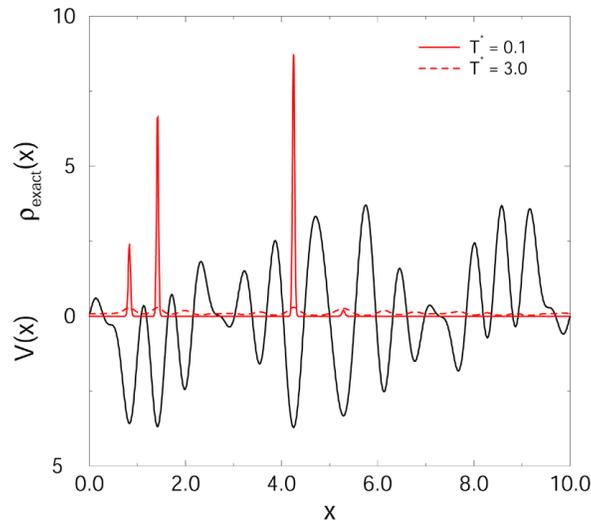
C_1	-0.466516	C_{11}	0.891462
C_2	-0.834376	C_{12}	-0.665239
C_3	-0.714529	C_{13}	0.810546
C_4	-0.0245586	C_{14}	0.198216
C_5	0.238837	C_{15}	-0.816637
C_6	0.0143649	C_{16}	-0.195351
C_7	0.271003	C_{17}	-0.573181
C_8	-0.374538	C_{18}	0.251745
C_9	0.873564	C_{19}	0.647615
C_{10}	-0.370258	C_{20}	0.201654

Supplementary Table 2: Replicas and exchange probabilities for TReX and CREx; annealing schedules for TCW and DW.

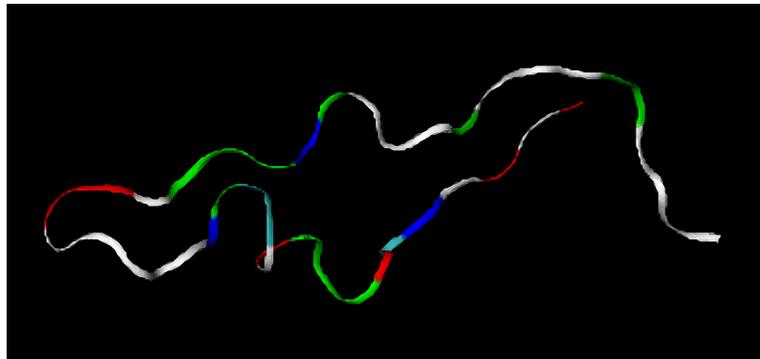
Alanine	TReX Replicas												
	T, (K)	300.0	314.5	331.0	347.5	367.0	387.5	410.0					
	Acc, %	20.9	19.0	22.5	18.7	20.2	19.9						
	CREx Replicas												
	Eps	0.69	1.00	1.78	4.00	16.00	Inf.						
	Acc, %	0.09	0.06	0.11	0.16	0.24							
	TCW Schedule												
	T, (K)	300.0	314.5	331.0	347.5	367.0	387.5	410.0					
	DCW Schedule												
	Eps	1.00	1.20	1.39	1.54	1.68	1.83	2.04	2.36	2.88	3.67	4.84	
Met-Enk	TReX Replicas												
	T, (K)	300.0	310.0	320.5	331.5	342.5	355.5	369.5	384.0	399.5	415.5	432.0	
	Acc, %	20.2	20.0	20.2	22.6	18.1	16.9	17.9	17.8	18.8	20.1		
	TCW Schedule												
	T, (K)	300.0	308.0	317.1	327.2	338.2	349.9	361.9	374.0	386.0	398.0	410.0	425.0

Supplementary Table 3: Replicas and exchange probabilities for TReX and annealing schedule for TCW for A β 42. The annealing schedule used for TCW was identical to the set of TReX replicas.

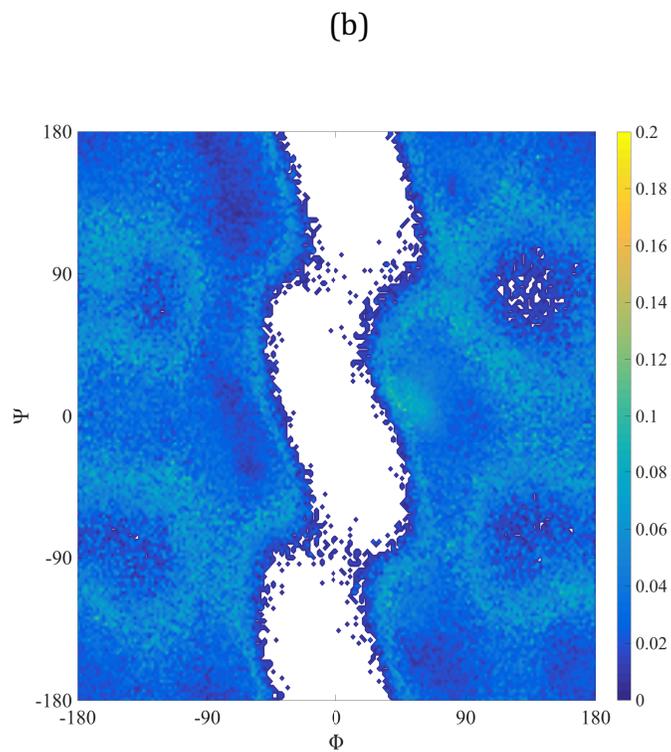
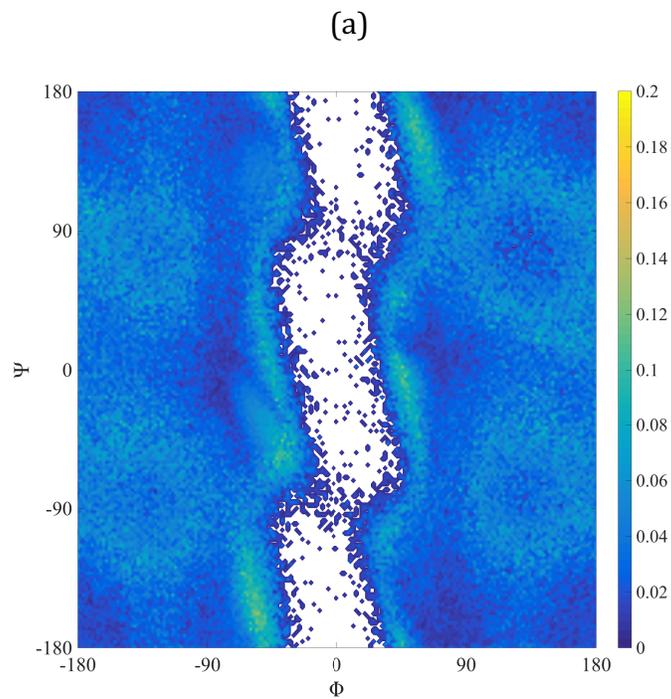
A β 42	TReX Replicas												
	T, (K)	287.0	289.1	291.2	293.3	295.5	297.6	299.8	302.0	304.2	306.5	308.7	311.0
	Acc, %	20.0	20.0	19.0	20.0	20.0	21.0	21.0	21.0	21.0	21.0	22.0	20.0
	T, (K)	313.3	315.6	318.0	320.3	322.7	325.2	327.6	330.1	332.7	335.3	337.9	340.6
	Acc, %	23.0	21.0	22.0	22.0	21.0	22.0	22.0	21.0	21.0	20.0	21.0	21.0
	T, (K)	343.3	346.0	348.8	351.7	354.5	357.4	360.4	363.4	366.4	369.4	372.6	375.7
	Acc, %	22.0	20.0	20.0	21.0	21.0	21.0	21.0	21.0	21.0	21.0	20.0	21.0
	T, (K)	378.9	382.2	385.4	388.7	392.1	395.5	398.9	402.4	405.9	409.5	413.1	416.8
	Acc, %	22.0	21.0	21.0	21.0	21.0	21.0	21.0	22.0	21.0	21.0	22.0	22.0
	T, (K)	420.5	424.3	428.1	431.9	435.9	439.8	443.9	447.9	452.0	456.2		
Acc, %	22.0	21.0	22.0	22.0	21.0	22.0	21.0	21.0	23.0				



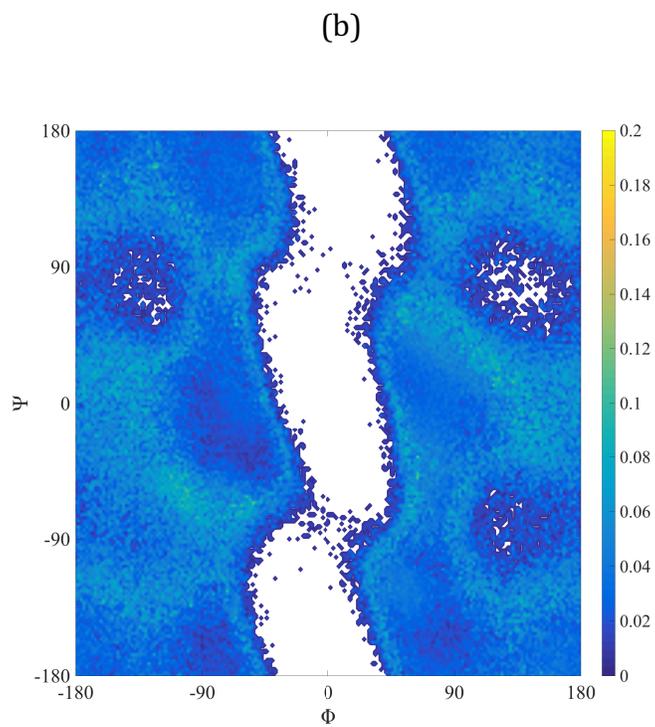
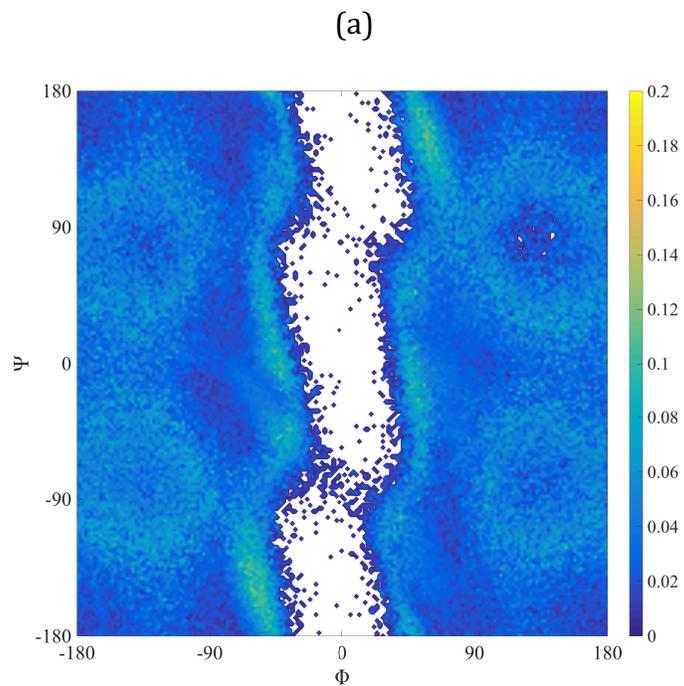
Supplementary Figure 1: *One-dimensional potential energy $V(x)$ versus position x for box length $L = 10$. Also shown in the plot is the ideal distribution function $\rho_{exact}(x)$ for temperatures of $T^* = 0.1$ and 3.0 . The units for the potential energy are arbitrary. At $T^* = 0.1$ the distribution is sharply peaked about the four lowest minima, whereas at the higher temperature $T^* = 3.0$ the peak heights for these minima are roughly equal.*



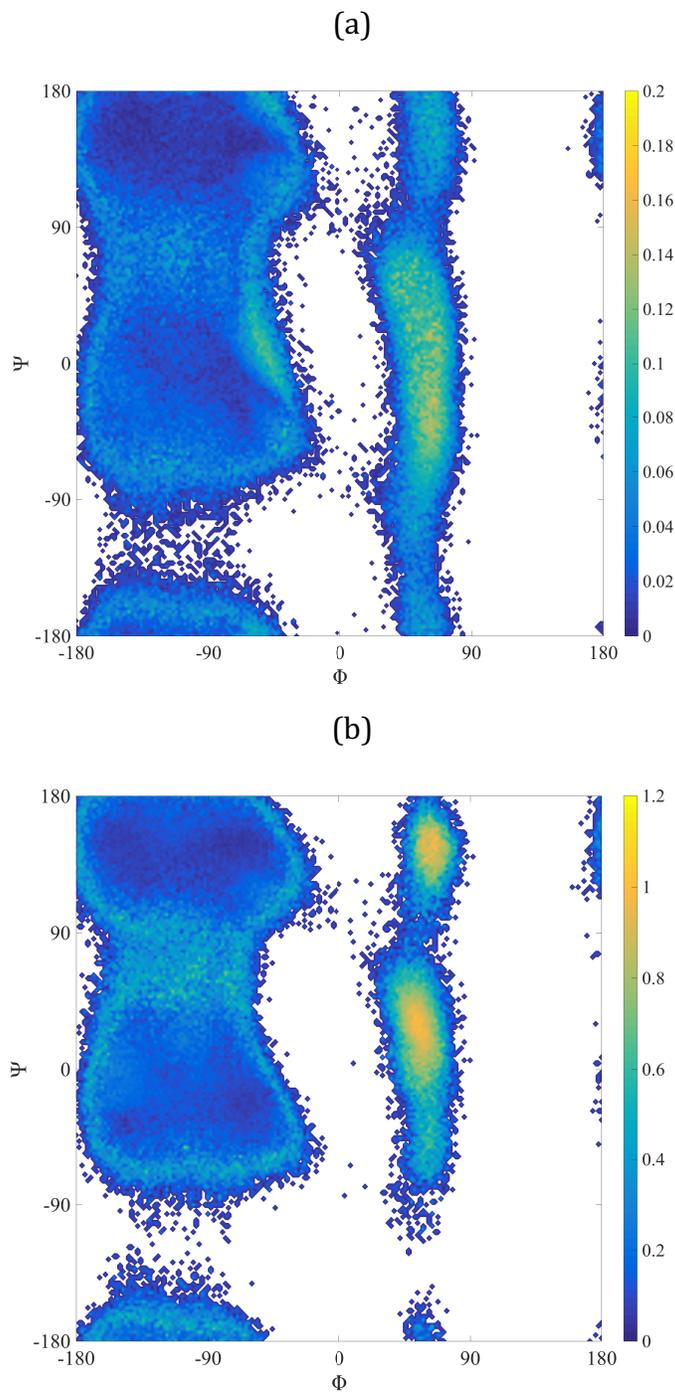
Supplementary Figure 2: *Starting structure for $A\beta 42$ simulations. The peptide is represented as a ribbon, and residues are colored by type to differentiate the C terminus (red and white striped) from the N terminus (white).*



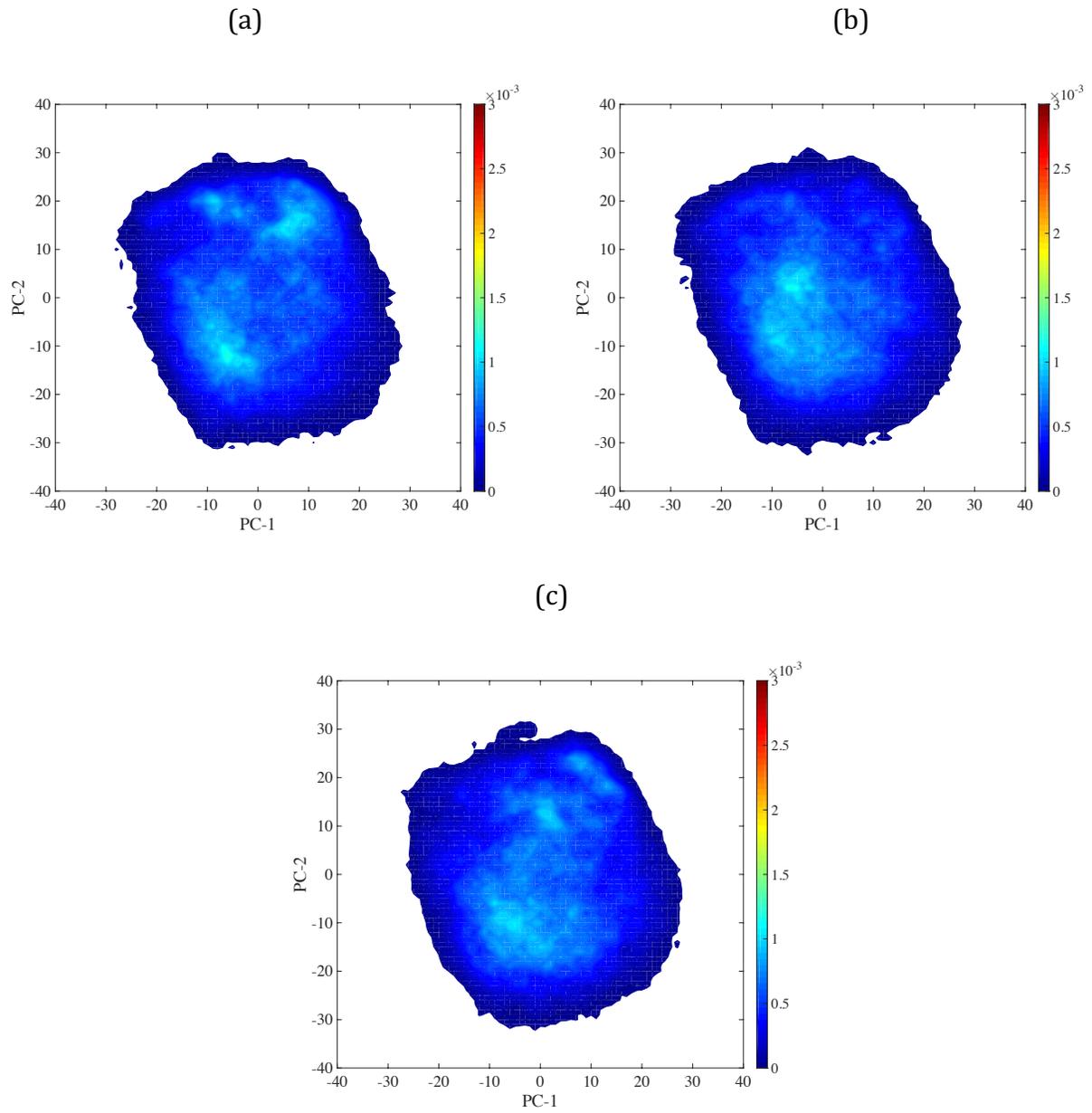
Supplementary Figure 3: *Normalized standard deviation between trajectories quantified for the Ramachandran phi, psi angles for Gly-2 of met-enkephalin. (a) TCW and (b) TReX.*



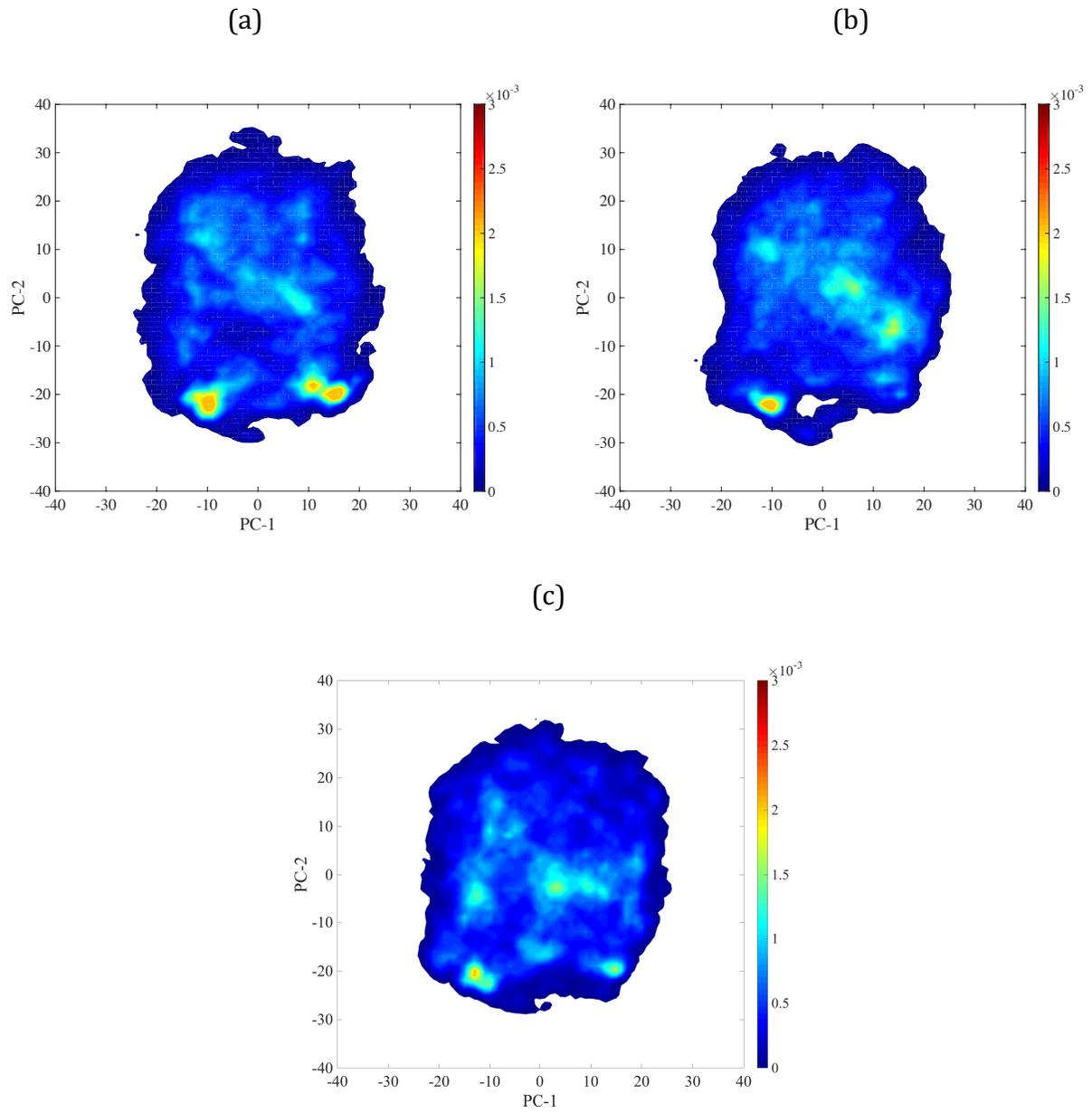
Supplementary Figure 4: Normalized standard deviation between trajectories quantified for the Ramachandran ϕ , ψ angles for Gly-3 of met-enkephalin. (a) TCW and (b) TREx.



Supplementary Figure 5: *Normalized standard deviation between trajectories quantified for the Ramachandran phi, psi angles for Phe-4 of met-enkephalin. (a) TCW and (b) TREx.*



Supplementary Figure 8: *Principal component analysis tests for three additional TCW runs (a) – (c) of met-enkephalin.*



Supplementary Figure 9: *Principal component analysis tests for three additional TREx runs (a) – (c) for met-enkephalin.*

Chapter 3

The Combined Force Field-Sampling Problem in Simulations of Disordered Amyloid- β Peptides*

Molecular dynamics simulations of intrinsically disordered peptides (IDPs) can provide high resolution structural ensembles if the force field is accurate enough and if the simulation sufficiently samples the conformational space of the IDP with the correct weighting of sub-populations. Here, we investigate the combined force field-sampling problem by testing a standard force field as well as newer fixed charge force fields, the latter specifically motivated for better description of unfolded states and IDPs, and comparing them with a standard temperature replica exchange (TREx) protocol and a non-equilibrium Temperature Cool Walking (TCW) sampling algorithm. The force field and sampling combinations are used to characterize the structural ensembles of the amyloid- β peptides A β 42 and A β 43, which should both be random coils as shown recently by experimental NMR and 2D FRET experiments. The results illustrate the key importance of the sampling algorithm: while the standard force field using TREx is in poor agreement with the NMR J-coupling and NOE and 2D FRET data, when using the TCW method the standard and optimized protein-water force field combinations demonstrate significantly improved agreement with the same experimental data, with the TCW sampling method producing qualitatively different ensembles than TREx that are indicative of the different levels of sampling acceleration for the two methods.

3.1 Introduction

Intrinsically disordered proteins (IDPs) are a class of biomolecules that do not adopt a well-defined equilibrium structure in solution, instead sampling an ensemble comprised of sub-populations of fully and/or partially disordered structures^{1, 2}. A classic IDP example is the amyloid- β (A β) peptide associated with Alzheimer's disease³, for which recent state-of-the-art solution-based nuclear magnetic resonance (NMR) and Förster resonance energy transfer (FRET) experiments have shown that the monomeric forms of the A β 40 and A β 42 peptides are largely random coils⁴⁻⁶. Roche and co-workers performed multiple types of J-coupling measurements on A β 40 and A β 42 to show there were no overt differences from random coil signatures for both peptides⁵. They further supported this result with very high resolution nuclear Overhauser effect (NOE) spectra that showed that both A β 40 and A β 42 are dominated by short ($i, i+1$ contacts) and to a lesser extent medium range ($i, i+2$ to $i, i+4$ contacts) NOEs, and thus any longer-range helical or β -hairpin formation would only be expected to appear at levels below the detection limit⁵. Conicella and Fawzi employed ¹H_N and ¹⁵N chemical shift data and H_N-H _{α} J-couplings, as well as ¹⁵N R₂, ¹⁵N R₁, and heteronuclear ¹⁵N-¹H NOE measurements, to show that there are no structural differences

* Reproduced with permission from: Lincoff, J.*; Sasmal, S.*; Head-Gordon, T., The Combined Force Field-Sampling Problem in Simulations of Disordered Amyloid- β Peptides. *J. Chem. Phys.* **2019**, *150* (10). (* denotes equal co-first authors)

in the N-terminus and central hydrophobic core (CHC) between the A β 42 and A β 43 peptides, with only small structural differences in the C-terminus⁴. Since the two peptides differ only by a threonine residue at the C-terminus, it implies that the A β 43 monomer is also largely random coil-like in structure, albeit with a greater propensity for aggregation⁴. Finally, using single molecule and 2D FRET, Meng *et al.* showed that both A β 40 and A β 42 have ensembles that are dominated by expanded conformations with no persistent sub-populations of secondary or tertiary structure (i.e. no long helices or β sheets above \sim 5-10 % population), in general agreement with the NOE spectra of Roche *et al.*⁶.

Computational techniques are often combined with such experimental information to create the structural ensemble and to characterize the sub-populations of an IDP of interest⁷⁻¹⁰. A complementary approach is to generate IDP ensembles using molecular dynamics (MD) simulation without experimental information as input, which therefore requires an accurate force field and sampling method that can describe the conformational substates of the IDP ensemble. In early simulation studies of IDPs, research groups relied on off-the-shelf and pairwise-additive protein and water force fields¹¹, such as the Amber¹², GROMOS¹³, OPLS-AA^{14, 15} and CHARMM¹⁶ protein force fields in combination with TIP3P¹⁷, TIP4P¹⁸, and TIP4P-Ew¹⁹ water models. But since standard pairwise additive force fields are parameterized using mostly folded protein crystallographic data²⁰, they have been thought to be insufficient for modeling of IDPs because they exhibit a bias toward overly collapsed and ordered structural ensembles, or poorly reproduce the equilibrium between unfolded and native conditions for globular proteins²¹⁻²³.

Furthermore, multiple groups have shown that unfolded and IDP structural ensembles generated using different standard force fields vary considerably in terms of secondary structure content²⁴⁻²⁷. For example, in studies of the A β ₁₆₋₂₂ peptide, Nguyen *et al.* have demonstrated that Amber99 predicts more helical structures and GROMOS96²⁸ favors more β -strand structures, whereas OPLS-AA demonstrates no particular secondary structure preference²⁴. As we have discussed before, the quality of the water model is also critical for accurate molecular simulations of peptides and proteins by balancing the relative strengths of water-water and water-solute interactions.²⁹ Several groups including Song and co-workers³⁰, Amini and co-workers³¹, and Viet and co-workers³² have conducted straight MD simulations on A β peptides on the order of microseconds with standard protein force fields and using a variety of three-site and four-site water models and multiple starting structures. Using brute force MD, Robustelli and co-workers simulated A β 40 using a single 30 μ s trajectory of a standard protein force field Amber99ffsb-ildn* and TIP3P, and found close to 90% β -sheet structure in different regions of the sequence³³. In general, all of these simulations produce overstructured ensembles, in disagreement with the experimental results of Roche *et al.*⁵ and Meng *et al.*⁶ for A β peptides. One might surmise from the accumulation of evidence that the standard force fields for the protein and water are failing to describe disordered or unfolded protein ensembles.

To make better and more uniform predictions across different IDPs, as well as more accurate models for protein folding equilibrium, a number of research labs have modified the parameters used in standard force fields^{21-23, 27, 34, 35}. These force field modifications include adjusting the water-protein London dispersion interactions to be more favorable^{21, 22}, refining the peptide backbone parameters to produce more expanded structures^{22, 23, 34} or reduce propensity to certain ordered conformations²¹, and/or changing the salt-bridge interactions²¹. Skepö and co-workers, using long MD simulations, have shown that some of

these modified force fields produce better agreement with experimental SAXS data in terms of radius of gyration, R_g , for the disordered Histatin 5 peptide^{36, 37}. Huang *et al.* also found better agreement with the SAXS profile for the RS peptide using the modified CHARMM force field²¹. However, Robustelli *et al.* simulated A β 40 with a single 30 μ s trajectory for the recently modified A03ws force field by Best and co-workers³⁸, and found close to 25-60% α -helix in large regions of the sequence that indicate that the newer IDP force fields may also be experiencing problems³³.

The other important and often not enough emphasized aspect of generating IDP structural ensembles is the sampling technique itself. Because IDPs have a relatively flat energy landscape with many local minima, it takes substantial sampling efficiency to determine all relevant configurations with the correct weighting of multiple small sub-populations. Enhanced sampling methods are therefore generally applied for the simulation of IDPs, as they accelerate the rate of convergence to timescales that are significantly less than possible with a brute force MD simulation^{39, 40}. The most common enhanced sampling technique used in the IDP field at present is temperature replica exchange (TREx) method³⁹⁻⁴³. In fact, TREx simulations were used to identify perceived errors in standard force fields, which led to some of the modified force fields developed to improve modeling of unfolded proteins and IDPs^{26, 27, 38, 44}.

However, a noted deficiency of TREx for large systems is the diffusiveness of barrier crossing due to many closely spaced intermediate replicas, when energy landscapes are dominated by entropic barriers^{39, 45, 46}. Several alternative enhanced sampling methods with better performance than TREx have been developed and applied to the study of IDPs, including replica exchange with solute tempering (REST)⁴⁷, metadynamics⁴⁸, and MD combined with Markov State Model (MSM) analysis⁴⁹. More specifically, Lin *et al.*⁴⁹ performed \sim 200 μ s MD simulations over many initial conditions, and combined them with MSM to reach even longer timescales, to characterize the A β 42 structural ensemble using the Amberff99SB protein force field and the TIP3P water model. Over this much longer sampling timescale they obtained far higher quantities of extended, largely unstructured conformations, with the only noteworthy structural component being \sim 10 – 20 % observed helical content between residues 12-18. While there may be concerns that the clustering protocols may have introduced error in secondary structure populations, and improved clustering methods are now available⁸⁹, it is evident that that the full weighted ensemble based on the MSM is equivalent to the raw α -helix and β propensities from their production MD runs (as seen in their supplementary information). Thus this more extensive sampling produced a significant improvement in generating random coil ensembles using a standard peptide-water force field, in good agreement with the experimental results of Roche *et al.*⁵ and Meng *et al.*⁶ for A β peptides, and in significant disagreement with previous shorter MD runs³³ and TREx studies^{26, 27, 38, 44}.

We have developed the temperature cool walking (TCW) technique^{46, 50}, a non-equilibrium alternative to TREx, which uses only one high temperature replica to generate trial moves for the target temperature replica. In previous studies we have shown that TCW converges more quickly to the proper equilibrium distribution than TREx, and at much lower computational expense, for a 1D rough surface⁴⁶, and for alanine dipeptide and met-enkephalin—sufficiently small systems where well-defined and quantitative metrics of convergence are available⁵⁰. More recently, we have been able to apply TCW to larger

systems such as A β through its implementation in OpenMM^{50, 51}. In this work we address both dimensions of the IDP problem by comparing different combinations of enhanced sampling techniques, TReX and TCW, and protein-water force fields, unmodified and newly-optimized, testing combinations of each on A β 42 and A β 43. Again summarizing the recent experiments⁴⁻⁶, the structural ensembles of A β 42 and A β 43 peptides should be largely the same and exhibit no persistent structural ordering or long-range contacts. Thus, one would expect the computationally generated structural ensembles to be highly similar for A β 42 and A β 43 in terms of back-calculations to experimental observables such as chemical shifts or J-couplings^{4, 5}, and that both peptides would display largely random coil configurations, lacking stable populations of organized structures such as α -hairpins and extended helices, to agree with reported NOE^{4, 5} and 2D FRET⁶ data for A β .

Among our set of results, we attain the biggest improvement in IDP ensemble generation by switching the sampling method from TReX to TCW, since the latter sampling algorithm provides much better agreement with the full range of NMR J-coupling and NOE data^{4, 5} as well as with the 2D FRET data⁴⁻⁶, compared to the same force fields simulated with TReX. More specifically, the unmodified force field when sampled with TCW yields A β ensembles that are largely unstructured, in qualitative agreement with the robust MD/MSM simulation⁴⁹ of the A β 42 peptide using a similar unmodified force field, with only small populations of structures containing longer-range contacts at levels (\sim 5-10 %) that would be undetectable by the NOE and 2D FRET experiments. We believe that the presented evidence supports the conclusion that TCW is more capable than TReX of sampling the disordered protein energy landscape, and is in support of recent work by Granata and co-workers that have shown that disordered conformations are lower in free energy than ordered structures⁵². Finally, using the TCW sampling protocol, we find that the newly modified force fields do produce more extended ensembles that are in better quantitative agreement with the 2D FRET data, whereas the standard force field is in better quantitative agreement with the NMR J-coupling data. We conclude that more work is needed in regards interpreting FRET data^{6, 53-57} and more extensive testing in general is needed before standard force fields are abandoned or more force field changes are pursued for IDPs.

3.2 Materials and Methods

Peptide Simulations. The starting A β 42 and A β 43 configurations were created using the *tleap* module in Amber⁵⁸, and the peptides were subsequently minimized and equilibrated in the NPT ensemble at 1 bar to obtain the correct density. Structures at the density of maximum probability were selected as initial structures for production, producing boxes that were approximately 60 Å on each side. Ewald summation was used for the long-range, with a cutoff of 9.0-9.5 Å for the real space electrostatics and Lennard-Jones forces. Trajectories were analyzed for results at 287 K.

TReX Simulation Protocol. The Amber14 molecular dynamics package⁵⁸ was used to perform 100 ns TReX simulations with 58 temperature replicas in the temperature range 287-450 K, testing both peptides with 1) the Amber ff99SB¹² force field with TIP4P-Ew¹⁹ water and 2) the CHARMM36m²¹ force field with CHARMM-TIP3P water¹⁶. The temperature schedule was chosen such that the exchange probability between each pair of replicas was in the range 18-22 %, which has been shown to be optimal for TReX⁵⁹.

Exchanges between neighboring replicas were attempted every 0.5 ps. The TReX simulations were performed in the NVT ensemble with a time step of 1 fs and with constraints on the heavy atom hydrogen bonds. A Langevin thermostat was used to maintain constant temperature. Each set of conditions—peptide + force field + water model—was run in duplicate. The first 50 ns of data were discarded as equilibration, with the last 50 ns being analyzed and presented, averaged across the two independent simulations. TReX simulations with this setup were also attempted for the Amber ff99SB-ILDN^{22, 60} force field with TIP4P-D²² water, but using the same temperature ladder and simulation package we were unable to obtain similar exchange probabilities across the entire temperature span. As we could not run an optimal TReX simulation with this force field combination, no data are presented for this simulation setup.

Over the last several years, several groups have run much longer TReX simulations for IDPs, on the order of 750 – 1000 ns, noting that even when optimally run, structural properties such as the radius of gyration and secondary structure propensities can take several hundred ns to reach apparent equilibration in TReX^{6, 44, 61, 62}. To evaluate this difference, we additionally performed one 800 ns TReX simulation each for A β 42 and A β 43 using Amber ff99SB + TIP4P-Ew water using the Amber16 molecular dynamics package⁶³. All parameters were kept the same as in the original 100 ns TReX simulations, with the exception of using a 2 fs timestep and attempting exchanges between adjacent pairs of replicas every 1 ps. For these simulations, the first 300 ns of data were discarded as equilibration, with the last 500 ns of data at 287 K analyzed. Averaging over two 250 ns blocks was done to mimic the averaging over two independent simulations performed for the 100 ns TReX simulations and the TCW simulations.

TCW Protocol. The TCW enhanced sampling method uses only two temperature replicas to generate an equilibrium ensemble at the target temperature^{46, 50}. Starting with expanded structures at the high temperature, sequential cooling is performed to obtain structures at the low temperature such that detailed balance is satisfied. The same set of maximum and minimum temperatures and cooling schedule was used as for all TReX simulations. Temperature was regulated using an Andersen thermostat. Trial moves were initiated every 8 ps. The cooling rate was set such that the peptides spent 40 fs on average at each intermediate temperature, with the first trial exchange per cooling run occurring after having annealed through 40 of the 56 intermediate temperature steps on average. Additional high temperature replica propagation was performed to further decorrelate subsequent configurations, at a ratio of 8:1 to each fs of annealing performed. TCW simulations were performed using modifications to the OpenMM software package⁶⁴ and code is available upon request from the authors. TCW simulations were performed with the 1) Amber ff99SB force field + TIP4P-Ew water model 2) CHARMM36m force field + CHARMM-TIP3P water and 3) Amber ff99SB-ILDN force field + TIP4P-D water. Simulations were run in duplicate for 200 ns, with the first 50 ns of each discarded as equilibration, and results from the remaining 150 ns of each simulation averaged together.

Trajectory Analysis. The structural ensembles were analyzed using both the *cpptraj*⁶⁵ module of Amber and in-house codes. Contact maps were generated by calculating the fraction of structures where pairs of residues had at least one pair of heavy atoms within 5 Å of each other. The DSSP criterion was used to assign secondary structures⁶⁶. Details about the back calculation of NMR observables have been reported in previous publications

by our group^{10, 29, 67, 68}, including chemical shifts from ShiftX2⁶⁹ and J-coupling constants⁷⁰. In this work we focus primarily on J-couplings using the Karplus equation

$$\langle J \rangle = \langle A \cos^2 \phi + B \cos \phi + C \rangle \quad (1)$$

where the $\langle \dots \rangle$ denote ensemble averages. We calculate the χ^2 parameter from the simulated J-coupling constants ($J_{\text{HN-H}\alpha}$) for each ensemble as compared to experiment,

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

where J_i is the scalar coupling constant for the i th residue, N is the total number of experimental $J_{\text{HN-H}\alpha}$ observables, subscripts *sim* and *expt* refer to the simulated and the experimental values respectively, and σ^2 is the RMSD error when using the Karplus parameters introduced by Vögeli *et al.*⁷¹ We additionally perform a Bayesian analysis on the scalar couplings, Experimental Inferential Structure Determination (EISD)⁷², that accounts for uncertainties in the values of the Karplus parameters as well as the individual per-coupling experimental uncertainties. The relative magnitudes of ensemble scores for a peptide represent the relative likelihood of that structural ensemble matching the experimental data against which they are compared, with a larger score corresponding to a higher probability.

The observable directly obtained in MD simulations of the FRET experiments, the efficiency distribution $E(t)$,

$$E_{FRET} = \frac{1}{1 + (R_{ee}/R_0)^6} \quad (3)$$

and its average E_{FRET} , is back-calculated from the simulated end-to-end distance, R_{ee} of the untagged peptide across the ensemble, with a Förster radius $R_0 = 5.2$ nm for the dye pair of Alexa 488 and 647⁶. As in Meng *et al.*, the R_{ee} is first calculated as the distance between the C_α atoms of the first and last residue of each peptide, and then scaled up to approximate the additional distance between the two fluorophores, relative to the distance between the first and last residues.

3.3 Results

We tested five different force field-sampling combinations for each of the two peptides: (1) duplicate TReX simulations using the standard Amber ff99SB¹² protein force field and the TIP4P-Ew¹⁹ water model; (2) duplicate TReX simulations using the modified CHARMM36m²¹ protein force field and the CHARMM-TIP3P¹⁶ water model; (3) duplicate TCW simulations using the standard Amber ff99SB protein force field and the TIP4P-Ew water model; (4) duplicate TCW simulations using the modified CHARMM36m protein force field and the CHARMM-TIP3P water model; and (5) duplicate TCW simulations using the Amber ff99SB-ILDN⁶⁰ protein force field and the TIP4P-D²² water model.

Figure 1 shows that the 0.1 μ s TReX + Amber ff99SB + TIP4P-Ew simulations predict that both peptides are very collapsed, as evidenced by their contact maps and R_g distributions, which originate from the abundance of organized backbone structure that

underlie the secondary structure propensities (Figure 1 and Table 1). In addition, the A β 42 and A β 43 ensembles show significant differences, with the A β 43 ensemble exhibiting increased α -helical structure in the central hydrophobic cluster (CHC) of residues 16-30, as well as increases in the turn populations at residues 36-38 and the 6-9 region with a simultaneous increase in the β strand formed by residues 3-5 and 10-12, structural sub-populations that are greatly diminished (but still present) in the A β 42 ensemble. This result is in direct contradiction with results reported by Conicella and Fawzi⁴, which showed that there are no major structural differences in the N-terminus and CHC region for the two peptide monomers.

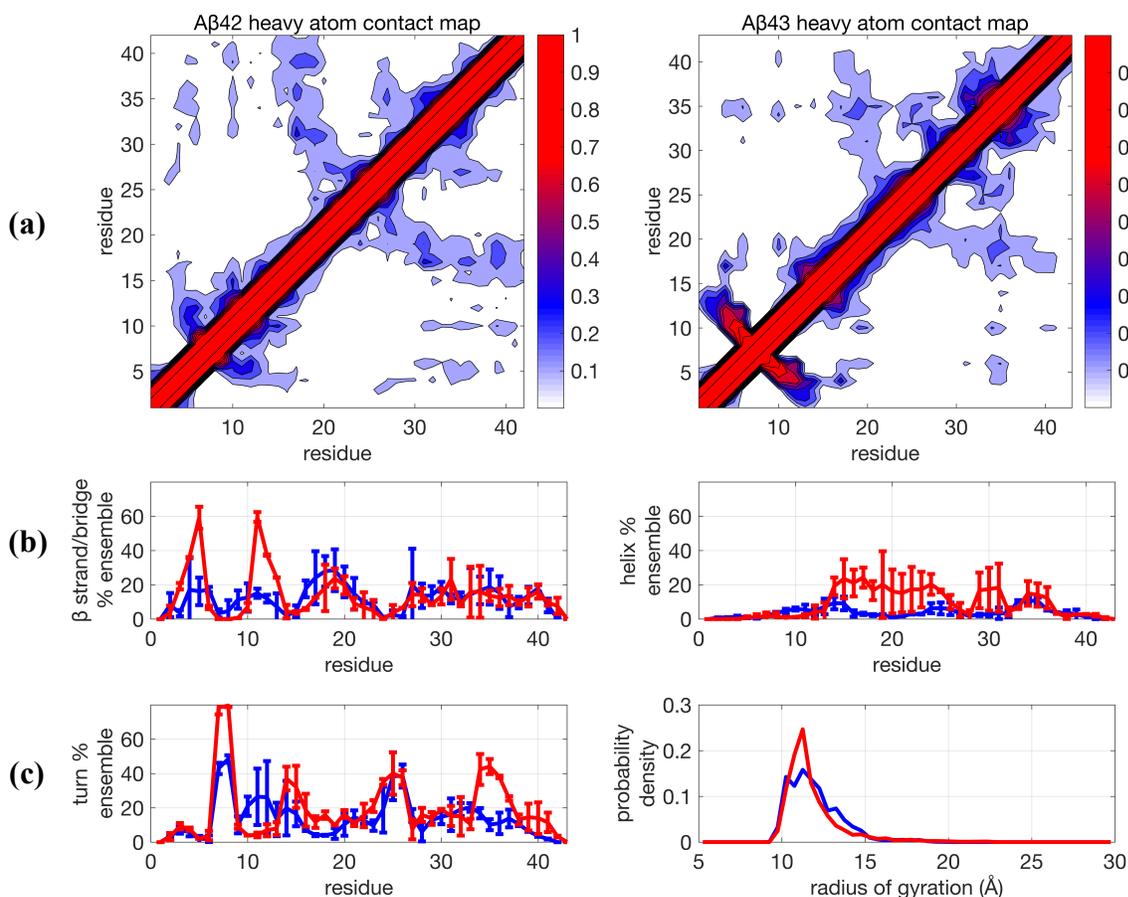


Figure 1. Comparison of structural properties for A β 42 and A β 43 using the force field Amber99SB + TIP4P-Ew and using TRES simulated at 0.1 μ s per replica. (a) Contact maps; (b) β -strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b)-(d) blue lines are for A β 42 and red lines are A β 43 and represent the average for two independent trajectories. Error bars are plus and minus one standard deviation of the calculated propensity at the given residue for the two trajectories, and generally represent the degree of agreement between the two trajectories.

When using the same TRES sampling method for 0.1 μ s, but changing to the CHARMM36m + CHARMM-TIP3P force field, there is significant reduction in long-range structure and secondary structure propensities for the two peptides, and hence the

ensembles are less collapsed compared to the standard force field combination for the two peptides (Table 1 and Figure S1). However, the TReX simulations using the newer protein-water force field produces ensembles that still predict differences in the sub-populations of structure for both A β 42 and A β 43 that are not observed in experiment,⁴ such as the relative enrichment in β content for A β 42. The long-range structure evident in the contact maps shown in Figure S1 directly contradict the NOE data of Roche *et al.*, which found no evidence for long-range contacts.⁵ The large error bars indicate that the two independent trajectories did not converge to the same result, a sign that one or both of the independent trajectories were stuck in local minima on the 0.1 μ s/replica timescale of the TReX simulation. We show this data to emphasize the point that TReX simulations on these non-converged timescales were used to identify errors in standard force fields, and thus informed the development of modified force fields for unfolded proteins and IDPs^{27, 38}.

Since more recently it has become standard to perform TReX simulations using on the order of 1 μ s/per replica in recent IDP studies^{6, 21, 44, 61, 62}, we conducted one additional trajectory using a TReX simulation out to 800 ns per replica for each peptide using Amberff99SB + TIP4P-Ew, with the resulting structural data in Figure 2. Our TReX simulation of 0.8 μ s is generally consistent with the 1.0 μ s per replica TReX simulations performed by Rosenman *et al.*⁴⁴ With longer simulation timescales, the previously significant differences between A β 42 and A β 43 are reduced to being within statistical error for all secondary structure categories. While there is a clear improvement in some structural properties at this longer timescale using TReX, such as reduction in the population of α -helices, the β -sheet propensities still show large variability of ± 20 % at residues 5-6, 19, and 31. As a result, their structural ensembles are in stark disagreement with the available NMR and 2D FRET data by being too collapsed and highly structured. It is therefore very understandable why one would continue to conclude that there is a deficiency in the standard force fields based on evidence such as Figure 2.

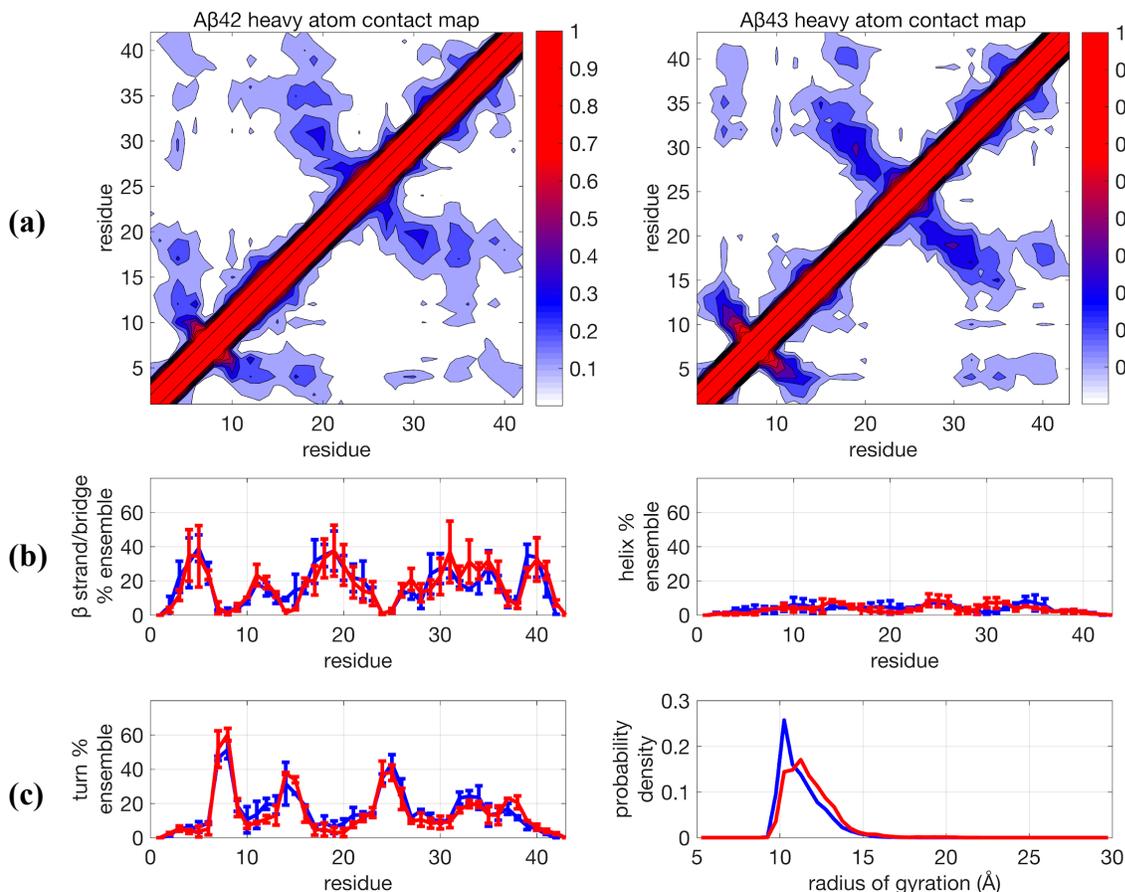


Figure 2. Comparison of structural properties for A β 42 and A β 43 using the force field Amberff99sb + TIP4P-Ew and using Temperature Replica Exchange (TREx) simulated at 0.8 μ s per replica. (a) Contact maps; (b) β -strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b) and (c) blue lines are A β 42, and red lines are A β 43. Further details are given in Figure 1 caption.

Furthermore, the TREx + Amberff99SB + TIP4P-Ew A β ensembles strongly disagree with the A β ensemble generated by Lin *et al.*⁴⁹ In their more extensive MD/MSM simulation they found negligible β -sheet content, with the only persistent secondary structure being \sim 10- 20 % α -helical content for residues 12-18. While the MD/MSM study used the same peptide force field as in this work (Amber ff99SB), they performed their simulations with the TIP3P water model¹⁷ instead of TIP4P-Ew, which likely accounts for some quantitative differences in the A β ensembles. A study of the folding of Trp-cage using Amber ff99SB found that TIP3P enriched sampling of helical content by \sim 10 % and reduced sampling of β -sheet content by a similar amount when compared to TIP4P-Ew.⁷³ These relatively minor population shifts are too small to explain the large discrepancies between the TREx and MD/MSM results, so much of the differences between the ensembles must result from the sampling efficiency. We believe this demonstrates that even on the μ s timescale, TREx simulations are not able to fully capture the structural ensemble of A β peptides.

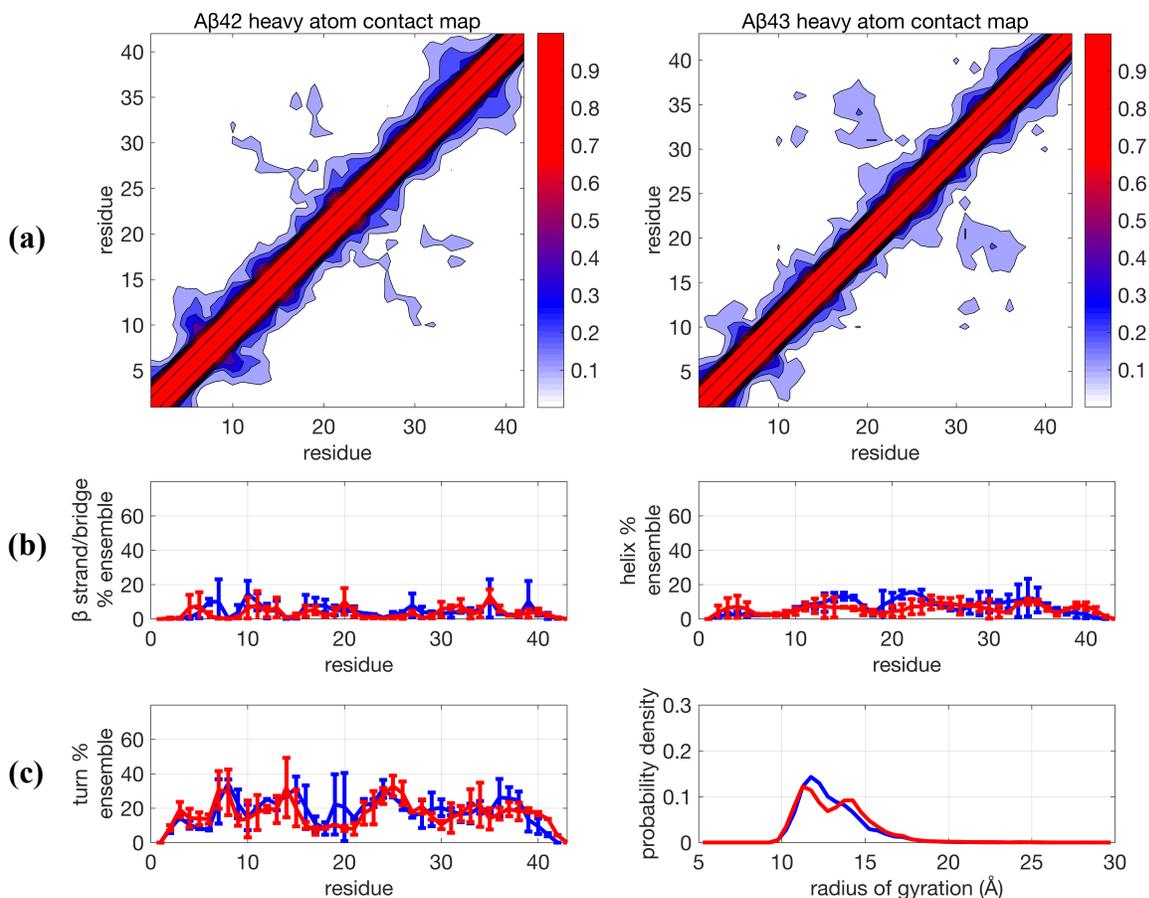


Figure 3. Comparison of structural properties generated for A β 42 and A β 43 using the force field combination Amber99SB + TIP4P-Ew and using Temperature Cool Walking (TCW). (a) Contact maps; (b) β strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b) and (c) blue lines are A β 42, and red lines are A β 43. Further details are given in Figure 1 caption.

We next consider whether an alternative sampling, namely the TCW method, would converge faster on the ~ 0.1 - $0.2 \mu\text{s}$ /replica timescale to evaluate the different force field combinations (Figure 3 and Table 1). We have established in previous work that TCW is superior to the TReX approach using test systems where quantitative measures of convergence are available^{46, 50}. In addition, Figures S2 and S3 show that the TCW method on the $\sim 0.2 \mu\text{s}$ timescale reaches comparable convergence to the $\sim 0.8 \mu\text{s}$ TReX simulation for all secondary structure categories. But as seen when comparing Figures 2 and 3, the TReX and TCW enhanced sampling methods yield very different A β peptide ensembles using the same unmodified protein-water force field combination. More specifically, the TCW + ff99SB + TIP4P-Ew result is far less structured than that found with TReX and is much more similar to the MD/MSM results by Lin *et al.*⁴⁹ Although there are regions of the sequence that exhibit larger uncertainties in the turn population, this variation in turn content suggests transient sampling of still overall unstructured conformations since the sampling of β and helical conformations is consistently low, rather than the more

significant sampling of highly structured conformations seen in the TReX simulations. For the unmodified force field combination using TCW, there is a very small amount ($\sim 5-10\%$) for the contact region formed by residues 16-20 and 30-37 for both A β 42 and A β 43. It is unlikely that the experimental NOEs could absolutely rule out the presence of such a small population of long-range structure, and in fact the simulated NOEs would support this conclusion since the average NOE distance for these residues would be dominated by the $\sim 90\%$ of the unstructured populations.

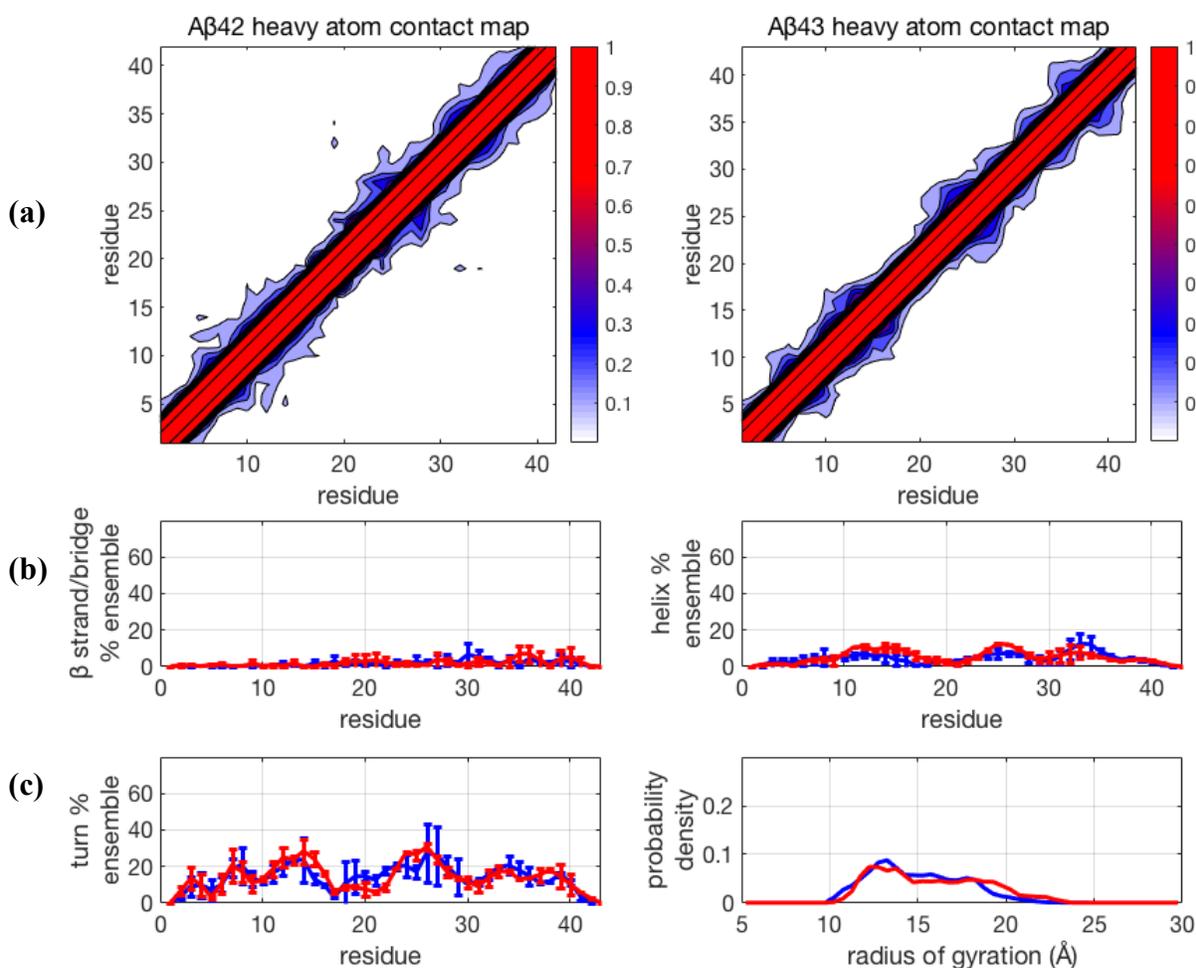


Figure 4. Comparison of structural properties generated for A β 42 and A β 43 using the force field Amber99SB-ILDN + TIP4P-D and using Temperature Cool Walking (TCW). (a) Contact maps; (b) β -strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b) and (c) blue lines are A β 42, and red lines are A β 43. Further details are given in Figure 1 caption.

For completeness, we consider the newer force fields that have been shown to improve descriptions of IDP ensembles using TReX, but now simulated with TCW. We first consider the Amber ff99SB-ILDN + TIP4P-D combination (Figure 4), where the force field optimization is centered on increasing the strength of attractive London dispersion forces on the water oxygen atoms by $\sim 50\%$ relative to that of other four-site water models.

Using TReX simulations, this has been reported to produce less structured IDP ensembles by increasing the strength of protein-water attractions relative to protein-protein attractions using standard Lennard-Jones mixing rules²². The results for TCW shown in Figure 4 also confirm that this recently introduced water model generally reduces structural order and maintains structural similarity between the ensembles of the two peptides, with $\langle R_g \rangle \sim 14.9 \text{ \AA}$ and 15.7 \AA for A β 42 and A β 43, respectively (Table 1). There is again, as in the TCW + ff99SB + TIP4P-Ew result, some variability in the turn population, suggesting sampling of some different unstructured substates in the different trajectories, but the β -sheet and α -helical populations are consistently small.

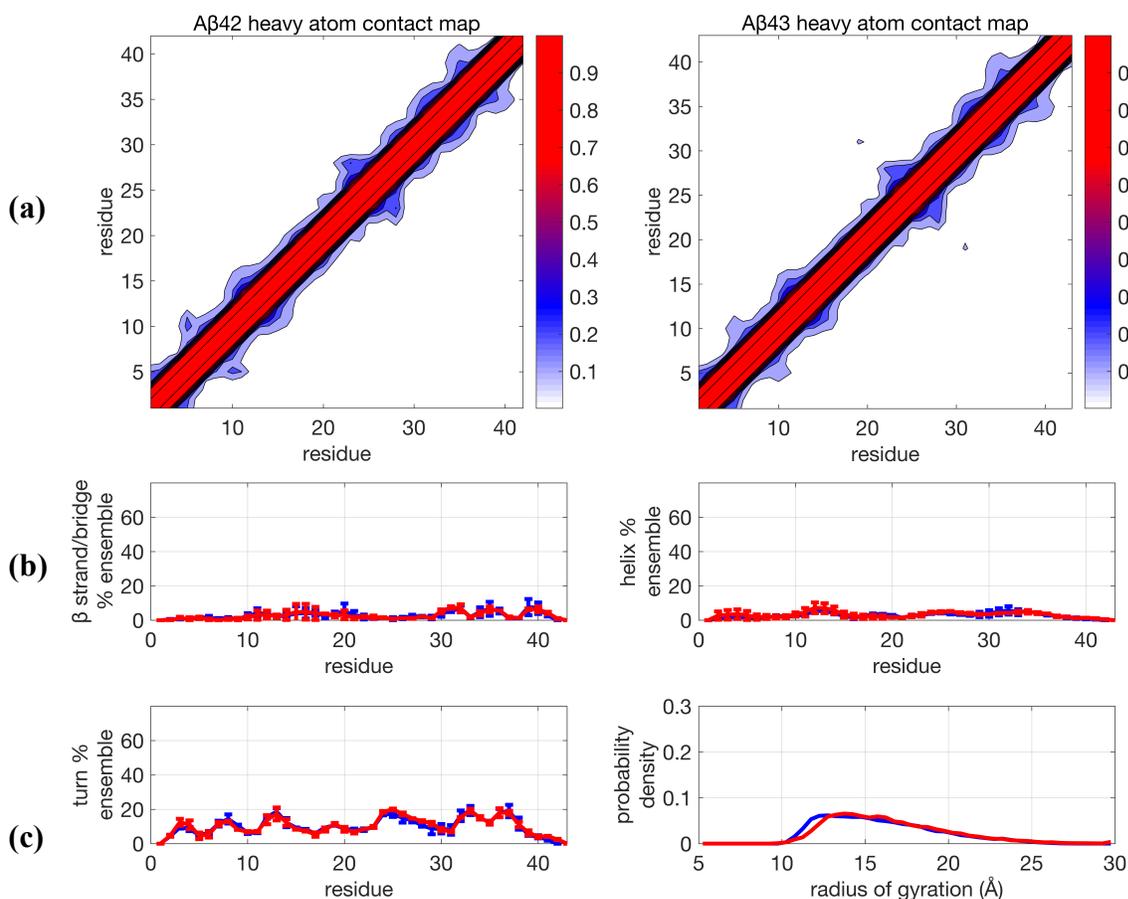


Figure 5. Comparison of structural properties generated for A β 42 and A β 43 using the force field CHARMM36m + CHARMM-TIP3P and using Temperature Cool Walking (TCW). (a) Contact maps; (b) β -strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b) and (c) blue lines are A β 42, and red lines are A β 43. Further details are given in Figure 1 caption.

Similar structural results are obtained for the CHARMM36m + CHARMM-TIP3P force field combination using the TCW sampling method (Figure 5), predicting more expanded structures as evidenced by a shifted and broader R_g distribution, with $\langle R_g \rangle \sim 15.9\text{-}16.4 \text{ \AA}$ for the two peptides (Table 1), with no long-range contacts found using TReX as demonstrated through the absence of secondary structure signatures as well as the

contact maps, in agreement with the NOEs for A β 42.^{5, 6} Furthermore, the structural ensemble of the A β 42 and A β 43 free monomers are seen to be nearly identical, in agreement with the experimental results of Conicella and Fawzi⁴.

3.4 Experimental Validation of Simulated Ensembles

In order to better validate these results for the different force fields, and to better understand the differences found between TReX and TCW, we consider back-calculations of the NMR and FRET data as a metric for comparing A β 42 and A β 43 ensembles to experiment (Table 1). Unfortunately, current state-of-the-art chemical shift calculators developed specifically for protein applications have large associated intrinsic back-calculation errors, making quantitative comparisons problematic for IDPs. For example, the RMS error for $^1\text{H}_\text{N}$ chemical shifts in the SHIFTX2 calculator is 0.17 ppm⁶⁹, which is much larger than the experimental difference (< 0.05 ppm) between the A β 42 and A β 43 ensembles⁴. Thus, chemical shifts are not an ideal metric to distinguish between the different simulated ensembles, and as expected, the simulations are unable to capture the subtle differences in the $^1\text{H}_\text{N}$ chemical shifts in the C-terminus of the two peptides (examples given in Figure S4)³³.

Table 1 shows the calculated χ^2 parameter from the simulated J -coupling constants ($J_{\text{HN-H}\alpha}$) for each ensemble as compared to experiment, in which lower values of the χ^2 metric indicate better agreement with experimental J -coupling constants, with values near one indistinguishable from the scalar coupling back-calculation error. We compare to the Roche *et al.* experimental dataset⁵ consisting of 38 J_i values for A β 42 and the Conicella and Fawzi⁴ experimental dataset consisting of 22 J_i values for A β 43. The simulated J -couplings from the computational ensembles support the general conclusion that the TCW sampling method is in better agreement with the NMR J -couplings than found for TReX, regardless of force field for both A β 42 and A β 43 peptides, with the standard force field in best quantitative agreement⁴. We note that the χ^2 values appear to be commensurate with the results of Meng *et al.* for A β 42 using 750 ns TReX simulations with Amber ff99SBws ($\chi^2 = 2.89$) and Amber ff03ws ($\chi^2 = 4.58$), two IDP-optimized force fields, using the same experimental data and set of Karplus parameters. However their χ^2 values include a block-averaging error due to variance within the trajectories that depresses the value of the χ^2 (up to $\sim 10\%$). We did not apply this block averaging error since this hides the intrinsic sampling problem we are investigating.

The conclusion that the TCW method yields better agreement with J -couplings is bolstered when using a Bayesian analysis we have developed, the Experimental Inferential Structure Determination (EISD) method⁷². The EISD method is designed to assess agreement given the available experimental J -coupling data (as well as chemical shifts, which we ignore here given their low predictive value) that takes into account the intrinsic experimental and back-calculation uncertainties, through optimization. In this case the optimization occurs within the variance of the Gaussian distributed model for the back-calculation error for Karplus parameters A, B and C in Eq. (1). We analyze each ensemble with EISD against the same set of experimental data used for the χ^2 analysis and find that the ranking of the ensembles does not change from the χ^2 analysis (Table 1), i.e. the available J -coupling data is sufficient for concluding that the TCW ensembles are in better

agreement with the NMR data, with significant better agreement for the standard force field.

Table 1. Simulated properties for the A β 42 and A β 43 peptides for each sampling method-force field combination. Mean and standard deviation averaged over two independent trajectories, except for the 0.8 ms TREx simulation, where block averaging over two 250 ns blocks was used. We noted a systematic shift in J-couplings between the two experimental datasets from Conicella and Fawzi⁴ and Roche *et al.*; we have applied a -0.4 Hz shift to the A β 43 J_i values from Conicella and Fawzi before comparison which would bring the A β 42 results in line with each other, to the simulated J-couplings.

Peptide	Sampling Method and Force Field Combination					
	TREx (0.1 \uparrow s) + Amberff99SB + TIP4P-Ew	TREx (0.1 \uparrow s) + CHARMM36m + CHARMM- TIP3P	TREx (0.8 \uparrow s) + Amberff99SB + TIP4P-Ew	TCW (0.2 \uparrow s) + Amberff99SB + TIP4P-Ew	TCW (0.2 \uparrow s) + CHARMM36m + CHARMM- TIP3P	TCW (0.2 \uparrow s) + Amber99SB- ILDN + TIP4P-D
χ^2 between simulated and experimental ^{4,5} J-coupling constants						
A β 42	3.70	4.65	3.70	2.70	3.01	2.89
A β 43	4.75	3.76	3.65	2.47	2.96	2.71
EISD Score						
A β 42	39.479	3.858	40.014	57.833	47.741	54.937
A β 43	23.234	25.090	36.530	47.488	39.328	45.338
Mean and standard deviation of the end-to-end-distance, $\langle R_{ee} \rangle$ (in \AA)						
A β 42	24.3 \pm 0.6	39.3 \pm 3.0	20.5 \pm 1.7	28.4 \pm 0.9	36.8 \pm 1.3	33.5 \pm 7.9
A β 43	26.5 \pm 3.5	44.4 \pm 1.1	20.3 \pm 0.7	29.1 \pm 2.0	38.2 \pm 1.3	31.9 \pm 0.9
Mean and standard deviation of FRET efficiencies, $\langle E_{FRET} \rangle$						
A β 42	0.93 \pm 0.002	0.64 \pm 0.084	0.96 \pm 0.007	0.88 \pm 0.009	0.71 \pm 0.031	0.77 \pm 0.150
A β 43	0.92 \pm 0.049	0.56 \pm 0.001	0.97 \pm 0.007	0.87 \pm 0.034	0.69 \pm 0.020	0.82 \pm 0.004
Mean and standard deviation of the radius of gyration, $\langle R_g \rangle$ (in \AA)						
A β 42	12.0 \pm 0.3	17.2 \pm 2.2	11.3 \pm 0.0	12.9 \pm 0.1	15.9 \pm 0.5	14.9 \pm 1.6
A β 43	11.8 \pm 0.6	17.8 \pm 0.4	11.8 \pm 0.1	13.2 \pm 0.4	16.4 \pm 0.2	15.7 \pm 2.0

Although we do not invoke a full scale simulation of NOE data as we have done in previous studies^{10, 29, 67, 68}, we can make some qualitative comparisons to the NOE data for A β 42⁵ using the contact maps. Clearly the newer force fields are in excellent agreement with the NOE data. Furthermore, the standard force field using the TCW simulations is also in good agreement with the NOE data, unlike the TREx simulations that contain a very high percentage of long-range contacts. For the unmodified force field combination using TCW, there is a very small amount (\sim 5-10 %) for the contact region formed by residues 16-20 and 30-37 for both A β 42 and A β 43, but as we have already stated above, it is unlikely that the experimental NOEs could absolutely rule out the presence of such a small population of transient long-ranged structure.

Next we consider the comparison of the different sampling and force field combinations to 2D FRET that have been reported recently for A β 42⁶; we again make the reasonable assumption that FRET efficiencies, E_{FRET} , for A β 43 will be nearly identical based on the results of Conicella and Fawzi⁴. In principle E_{FRET} should be calculated for all

conformations using the end-to-end distance between the dyes or tags, $R_{ee}^{tag}(t)$, *i.e.* the simulations should use the same sequence construct as the experiment that includes additional residues and the chemical specifics of the covalently bound dye molecules. The possibility that the IDP ensemble will be perturbed to some degree by these tags, as we have seen previously for the MTSL tag used in EPR studies⁵¹, obscures the means to compare R_{ee}^{untag} , from simulation data of untagged peptides, with experimentally derived values of R_{ee}^{tag} . Hence a model for the missing tags must be developed to make contact with the 2D FRET data.

It is useful to consider the FRET model used by Meng *et al.*⁶ in which the distance between the dyes $R_{ee}^{tag}(t)$ is implicitly accounted for by scaling the R_{ee}^{untag} with an approximation that the effect of the dyes is equivalent to adding N_{tag} additional residues to the sequence length⁷⁴.

$$R_{ee}^{tag}(t) = R_{ee}^{untag}(t) \left(\frac{N+N_{tag}}{N} \right)^{0.5} \quad (4)$$

The origin of the scaling model used for the missing tags (Eq. (4)) and value for N_{tag} has precedent in the literature from a study by McCarney and co-workers, who conducted 1 ns standard MD simulations on a model of Alexa 488 attached to a single cysteine residue; they obtained an ensemble average that does provide support for $N_{tag} = 12$ for a pair of dyes⁷⁵. However, they also noted that this average conceals the chemical nature of the bimodal distribution, which results from an extended conformation corresponding to $N_{tag} \sim 16$, and a collapsed conformation due to hydrophobic interactions between the dye and its linker, with $N_{tag} \sim 4.4$. While McCarney *et al.* note that the collapsed conformation would not likely persist in experiments with sufficient denaturant present, as is the case in many FRET studies including their own⁷⁵, that would not be and is not the case for native, denaturant-free studies, as in the experiments of Meng *et al.* for A β 42⁶ that study the peptide in more native-like conditions.

The resulting untagged peptide simulations that have been scaled using Eq. (4) with $N_{tag}=12$, show a highly skewed distribution with a dominant peak at $E_{FRET} = 1$ (Figure 6) that is in disagreement with the experiment which is peaked around the average FRET efficiency of ~ 0.63 for A β 42. This same difference between the experimental and simulated result is also evident in the supplementary material in Meng *et al.*⁶ using the new IDP force fields developed by Best and co-workers^{23, 27, 38}. The extremely high FRET efficiency is as expected given the relatively short length of the A β peptides and the large Förster radius for the dye pair of Alexa 488 and 647, *i.e.* per Eq. (3) all conformations with the scaled values of R_{ee}^{untag} less than ~ 40 Å will yield $E_{FRET} \sim 1$.

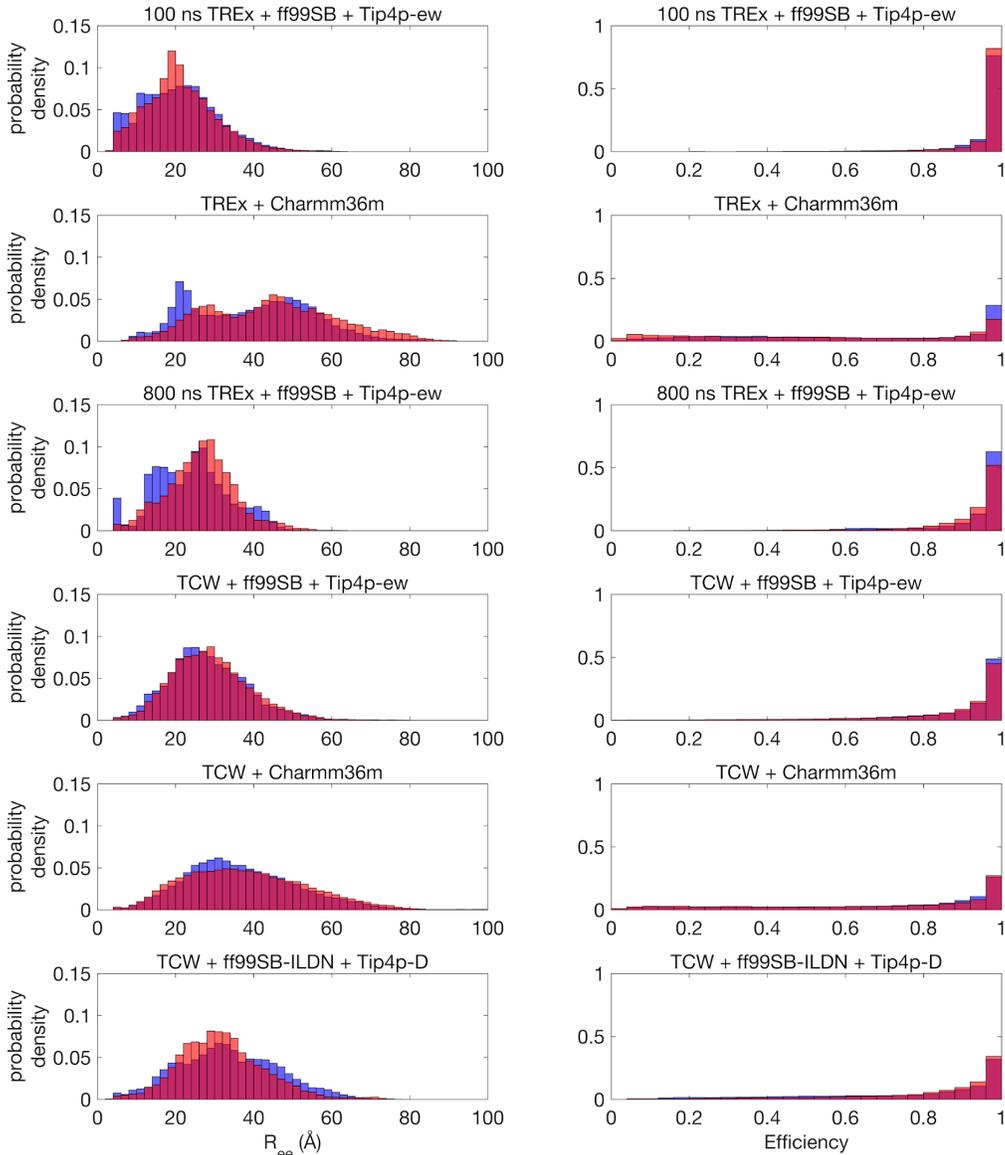


Figure 6. Simulated end-to-end distance histograms (left) and corresponding FRET efficiency histograms using Eq. (3) (right) for all force field- sampling combinations. For all plots, blue bars are for A β 42 and red bars are for A β 43. All simulations are for the untagged peptides, with the results shown having incorporated a shift to model the additional residues and tags.

Even so, the TCW simulations using the different force fields yield FRET efficiencies for A β 42 of $\langle E_{FRET} \rangle = 0.71 - 0.88$ (Table 1); to compare to the work of Meng *et al.* for A β 42, the range of simulated FRET efficiencies they found using two different force fields designed for IDPs with the same value of $N_{tag} = 12$ gave values of $E_{FRET} = 0.68 - 0.83$ which they state is in good agreement with the experimental results.⁶ By contrast the 0.8 μ s TREx simulation using a standard force field yields $E_{FRET} = 0.96 - 0.97$. Increasing N_{tag} to 16 or even 20 would only reduce E_{FRET} by a few percent, hence this general

difference between sampling methods would remain the same. Among the TCW simulations, those using modified force fields produce values of E_{FRET} that are lower and closer to the experimental values of Meng *et al.*, consistent with those force fields producing more expanded ensembles than standard ones.

Given the complete summary of the experimental validation, we also must conclude that the secondary structure propensities for the standard force field for the A β peptides are not as egregiously incorrect as ascertained from the TReX sampling method using a standard 0.8 μ s per replica simulation. There is no question that the newer force fields agree well with the experimental data, but the TCW + Amber ff99SB + TIP4P-Ew also agree as well, and in fact even slightly better for the NMR J-coupling χ^2 and EISD evaluation. Given the uncertainties in the 2D FRET models, we can only conclude that the TCW simulations for the standard force fields are acceptable based on the upper bound value reported by Meng *et al.* ($E_{FRET} = 0.83$)⁶ for one of their modified force fields. Using TCW, the standard force field yields negligible β -sheet and α -helical content, and the only observed long-range contacts are at very low population, and hence qualitatively consistent with the 2D FRET and NOEs taken on the A β 42 peptide, where the 5-10 % population of long-range contacts will likely not be captured in experiment^{5, 6}. As a result the structural ensemble of both peptides generated using the TCW sampling method are more extended than what was found under the TReX protocol for the standard force field, with $\langle R_g \rangle \sim 12.9 - 13.2$ Å for the two peptides (Table 1).

3.5 Discussion

Given the better agreement with all of the experimental data using the TCW protocol over the TReX simulations, we use the TCW results to next address the question as to whether the new modified force fields introduce a genuine improvement over the standard force field. One assessment is whether the force fields are yielding structural ensembles consistent with a random coil ensemble modeled as a Gaussian random chain or a self-avoiding random walk (SARW). Although the A β ensemble using the standard force field does contain a small amount of residual structure, a rapidly interconverting unfolded or IDP ensemble will sample both extended and compact conformations containing regions of secondary structure seen in folded proteins that are still consistent with a random coil ensemble⁷⁶.

The chain length scaling exponents for polymer models used to interpret SAXS, NMR and FRET measurements for unfolded proteins and IDPs have been shown to be dependent on sequence characteristics such as charge and hydrophobicity^{77, 78}. Given that A β has both a net positive and relatively high hydrophobicity in the CHC region of its sequence, the scaling exponent might be expected to reduce to the Θ -limit for this IDP⁶,

$$R_g = R_0 N^\nu \quad (5)$$

where $R_0 = 2.0$ Å as given by Fitzkee and Rose⁷⁹ and $\nu = 0.5$ for the Θ -limit which would yield an $\langle R_g \rangle \sim 13$ Å that is consistent with the value calculated from the unmodified force field results simulated with TCW given in Table 1. Under the Gaussian random coil model we would conclude that the modified force fields have resulted in an over-correction by

producing much more expanded ensembles than is warranted. However, if we assume a SARW model we determine a larger value of for R_g , *i.e.* using $\nu = 0.54 - 0.6$ in a good solvent⁵⁶ would yield $\langle R_g \rangle \sim 15 - 19 \text{ \AA}$, in better agreement with the modified force fields, and thereby suggesting that the standard force fields are in fact too collapsed. Furthermore, for a Gaussian random chain we would expect the following relationship to hold

$$R_g = R_{ee}/\sqrt{6} \quad (6)$$

But this correlation between R_g and R_{ee} is poor given the simulated data in Table 1, except for the TCW + CHARMM36m + CHARMM-TIP3P combination, which is inconsistent with the Gaussian model based on R_g .

Thus we view polymer physics models and analyses as largely inconclusive for differentiating the quality of force fields for this IDP system for several reasons. First is that the differences between the Gaussian and SARW models are better differentiated for much longer polymers than the small A β peptides investigated here. Furthermore, Fuertes *et al.* have presented some novel analytical techniques on a set of IDPs and denatured proteins using a variety of dyes and denaturant conditions to evaluate the relationship between FRET and SAXS⁵⁷. They suggest that one should generally decouple R_g from R_{ee} , and thus avoid using a simple scaling law such as Eq. (6) that is independent of peptide chemistry⁵⁷. Finally, the field is in need of new heteropolymer-centric theoretical models that are able to capture sequence details to extend beyond simple scaling laws and empirical relations, models which are now starting to be developed in recent work for the IDP class of proteins^{80, 81}.

We next turn to a more quantitative assessment using the 2D FRET data to ascertain the differences in force fields. The simulation of FRET efficiencies, especially for IDPs, involves a series of assumptions that introduce uncertainty that must be acknowledged when comparing to the FRET observable given the presence of the fluorescent dyes. As of late 2018, there is active debate on the perturbations introduced by fluorophore tags⁵³⁻⁵⁷. Fuertes *et al.*⁵⁴ assert that there is no perturbation of the structural ensemble across a series of disordered peptides upon addition of dye labels. Part of the basis of their conclusion is that they did not find significant shifts in the SAXS profiles or $\langle R_g \rangle$ of tagged and untagged peptides at high denaturant conditions⁵⁷. If this is the case, then our 2D FRET results and analysis support the view that the standard force field for the untagged A β ensembles and tags modeled using Eq. (4) are in adequate to good agreement with the experiment, with the modified force fields performing only slightly better.

By contrast, Riback *et al.* concluded that at native conditions with no denaturant, the addition of tags in FRET experiments leads to interactions with the IDP that will contribute to FRET signals that overemphasize its contraction, and thus artificially increases its FRET efficiency under denaturant-free conditions⁵⁶. Furthermore, the perturbative effect of the dyes was seen to be larger for the smaller peptides in their experiments, where the addition of labels and the residues to which they are bonded has a greater effect on the mass and resulting dynamics and structure of the peptide. For example, the addition of the tags produced shifts from - 0.3 to + 0.5 nm in the average R_g of the two smallest peptides studied, N49 and NLS, which are natively 36 and 44 residues⁵⁶. Even the results of Meng *et al.* found an increase in E_{FRET} upon explicit representation of the Alexa fluorescent tags

using a standard MD calculation for A β 42⁶, which is consistent with greater compaction of the ensemble. Our own recent work demonstrated that addition of a hydrophobic MTSL-Cys tag to A β 42 can perturb the structural ensemble through interactions between the dye and the peptide, that in turn leads to more a collapsed structural ensemble compared to the original peptide⁵¹. The simulations of Fuertes *et al.*, however, modeled the unlabeled peptide using implicit solvent, then built up an ensemble of structures of the labeled peptide by stochastically adding the dyes to the unlabeled conformers⁵⁷. While this produces a useful test for verifying that many different values of R_{ee} can be obtained from an ensemble with the same R_g , it does not address the potential phenomenon of the dyes affecting the chain dynamics and structure directly. If the tags do induce an artificial compaction of the ensemble, then the true experimental ensembles would be more expanded with even lower FRET efficiencies than reported for A β 42. If that is the case, then the standard force fields yield IDP ensembles that are in fact too collapsed, and force field modifications are warranted, especially for small A β 42 and A β 43 peptides studied in this work.

3.6 Conclusions

We have simulated the disordered structural ensembles of the A β 42 and A β 43 peptides, which according to recent experiments^{4, 5,6} should be largely the same and exhibit no persistent structural ordering or long-range contacts. But two types of error can occur during computational studies of IDP structural ensembles that prevent connections to such experiments, namely statistical sampling error and systematic error in energy and forces. Statistical sampling error occurs when the simulations have not been run sufficiently long to achieve convergence, while systematic error happens when the energy surface of the peptide-water system is not modeled with accurate molecular interactions. Not surprisingly, these two potential errors are intertwined, and hence we have attempted to consider them both by comparing two sampling methods, TREx and TCW, as well as comparing a standard protein-water force field and those that have been recently modified to yield better modeling of disordered states. We have also attempted to validate the various simulated ensembles by comparing to recent state of the art NMR and 2D FRET experiments^{4,5, 6}.

While it is starting to become established that long MD trajectories of ~ 100 μ s to 1 ms are often necessary to reveal force field deficiencies^{11, 49}, simulation timescales that are largely routine only on specialized hardware such as Anton⁸² or Folding@home⁸³, the hope is that better enhanced or accelerated sampling methods might converge more quickly, *i.e.* with one to two orders of magnitude less effort. In this work we have shown that even at μ s timescales there appear to be limitations in the TREx sampling method, producing far more structured ensembles that are in disagreement with NMR and 2D FRET validation data on A β peptides. Changing the sampling method from TREx to TCW produces ensembles that are qualitatively in agreement with J-couplings, NOEs, and FRET efficiencies, regardless of the force field that is simulated. The TCW results for A β 42 are also in very good agreement with the very long sub-millisecond MD/MSM results by Lin and co-workers⁴⁹. Thus our evidence has shown that what was thought to be primarily a force field problem was masked by what is at least in part a problem of poor sampling. This work establishes that the TCW method is more effective than TREx when entropic barriers dominate, and when

applied to IDPs, supports the recent hypothesis that IDPs have an inverted free energy landscape in which disordered conformations are lower in free energy than ordered structures⁵².

Sampling is an important consideration for establishing the transferability of any force field, by demonstrating that appropriate conformational equilibria is reached across a range of thermodynamic conditions in order to describe folded and unfolded states of globular proteins as well as IDP sequences. For example, some standard protein and water force field combinations have proven robust for understanding mechanistic questions about protein folding⁸⁴, which requires a good model of the structure and internal dynamics of the unfolded states in addition to the folded state of a globular protein⁸⁵. But historically, it was the ability to sample multiple folding and unfolding events using these standard force fields which allowed them to gain validation through direct comparisons to robust folding experiments^{86, 87}. This work rescues some of these standard force fields in the sense that they require extensive sampling to definitively show whether they are also capable of simulating accurate IDP ensembles.

The modified force fields may have had the effect of inherently lowering the entropic barriers, and in the best case, still maintaining the folded-unfolded equilibrium. However, unlike the case of protein folding, experimental validation is more limited and underdetermined for IDPs for a variety of reasons. At present there is an impedance mismatch for chemical shifts and FRET data; for chemical shifts the fault is on the theoretical side because of the reduced capability to back-calculate shifts from structure⁷², whereas for FRET measurements on very small IDPs there remains the possibility that the presence of the fluorescent tags might perturb the IDP ensemble from its equilibrium state⁵⁶. However, the impedance match between theory and experiment for scalar couplings and NOEs have provided strong support for the conclusion that, when simulated with TCW or using a very long MD/MSM simulation, the Amberff99SB + TIP4P-Ew, CHARMM36m + CHARMM-TIP3P, and Amberff99SB-ILDN + TIP4P-D are all appropriate force fields for IDPs. While clearly some of the new force field modifications can promote more expanded monomer ensembles to reproduce many experimental IDP properties more expediently, it is important to remember that the IDP-specific force fields³⁸ and other modified force fields may come with their own limitations, such as now manifesting native state instability³³, thereby forgoing the ability to simulate disorder to order transitions in folding upon binding events that are part of the greater functional repertoire of proteins with intrinsic disorder².

3.7 Acknowledgments

We would like to thank Ad Bax, Robert Best, Nick Fawzi, Julie Forman-Kay, Alex MacKerell, Stefano Piana-Agostinetti, and Kevin Plaxco for helpful discussions. We thank Reviewer 1 for very helpful comments and David H. Brookes for early calculations using EISD. SS and THG were supported by the National Institutes of Health (NIH) grant U01GM121667-02 and JL was partially supported on 5R01GM127627-01. J.L. also acknowledges partial support from an NIH Molecular Biophysics training grant, T32-GM008295 during early stages of this study. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This

research also used GPU resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

3.8 References

1. Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L., Function and Structure of Inherently Disordered Proteins. *Curr. Opin. Struct. Bio.* **2008**, *18* (6), 756-64.
2. Dyson, H. J.; Wright, P. E., Intrinsically Unstructured Proteins and Their Functions. *Nature Reviews: Mol. Cell Biol.* **2005**, *6* (3), 197-208.
3. Goedert, M.; Spillantini, M. G., A Century of Alzheimer's Disease. *Science* **2006**, *314* (5800), 777-781.
4. Conicella, A. E.; Fawzi, N. L., The C-Terminal Threonine of A β 43 Nucleates Toxic Aggregation Via Structural and Dynamical Changes in Monomers and Protofibrils. *Biochemistry* **2014**, *53* (19), 3095-3105.
5. Roche, J.; Shen, Y.; Lee, J. H.; Ying, J.; Bax, A., Monomeric A β 1-40 and A β 1-42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochem.* **2016**, *55* (5), 762-775.
6. Meng, F.; Bellaiche, M. M. J.; Kim, J.-Y.; Zerze, G. H.; Best, R. B.; Chung, H. S., Highly Disordered Amyloid-B Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys. J.* **2018**, *114* (4), 870-884.
7. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinform.* **2012**, *80* (2), 556-572.
8. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinformatics* **2013**, *29* (3), 398-399.
9. Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T., Finding Our Way in the Dark Proteome. *J. Amer. Chem. Soc.* **2016**, *138* (31), 9730-42.
10. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-B Peptides. *J. Phys. Chem. B* **2014**, *118*, 6405-6416.
11. Nerenberg, P. S.; Head-Gordon, T., New Developments in Force Fields for Biomolecular Simulations. *Curr. Opin. Struct. Bio.* **2018**, *49*, 129-138.
12. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712-25.
13. Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürki, R.; Geerke, D. P.; Heinz, T. N.; Kastenholtz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F., The Gromos Software for Biomolecular Simulation: Gromos05. *Journal of Computational Chemistry* **2005**, *26* (16), 1719-1751.
14. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the Opls All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225-11236.

15. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B* **2001**, *105* (28), 6474-6487.
16. 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.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998**, *102* (18), 3586-3616.
17. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926-935.
18. Jorgensen, W. L.; Madura, J. D., Temperature and Size Dependence for Monte Carlo Simulations of Tip4p Water. *Molecular Physics* **1985**, *56* (6), 1381-1392.
19. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665-78.
20. Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H.-F., New Force Field on Modeling Intrinsically Disordered Proteins. *Chem. Bio. & Drug Design* **2014**, *84* (3), 253-269.
21. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2016**, *14*, 71.
22. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119* (16), 5113-5123.
23. Best, R. B.; Mittal, J., Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *The Journal of Physical Chemistry B* **2010**, *114* (46), 14916-14923.
24. Nguyen, P. H.; Li, M. S.; Derreumaux, P., Effects of All-Atom Force Fields on Amyloid Oligomerization: Replica Exchange Molecular Dynamics Simulations of the a[Small Beta]16-22 Dimer and Trimer. *Physical Chemistry Chemical Physics* **2011**, *13* (20), 9778-9788.
25. Siwy, C. M.; Lockhart, C.; Klimov, D. K., Is the Conformational Ensemble of Alzheimer's A β 10-40 Peptide Force Field Dependent? *PLoS Comp. Bio.* **2017**, *13* (1), e1005314.
26. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11* (11), 5513-5524.
27. Best, R. B.; Mittal, J., Free-Energy Landscape of the Gb1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79* (4), 1318-1328.
28. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennel, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F., The Gromos Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **1999**, *103* (19), 3596-3607.

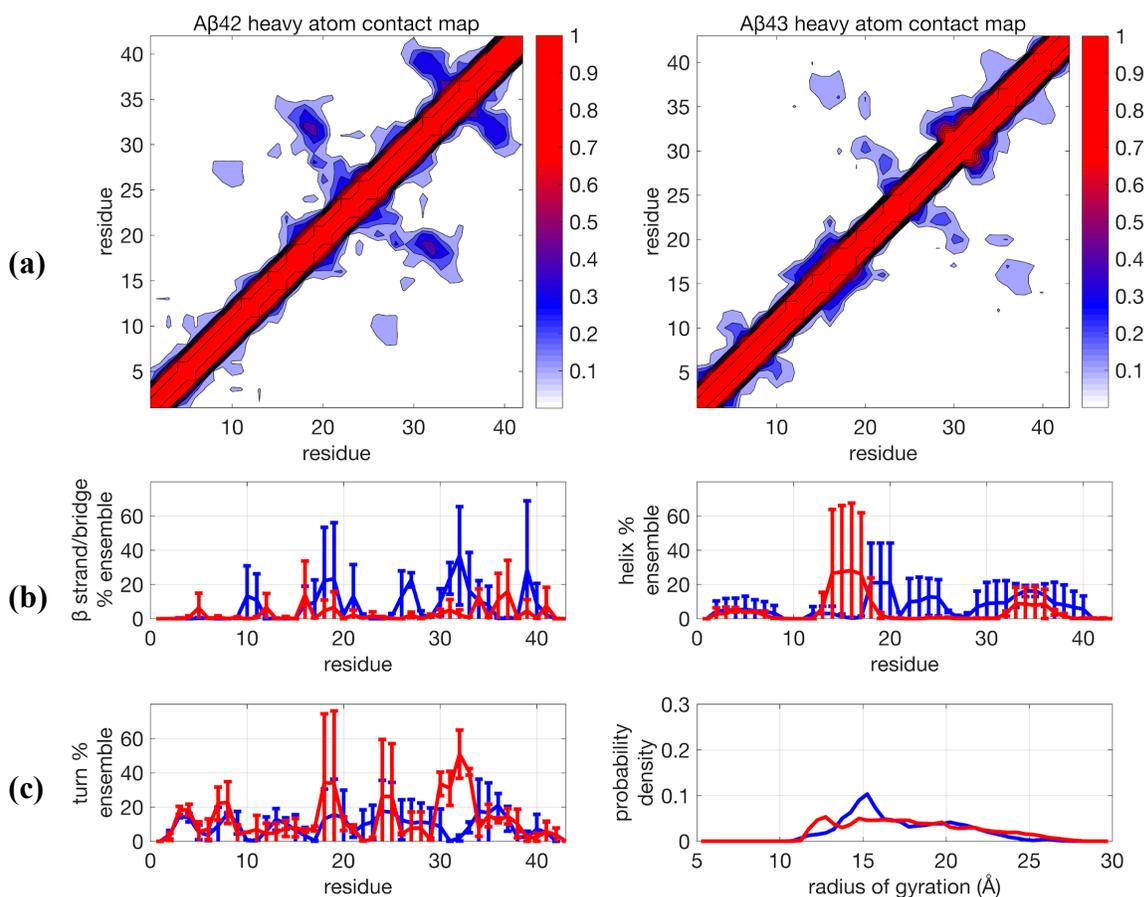
29. Fawzi, N. L.; Phillips, A.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T., Structure and Dynamics of the Alzheimer's Ab21-30 Peptide from the Interplay of Nmr Experiments and Simulation. *Journal of the American Chemical Society* **2008**, *130*, 6145.
30. Song, W. L.; Wang, Y. Y.; Colletier, J. P.; Yang, H. Y.; Xu, Y. C., Varied Probability of Staying Collapsed/Extended at the Conformational Equilibrium of Monomeric a Beta(40) and a Beta(42). *Sci. Rep.* **2015**, *5*.
31. Amini, Z.; Fatemi, M. H.; Rauk, A., Molecular Dynamics Studies of a Beta-Sheet Blocking Peptide with the Full-Length Amyloid Beta Peptide of Alzheimer's Disease. *Can. J. Chem.* **2016**, *94* (10), 833-841.
32. Viet, M. H.; Nguyen, P. H.; Derreumaux, P.; Li, M. S., Effect of the English Familial Disease Mutation (H6r) on the Monomers and Dimers of a Beta 40 and a Beta 42. *ACS Chem. Neurosci.* **2014**, *5* (8), 646-657.
33. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (21), E4758-E4766.
34. Nerenberg, P. S.; Head-Gordon, T., Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J Chem Theory Comput* **2011**, *7* (4), 1220-30.
35. Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T., Optimizing Solute-Water Van Der Waals Interactions to Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116* (15), 4524-34.
36. Henriques, J.; Cragnell, C.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11* (7), 3420-3431.
37. Henriques, J.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the Tip4p-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016**, *12* (7), 3407-3415.
38. Best, R. B.; Zheng, W.; Mittal, J., Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113-5124.
39. Abrams, C.; Bussi, G., Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16* (1).
40. Bernardi, R. C.; Melo, M. C. R.; Schulten, K., Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta* **2015**, *1850* (5), 872-877.
41. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314* (1-2), 141-151.
42. Hansmann, U., Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140-150.
43. Zuckerman, D. M.; Lyman, E., A Second Look at Canonical Sampling of Biomolecules Using Replica Exchange Simulation. *Journal of Chemical Theory and Computation* **2006**, *2*, 1200-1202.
44. Rosenman, D. J.; Wang, C.; García, A. E., Characterization of A β Monomers through the Convergence of Ensemble Properties among Simulations with Multiple Force Fields. *J. Phys. Chem. B* **2016**, *120* (2), 259-277.

45. Nymeyer, H., How Efficient Is Replica Exchange Molecular Dynamics? An Analytic Approach. *Journal of Chemical Theory and Computation* **2008**, *4* (4), 626-636.
46. Brown, S.; Head-Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J. Comput. Chem.* **2003**, *24* (1), 68-76.
47. Hicks, A.; Zhou, H.-X., Temperature-Induced Collapse of a Disordered Peptide Observed by Three Sampling Methods in Molecular Dynamics Simulations. *J. Chem. Phys.* **2018**, *149* (7), 072313.
48. Do, T. N.; Choy, W.-Y.; Karttunen, M., Binding of Disordered Peptides to Kelch: Insights from Enhanced Sampling Simulations. *J. Chem. Theory Comput.* **2016**, *12* (1), 395-404.
49. Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S., Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophysical Journal* **2012**, *102* (2), 315-324.
50. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J. Chem. Phys.* **2016**, *145* (17).
51. Sasmal, S.; Lincoff, J.; Head-Gordon, T., Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-Beta. *Biophys. J.* **2017**, *113* (5), 1002-1011.
52. Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M., The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments. *Sci. Rep.* **2015**, *5*, 15449.
53. Riback, J. A.; Bowman, M. A.; Zmyslowski, A.; Knoverek, C. R.; Jumper, J.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R., Response to Comment on "Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water". *Science* **2018**, *361* (6405), eaar7949.
54. Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A., Comment on "Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water". *Science* **2018**, *361* (6405), eaau8230.
55. Best, R. B.; Zheng, W.; Borgia, A.; Buholzer, K.; Borgia, M. B.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Schuler, B., Comment on "Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water". *Science* **2018**, *361* (6405), eaar7101.
56. Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Knoverek, C. R.; Jumper, J. M.; Hinshaw, J. R.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R., Innovative Scattering Analysis Shows That Hydrophobic Disordered Proteins Are Expanded in Water. *Science* **2017**, *358* (6360), 238.
57. Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Estrada Girona, G.; Milles, S.; Mishra, A.; Onck, P. R.; Gräter, F.; Esteban-Martín, S.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A., Decoupling of Size and Shape Fluctuations in Heteropolymeric Sequences Reconciles Discrepancies in Saxs Vs. Fret Measurements. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (31), E6342.
58. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry* **2005**, *26* (16), 1668-1688.
59. Rathore, N.; Chopra, M.; de Pablo, J. J., Optimal Allocation of Replicas in Parallel Tempering Simulations. *The Journal of Chemical Physics* **2004**, *122* (2), 024111.

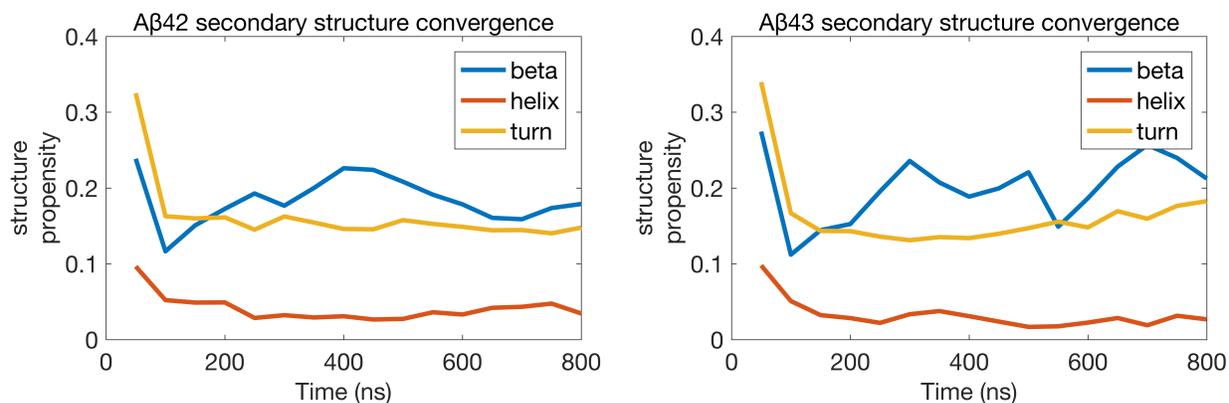
60. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins: Struct., Func., Bioinform.* **2010**, *78* (8), 1950-1958.
61. Sgourakis, N. G.; Yan, Y. L.; McCallum, S. A.; Wang, C. Y.; Garcia, A. E., The Alzheimer's Peptides a Beta 40 and 42 Adopt Distinct Conformations in Water: A Combined Md/Nmr Study. *Journal of Molecular Biology* **2007**, *368* (5), 1448-1457.
62. Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C. Y.; Garcia, A. E., Abeta Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined Md/Nmr Approach. *J. Mol. Bio.* **2013**, *425* (18), 3338-3359.
63. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; T.E. Cheatham, I.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., Amber 14. *University of California, San Francisco* **2014**.
64. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D., Openmm 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9* (1), 461.
65. Roe, D. R.; Cheatham, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-3095.
66. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577-2637.
67. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-Beta Peptides. *Biochemistry* **2011**, *50* (35), 7612-28.
68. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in B-Strand Populations of Monomeric Amyloid-B 40 and Amyloid-B 42. *Biophys. J.* **2013**, *104* (12), 2714-2724.
69. Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S., Shiftx2: Significantly Improved Protein Chemical Shift Prediction. *Journal of biomolecular NMR* **2011**, *50*, 43-57.
70. Vuister, G. W.; Bax, A., Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(Hnh.Alpha.) Coupling Constants in ¹⁵N-Enriched Proteins. *Journal of the American Chemical Society* **1993**, *115*, 7772-7777.
71. Vögeli, B.; Ying, J.; Grishaev, A.; Bax, A., Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *Journal of the American Chemical Society* **2007**, *129* (30), 9377-9385.
72. Brookes, D. H.; Head-Gordon, T., Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. *J. Amer. Chem. Soc.* **2016**, *138* (13), 4530-8.
73. Paschek, D.; Day, R.; García, A. E., Influence of Water-Protein Hydrogen Bonding on the Stability of Trp-Cage Miniprotein. A Comparison between the Tip3p and Tip4p-Ew Water Models. *Physical Chemistry Chemical Physics* **2011**, *13* (44), 19840-19847.

74. Zheng, W.; Borgia, A.; Borgia, M. B.; Schuler, B.; Best, R. B., Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *J. Chem. Theory Comput.* **2015**, *11*, 5543.
75. McCarney, E. R.; Werner, J. H.; Bernstein, S. L.; Ruczinski, I.; Makarov, D. E.; Goodwin, P. M.; Plaxco, K. W., Site-Specific Dimensions across a Highly Denatured Protein; a Single Molecule Study. *Journal of Molecular Biology* **2005**, *352* (3), 672-682.
76. Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M., The Concept of a Random Coil: Residual Structure in Peptides and Denatured Proteins. *Folding and Design* **1996**, *1* (5), R95-R106.
77. Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J.-r.; Labhardt, A. M.; Grzesiek, S.; Schuler, B., Comprehensive Structural and Dynamical View of an Unfolded Protein from the Combination of Single-Molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (37), E5389-E5398.
78. Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B., Polymer Scaling Laws of Unfolded and Intrinsically Disordered Proteins Quantified with Single-Molecule Spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (40), 16155-16160.
79. Fitzkee, N. C.; Rose, G. D., Reassessing Random-Coil Statistics in Unfolded Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (34), 12497-12502.
80. Lin, Y.-H.; Chan, H. S., Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. *Biophys. J.* **2017**, *112* (10), 2043-2046.
81. Firman, T.; Ghosh, K., Sequence Charge Decoration Dictates Coil-Globule Transition in Intrinsically Disordered Proteins. *J. Chem. Phys.* **2017**, *148* (12), 123305.
82. Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E., Systematic Validation of Protein Force Fields against Experimental Data. *PLOS ONE* **2012**, *7* (2), e32131.
83. Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B., Atomistic Protein Folding Simulations on the Submillisecond Time Scale Using Worldwide Distributed Computing. *Biopolymers* **2002**, *68* (1), 91-109.
84. Piana, S.; Klepeis, J. L.; Shaw, D. E., Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Bio.* **2014**, *24*, 98-105.
85. Echeverria, I.; Makarov, D. E.; Papoian, G. A., Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *J. Amer. Chem. Soc.* **2014**, *136* (24), 8708-8713.
86. Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späh, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B., Quantifying Internal Friction in Unfolded and Intrinsically Disordered Proteins with Single-Molecule Spectroscopy. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (44), 17800-17806.
87. Eaton, W. A.; Wolynes, P. G., Theory, Simulations, and Experiments Show That Proteins Fold by Multiple Pathways. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (46), E9759-E9760.

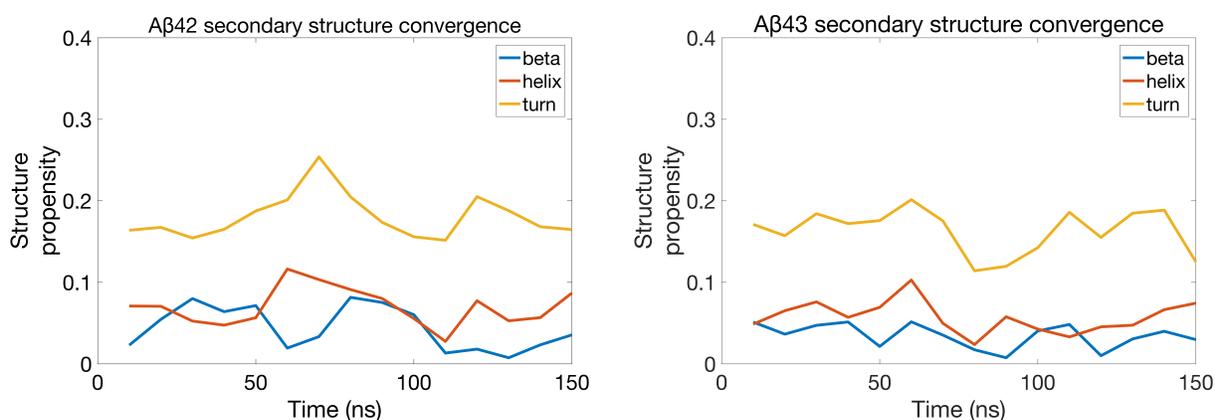
3.9 Appendix



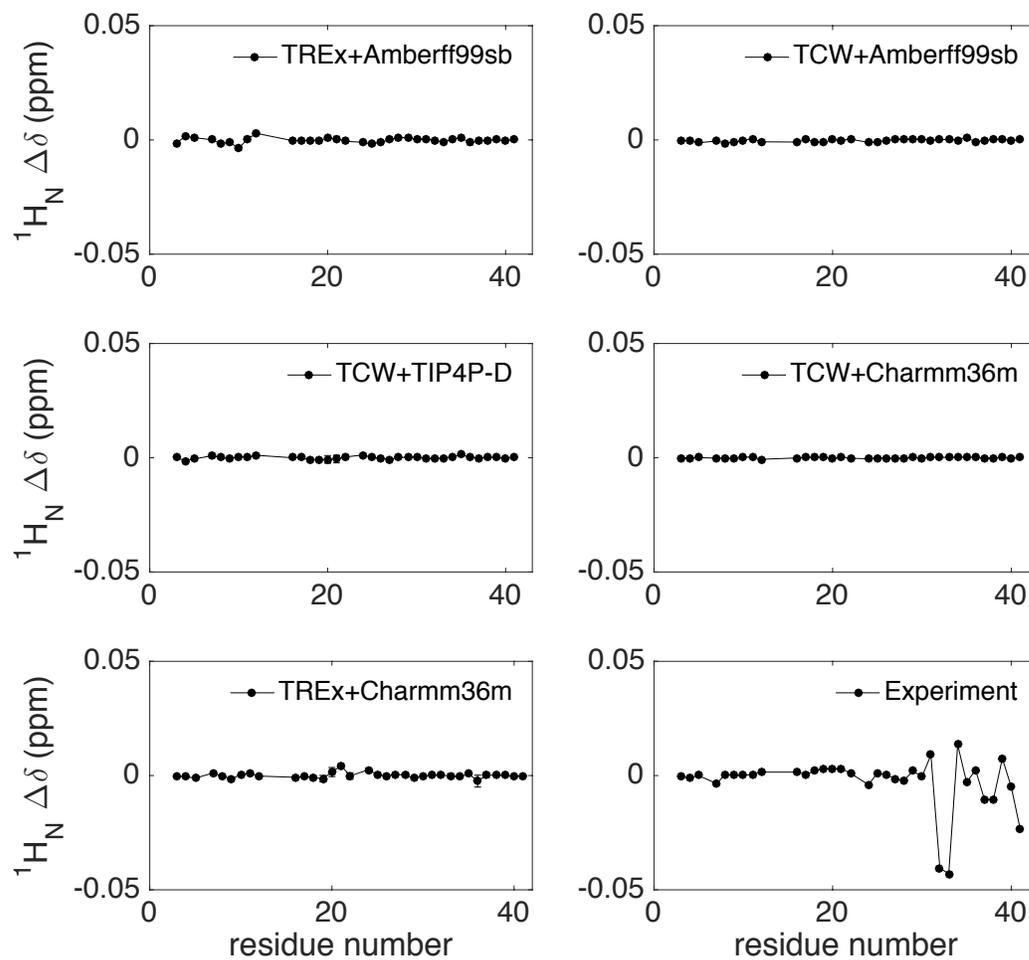
Supplementary Figure 1: Comparison of structural properties for Aβ42 and Aβ43 using the force field *Charmm36m + Charmm-TIP3P* and using *Temperature Replica Exchange (TREx)* simulated at $0.1 \mu\text{s}$ per replica. (a) Contact maps; (b) β-strand and helix propensities, (c) turn propensity and radius of gyration distribution. For (b) and (c) blue lines are Aβ42, and red lines are Aβ43. Further details are given in Figure 1 caption.



Supplementary Figure 2: Convergence of secondary structure propensities as a function of time using the force field Amber99SB + TIP4P-Ew with TReX simulations carried out to 800 ns. (a) Aβ42 and (b) Aβ43. Secondary structure propensities are block-averaged in 50 ns bins. While local secondary structure appears stable, properties such as β -sheet populations are not fully converged.



Supplementary Figure 3: Convergence of secondary structure propensities as a function of time using the force field Amberff99SB + TIP4P-Ew with Temperature Cool Walking (TCW) simulations carried out to 200 ns. (a) Aβ42 (b) Aβ43. First 50 ns discarded as equilibration, production segment of 150 ns block-averaged into 10 ns bins. Secondary structure content appears stable over course of simulation with each type for the two peptides fluctuating about a mean.



Supplementary Figure 4: Comparison of simulated chemical shifts difference between the $A\beta_{43}$ and $A\beta_{42}$ structural ensembles using different combinations of force field and sampling techniques with experiments. The simulated values were calculated using the SHIFTX2 calculator.

Chapter 4

Modeling of α -Helix Formation Across a Range of Temperatures by a Polarizable Force Field in Comparison with Diverse Fixed-Charge Force Fields

We present initial results evaluating the ability of standard and modified fixed charge force fields, as well as one polarizable force field, to model the significant increase in α -helix content sampled by the model (AAQAA)₃ with decreasing temperature, a commonly used though rarely perfectly modeled target in force field parameterization. This testing is meant to evaluate the transferability of the different force fields, inspired by recent work noting deficiencies of standard fixed charge force fields in modeling disordered states, and of force fields modified to induce disorder in modeling more folded states. Inclusion of polarization is hypothesized to improve modeling along the full disorder to order spectrum due to the more sophisticated treatment of electrostatics than fixed charge force fields, potentially leading the way to a truly generalizable force field for all proteins that may have high applicability once greater computational efficiency for these models is achieved. Our initial results using the polarizable force field are limited due to lack of convergence resulting from this high computational expense, and as such this work is being continued in the group. We do nonetheless confirm that most fixed charge models do not demonstrate significant change across temperature, and find some initial evidence that the polarizable model tested may, pending confirmation with more extensive testing, outperform them.

4.1 Introduction

Simulation of intrinsically disordered proteins (IDPs)¹ is necessary for their characterization due to strong time- and ensemble-averaging present in collection of experimental results², which complicates and generally renders difficult the generation of a fully resolved conformational ensemble from experiment alone^{3, 4}. As established in previous chapters, however, simulation of IDPs is not without its own model and methodological uncertainties. In Chapter 2 we describe development and validation of a novel enhanced sampling algorithm, temperature cool walking (TCW)⁵, designed to overcome limitations in computational cost and sampling efficiency⁶ of the most commonly used algorithm for simulation of IDPs^{7, 8}, temperature replica exchange (TREx)⁹. Such improvement in sampling efficiency, combined with the ability to run simulations longer due to lower cost, are key in ensuring that the resulting ensembles are properly converged. In Chapter 3 we apply TCW¹⁰, in comparison to TREx, to studying the differences in performance between a standard force field^{11, 12} and recently developed force fields¹³⁻¹⁶ that were modified to improve modeling of disordered states. Our findings¹⁰ that the force fields were not as divergent from each other upon improvement using TCW sampling suggest that the previously identified issues with standard force fields producing overly compact ensembles may have been over-estimated since the agreed better with J-coupling data, while the CHARMM36 and D.E. Shaw group force fields did produce more extended

ensembles that may be in better agreement with FRET experiments. This continued ambiguity in the modeling of the fully solvated monomer state of IDPs has also arisen in characterization of the disorder to the folded state of proteins⁸. In fact the utility of the modified force fields for modeling partially ordered IDP states that are key in their interaction with stable folded receptors¹⁷, or during aggregation¹⁸ brings into question whether any standard fixed charge force field would be capable of modeling the full range of IDP behavior.

One possible change in the functional form used that might improve modeling of the disorder-to-order transition is the inclusion of polarization physics to account for the variety in environmental responses¹⁹. Classical force fields use a point charge model for electrostatics. While this model is highly successful in many cases, it is possible that—as is suggested by the difficulty in developing a general-use fixed charge force field⁸—a more sophisticated treatment of electrostatics may be needed in order to accurately model the disorder-to-order spectrum. Polarizable force fields such as the AMOEBA¹⁹ and CHARMM Drude²⁰ force fields may, albeit at much higher computational expense, provide the key improvement in physical accuracy because they are formulated to allow for the charge density of molecules to change and respond to, and in turn affect, their electrostatic environment. Indeed, one study²¹ of the 15-residue synthetic peptide (AAQAA)₃²², which is often used as a model of increasing helix formation at lower temperature²³, found evidence of cooperative helix formation using the CHARMM-Drude force field that was not found using the fixed charge equivalent. Our own work²⁴ on the disordered Histatin-5 peptide found that the polarizable Amoeba force field, run with TCW, better modeled the radius of gyration distribution and secondary structure of the peptide than recently developed force fields optimized for the simulation of IDPs.

Here I present preliminary results continuing such work, making use of the lower computational cost of TCW^{5, 10} relative to TReX or long MD simulations to investigate modeling of helix formation by two standard force fields, two IDP-optimized force fields, and one polarizable force field in the model (AAQAA)₃ peptide across a wide range of temperatures. We compare against the expected fraction of α -helix based on chemical shifts from nuclear magnetic resonance (NMR) spectroscopy²², in which the synthetic peptide displays a strong response in helix content with changing temperature, going from no helix above 360 K to above 50 % at 280 K.

Our results confirm the difficulty in accurately modeling the helix formation of this peptide, with the fixed-charge force fields displaying no increase in helical content with decreasing temperature, in agreement with other simulation studies on the same and similar force fields^{8, 25}. While the results from our simulations using the polarizable force field¹⁹ need further confirmation with longer and repeated simulations, our initial study suggests that—as has been found with another polarizable force field²¹—these models may be better-suited to the modeling of these complex processes where a peptide samples electrostatically distinct environments as it undergoes folding.

4.2 Materials and Methods

The initial fully α -helical structure of (AAQAA)₃, capped with an acetyl group at the N-terminus and a methylamine group at the C-terminus, was created in the *tleap*¹¹ module of Amber, and solvated in water to produce a periodic box with sides of length 6.0 nm using

either *tleap* or the web interface of CHARMM-GUI²⁶. The force fields tested were 1) Amber ff99SB¹¹ for the protein with TIP4P-Ew¹² water, a standard fixed-charge force field combination that is widely used, though expected to under-sample the α -helix for this system²⁵ 2) Amber ff99SB-ILDN¹⁵ with TIP4P-d¹⁵ water, a recent disorder-optimized combination that in one study using TReX sampled negligible helical content across all temperatures⁸ 3) CHARMM36²⁷ + C-TIP3P¹⁴, another standard combination that has in various studies sampled 12 % α -helix at 300 K^{13, 28} 4) CHARMM36m¹³ + C-TIP3P¹⁴, a modified version of CHARMM36 designed to improve sampling of disordered states while maintaining a stable fold for other systems, with observed α -helix content ranging from 11 to 17 % at 300 K depending on the study^{8, 13} and 5) AMOEBA-13¹⁹ + Water03²⁹, a polarizable force field that was shown to sample between 10 and 30 % α -helix content per residue at 303 K¹⁹, albeit using much shorter TReX simulations than are now standard.

All production simulations were conducted using TCW^{5, 10} implemented in OpenMM³⁰, using a high walker temperature of 420 K. A disadvantage of TCW in comparison to TReX is that while TReX naturally leads to the generation of equilibrium ensembles at every simulated temperature, TCW only does so for the low temperature of interest. This is only a burden for tests such as these that involve study of behavior across a range of temperatures, as opposed to that described in previous chapters^{5, 10}. However, the overall cost savings of TCW are still apparent, and we benefit from greatly reducing the amount of simultaneously needed computational power, as each simulation can be run on a single GPU rather than ~ 50 . For each force field, we perform 6 different TCW simulations, with low temperatures of either 280, 300, 320, 340, 360 or 380 K, to generate ensembles that span the variation in helical content for the peptide²². Cool walking exchanges are initiated an average of once per ns of low temperature propagation, with 200 fs of annealing at each temperature. For each force field and low temperature, the system was first equilibrated in the NPT ensemble for 1 ns with restraints on peptide heavy atoms to maintain the initial conformation, and then the restraints relaxed from 10 kcal/mol/Å² to 0 kcal/mol/Å² over 100 ps before beginning production simulations.

Trajectories were analyzed using the Amber program *cpptraj*³¹. As is consistent with the literature on simulations of this system^{8, 13, 21, 23, 25, 28}, a residue was defined as being in an α -helical combination in a conformation if it was in a span of three continuous residues all within a defined (φ, ψ) dihedral range of φ between -160 and 20° and ψ between -120 and 50° . This definition has been found to be generally equivalent to use of DSSP³². The helical content for the entire peptide was calculated as the average of the per-residue helical content. These results were compared to previous results from the literature using the same force fields, and the experimentally predicted amounts of helical content based on NMR chemical shifts²².

4.3 Results

Simulations using fixed charge force fields were all run for a minimum of 400 ns, which should, except potentially in the case of the lowest temperatures, be sufficient for decent convergence based as estimated from previous work using TCW^{5, 10, 33}. The α -helix fractions averaged across the peptide and simulation length for each simulation are reported in Table 1. Uncertainties in the helix fraction for fixed-charge trajectories were calculated as the standard deviation in the cumulative average calculated using 10 ns blocks and

discarding the first 50 ns as equilibration, as has been done in previous work on this system²⁸.

Simulations using AMOEBA, given the much higher computational cost of using a polarizable model¹⁹, were run for 100 ns for the lower temperatures and 80 ns for 360 and 380 K. This is not expected to be fully sufficient to converge. A very rough estimate of behavior upon better convergence can be obtained by calculating the change in helix fraction for the fixed-charge force fields between the first 100 ns and the entirety of the simulation, as shown in Supplementary Table 1. Helix fraction tends overall to decrease upon changing from an estimate based on the first 100 ns relative to the whole simulation, as expected given the initial condition. However, a wide range of shifts is observed, with no strong trend apparent over either temperature or force field, except for slightly greater variability for TIP4P-D than the others. The mean decrease for the fixed-charge simulations is to a helix fraction 2/3 the value of the first 100 ns simulation, albeit with wide variability.

Table 1: *Simulation lengths and ensemble % α -helix for all force fields and temperatures studied.* For fixed-charge force fields, uncertainties were calculated as the standard deviation in 10 ns-blocked values of the cumulative average α -helix fraction, discarding the first 50 ns. For AMOEBA, given the short simulation time, no time was discarded in calculating the cumulative averages and uncertainties, reflecting the relatively lesser convergence and stability.

Force Field		Temperature (K)					
		280	300	320	340	360	380
CHARMM36 + C-TIP3P	Length (ns)	800	900	900	900	450	500
	% helix	4.2 (1.0)	4.2 (0.66)	4.0 (0.60)	4.2 (0.58)	4.8 (1.5)	5.2 (0.63)
CHARMM36m + C-TIP3P	Length (ns)	600	650	850	800	750	700
	% helix	4.4 (0.84)	5.8 (0.97)	4.4 (0.74)	4.5 (1.9)	5.0 (1.7)	4.6 (0.42)
Amber ff99SB + TIP4P-Ew	Length (ns)	550	550	450	550	450	550
	% helix	3.7 (2.4)	3.7 (0.73)	3.4 (0.75)	3.1 (0.46)	3.1 (0.41)	3.2 (0.88)
Amber ff99SB-ILDN + TIP4P-d	Length (ns)	450	500	500	450	550	400
	% helix	4.2 (1.4)	3.9 (0.32)	3.3 (1.6)	3.6 (0.53)	3.8 (0.95)	2.6 (0.22)
AMOEBA + Water03	Length (ns)	100	100	100	100	80	80
	% helix	29 (15)	5.7 (15)	6.7 (15)	6.6 (18)	7.5 (18)	1.8 (4.1)

Our results for the fixed-charge force fields generally agree with the literature results^{8, 13, 21, 23, 25, 28} in trend for the α -helix fraction, which remains relatively flat across all temperatures for nearly all force fields regardless of water model. We and others find less α -helical content with the CHARMM36 + C-TIP3P and CHARMM36m + C-TIP3P force fields than the $\sim 17\%$ at 300K reported using long standard MD¹³; the DE Shaw group using a $\sim 20 \mu\text{s}$ TReX simulation found that percentage dropped to 12%, whereas with TCW we find it to be 4-6%. Interestingly, we find more α -helical content for Amber ff99SB-ILDN with TIP4P-D than found with a previous TReX simulation which obtained zero helix⁸, although again the trends with temperature are flat. Our results for Amber ff99SB with TIP4P-Ew produce α -helical percentages of 2-4%, consistent with previous results using Amber ff99SB and TIP3P, a similar force field pair²⁵. In aggregate, we can rule out the sampling method used as a systematic source of any error, but appears to be a problem for all fixed charge force fields that there is an absence of cooperative disorder to order response over temperature.

At the time of the completion of this chapter, the short length of the AMOEBA simulations and resulting high uncertainties in helical content, render that data inconclusive at present.

4.4 Discussion and Conclusions

We have performed a set of enhanced sampling simulations using TCW^{5,6,10} method on a variety of fixed charge and polarizable force fields, seeking to evaluate whether a polarizable model¹⁹ can at least qualitatively model the sharp increase in helicity as temperature decreases. Despite the computational efficiency of TCW, we now must grapple with the increased computational expense of the polarizable model, highlighting the importance of other work that seeks to make polarization less expensive in simulation, such as implementation of novel integrators from our lab³⁴⁻³⁹.

Our results confirm the findings in the literature that many fixed charge force fields cannot model the sharp increase in helical content with changing temperature. The fact that this same result occurs across such disparate parameterizations of the same functional forms of a fixed-charge model, even given the stated goal of developing a universally appropriate fixed charge force field for proteins across the order-to-disorder spectrum⁸, the (AAQAA)₃ continues to be a challenging system. In fact that group eventually discarded (AAQAA)₃ as a benchmark given the difficulty in designing a force field capable of modeling both it and the other natural peptide systems used.

It is likely that more extensive changes to the functional form of classical force fields, such as the addition of polarization, may be necessary in order to overcome these problems. While our simulations using the polarizable AMOEBA¹⁹ are certainly not at full convergence, we find some evidence that the temperature response for that force field may be somewhat stronger than that of the fixed-charge force fields. This needs to be verified with further simulation to produce better-converged results, in work that is continuing in our lab. This suggests, along with other studies of polarizable models^{21,24}, that while the computational expense remains a limiting factor in the use of polarizable force fields, they may be entirely appropriate and perhaps even most capable of modeling the complex behavior of IDPs.

4.5 References

1. Tompa, P., Intrinsically Disordered Proteins: A 10-Year Recap. *Trends Biochem Sci* **2012**, *37* (12), 509-16.
2. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins* **2012**, *80* (2), 556-72.
3. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-Beta Peptides. *Biochemistry* **2011**, *50* (35), 7612-28.
4. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-Beta Peptides. *J Phys Chem B* **2014**, *118* (24), 6405-16.

5. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J Chem Phys* **2016**, *145* (17), 174107.
6. Brown, S.; Head-Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J Comput Chem* **2003**, *24* (1), 68-76.
7. Meng, F.; Bellaiche, M. M. J.; Kim, J. Y.; Zerze, G. H.; Best, R. B.; Chung, H. S., Highly Disordered Amyloid-Beta Monomer Probed by Single-Molecule Fret and Md Simulation. *Biophys J* **2018**, *114* (4), 870-884.
8. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc Natl Acad Sci U S A* **2018**, *115* (21), E4758-E4766.
9. Yoda, T.; Sugita, Y.; Okamoto, Y., Protein Folding Simulations by Generalized-Ensemble Algorithms. *Adv Exp Med Biol* **2014**, *805*, 1-27.
10. Lincoff, J.; Sasmal, S.; Head-Gordon, T., The Combined Force Field-Sampling Problem in Simulations of Disordered Amyloid-Beta Peptides. *J Chem Phys* **2019**, *150* (10), 104108.
11. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712-25.
12. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J Chem Phys* **2004**, *120* (20), 9665-78.
13. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71-73.
14. 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.; 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., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998**, *102* (18), 3586-616.
15. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins* **2010**, *78* (8), 1950-8.
16. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J Phys Chem B* **2015**, *119* (16), 5113-23.
17. Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D., Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an Scf Ubiquitin Ligase. *Structure* **2010**, *18* (4), 494-506.
18. Sasmal, S.; Schwierz, N.; Head-Gordon, T., Mechanism of Nucleation and Growth of Abeta40 Fibrils from All-Atom and Coarse-Grained Simulations. *J Phys Chem B* **2016**, *120* (47), 12088-12097.
19. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-Based Amoeba Force Field for Proteins. *J Chem Theory Comput* **2013**, *9* (9), 4046-4063.

20. Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D., Jr., Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J Chem Theory Comput* **2013**, *9* (12), 5430-5449.
21. Huang, J.; MacKerell, A. D., Jr., Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (Aaqa)3 Peptide. *Biophys J* **2014**, *107* (4), 991-7.
22. Shalongo, W.; Dugad, L.; Stellwagen, E., Distribution of Helicity within the Model Peptide Acetyl(Aaqa)3amide. *J Am Chem Soc* **1994**, *116*, 8288-8293.
23. Best, R. B.; Zheng, W.; Mittal, J., Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput* **2014**, *10* (11), 5113-5124.
24. Sasmal, S. Understanding Structure and Kinetics of A β Monomer and Fibril Ensembles Using Molecular Simulations. University of California, Berkeley, 2017.
25. Best, R. B.; Hummer, G., Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J Phys Chem B* **2009**, *113* (26), 9004-15.
26. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W., Charmm-Gui Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys J* **2009**, *97* (1), 50-8.
27. Huang, J.; MacKerell, A. D., Jr., Charmm36 All-Atom Additive Protein Force Field: Validation Based on Comparison to Nmr Data. *J Comput Chem* **2013**, *34* (25), 2135-45.
28. Boonstra, S.; Onck, P. R.; Giessen, E., Charmm Tip3p Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J Phys Chem B* **2016**, *120* (15), 3692-8.
29. Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J Phys Chem B* **2003**, *107* (24), 5933-5947.
30. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., Openmm 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659.
31. Roe, D. R.; Cheatham, T. E., 3rd, Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9* (7), 3084-95.
32. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577-637.
33. Sasmal, S.; Lincoff, J.; Head-Gordon, T., Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-Beta. *Biophys J* **2017**, *113* (5), 1002-1011.
34. Albaugh, A.; Demerdash, O.; Head-Gordon, T., An Efficient and Stable Hybrid Extended Lagrangian/Self-Consistent Field Scheme for Solving Classical Mutual Induction. *J Chem Phys* **2015**, *143* (17), 174104.
35. Albaugh, A.; Head-Gordon, T., A New Method for Treating Drude Polarization in Classical Molecular Simulation. *J Chem Theory Comput* **2017**, *13* (11), 5207-5216.
36. Albaugh, A.; Head-Gordon, T.; Niklasson, A. M. N., Higher-Order Extended Lagrangian Born-Oppenheimer Molecular Dynamics for Classical Polarizable Models. *J Chem Theory Comput* **2018**, *14* (2), 499-511.
37. Albaugh, A.; Niklasson, A. M. N.; Head-Gordon, T., Accurate Classical Polarization Solution with No Self-Consistent Field Iterations. *J Phys Chem Lett* **2017**, *8* (8), 1714-1723.

38. Albaugh, A.; Tuckerman, M. E.; Head-Gordon, T., Combining Iteration-Free Polarization with Large Time Step Stochastic-Isokinetic Integration. *J Chem Theory Comput* **2019**, *15* (4), 2195-2205.
39. Vitale, V.; Dzedzic, J.; Albaugh, A.; Niklasson, A. M.; Head-Gordon, T.; Skylaris, C. K., Performance of Extended Lagrangian Schemes for Molecular Dynamics Simulations with Classical Polarizable Force Fields and Density Functional Theory. *J Chem Phys* **2017**, *146* (12), 124115.

4.6 Appendix

Supplementary Table 1: Experimentally derived % α -helix from chemical shifts across the range of temperatures.

Temperature (K)	% α -helix
280	42
300	21
320	8.5
340	3.5
360	0.96
380	0.78

Supplementary Table 2: Comparison of % α -helix as calculated using only the first 100 ns of simulation vs. the full simulation for fixed charge force fields. Used to very roughly estimate how the amount of helix content may change upon extending the shorter AMOEBA simulations.

Force Field	% helix	Temperature (K)					
		280	300	320	340	360	380
CHARMM36 + C-TIP3P	First 100 ns	7.0	7.5	5.5	11	9.0	8.3
	Full	4.2	4.2	4.0	4.2	4.8	5.2
CHARMM36m + C-TIP3P	First 100 ns	7.2	14	6.9	9.4	6.8	5.2
	Full	4.4	5.8	4.4	4.5	5.0	4.6
Amber ff99SB + TIP4P-Ew	First 100 ns	7.5	4.7	5.3	3.5	4.6	6.1
	Full	3.7	3.7	3.4	3.1	3.1	3.2
Amber ff99SB-ILDN + TIP4P-d	First 100 ns	3.4	12	7.2	6.9	3.5	3.4
	Full	4.2	3.9	3.3	3.6	3.8	2.6

Chapter 5

A Bayesian Method for Validating Conformational Ensembles with Experimental Data: Application to Ensemble Refinement for Intrinsically Disordered Proteins

Given the difficulties that remain in guaranteeing convergence and force field accuracy for simulation of intrinsically disordered proteins (IDPs), validation of results by comparison to experimental data is required. Back-calculation methods for computing the experimental signal expected from a set of simulated structures, however, can be imprecise to the point that structurally distinct ensembles produce nearly identical back-calculated experimental signals. Because of this, standard comparison metrics such as squared errors that do not capture the effects of uncertainty in back-calculation and experiment are not always able to distinguish between even very different ensembles of IDP conformations. Here we present the development of improved scoring methods, based on the experimental inferential structure determination (EISD) formalism previously developed in the group for J couplings and chemical shifts, for a wide range of observables from nuclear magnetic resonance (NMR) and Förster resonance energy transfer (FRET) experiments. The EISD method uses a Bayesian analysis to calculate log-likelihood scores of a simulated ensemble corresponding to set of experimental data of various types, accounting for varying levels of uncertainty in back-calculation and experiment for each. We test the ability of each scoring metric to refine a structural ensemble for the disordered state of the drkN Sh3 domain, demonstrating the applicability of EISD in characterizing and refining IDP ensembles.

5.1 Introduction

The utility of a simulated IDP conformational ensemble^{1, 2}—the degree to which it can be trusted to faithfully include relevant conformations—is judged on the extent to which it can be shown to correctly model the structural information available from experimental data. The most straightforward approach to validating an IDP structural ensemble is to back-calculate the experimental data from the set of simulated conformers, and then quantify the overall degree of agreement using standard global comparisons such as squared errors, figures of merit, and correlation coefficients.

Such standard comparisons suffer from a lack of ability to account for uncertainty and systematic error in the underlying experiments and the possible inaccuracies of the back-calculation methods. As an example, chemical shifts from a nuclear magnetic resonance (NMR) experiment can be measured to very high accuracy resolved to differences of hundredths of ppm between related or mutated peptides³. Back-calculation of chemical shifts^{4, 5} from a set simulated structures, however, is less accurate by orders of magnitude, to the point of completely obscuring the experimentally determined differences (or lack thereof) between the peptides. Thus, while chemical shifts are an extremely sensitive experimental reporter of IDP structure, their utility for comparing simulated ensembles is limited in ways unaccounted for by many standard evaluation metrics.

A more relevant scoring method that does account for sources of uncertainty is furthermore important given the need to compare against multiple types of experimental observables. In contrast to folded proteins, the structural characterization of IDPs is highly underdetermined², and structurally distinct ensembles can be produced that agree with the same experimental signals. Comparison to multiple types of experimental data that report on different aspects of structure, such as local secondary structure vs. overall chain expansion, or that have different signal responses to structural characteristics, is therefore desired⁶. Given the potential differences in the level of uncertainties in both experiment and back-calculation, there is a need to be able to weight the relative importance of these comparisons. An ideal scoring method should therefore have at least the following capabilities:

1. Compare against multiple data types, with a scoring metric that accounts for uncertainties in both back-calculation and experiment.
2. Handle these comparisons simultaneously, to be able to account for the different structural information they provide.
3. In the event of disagreement between experimental restraints, weight the overall score more by the experiment and back-calculation methods with lower uncertainty.

Our group has previously developed such a method, experimental inferential structure determination (EISD)⁷, and applied it to chemical shifts and J couplings. It uses a maximum likelihood estimator formalism to assign a log likelihood score of a simulated ensemble matching an input set of experimental data, accounting for the uncertainties in both back-calculation and experiment by optimizing over the set of “nuisance parameters” ξ that are involved, which are treated as Gaussian random variables. EISD can be applied to multiple data types simultaneously to generate an aggregated probabilistic score of the form

$$\ln p(X, \xi | D, I) = \ln p(X | I) + \sum_{j=1}^M \ln [p(d_j | X, \xi_j, I) p(\xi_j | I)] \quad (1)$$

where $\ln p(X, \xi | D, I)$ is the log likelihood that the ensemble of N conformations $X = \{x_i\}_{i=1}^N$ and values of ξ correspond to the set of M experimental values $D = \{d_j\}_{j=1}^M$ and information I . The structural prior $p(X | I)$ can be treated as either a uninformative prior or a Boltzmann prior to incorporate an energetic component to the scoring, as explored in the original EISD publication⁷. Here, we use a uniform structural prior. The second component consists of $p(d_j | X, \xi_j, I)$ which models the experimental data given a set of conformers where the prior distribution $p(\xi_j | I)$ is the set of nuisance parameters ξ_j for a restraint j is modeled as a set of Gaussian random variables for an experimental or back-calculation uncertainty.

When tested on globular protein structures to distinguish the native state from misfolded decoys, EISD is able to clearly distinguish the correct native state by evaluating the degree of agreement with chemical shifts and J couplings. Furthermore, the formulation of the EISD scoring metric has $O(N)$ scaling with the number of conformations, allowing for much more rapid scoring of large ensembles than other methods that similarly incorporate a Bayesian approach⁸⁻¹¹. More importantly, EISD was able to show that the chemical shifts

and scalar coupling data types could not distinguish between qualitatively different IDP ensembles – for example the difference between a protein in a good vs. a bad solvent – because of the large back-calculation errors for chemical shifts. An important conclusion of the study is that other experimental data types and improved accuracy of back-calculation is needed to better validate and differentiate between IDP structural ensembles.

Here we extend the current use of EISD Bayesian model by developing scoring formalisms for additional experimental data types: homonuclear Nuclear Overhauser Effects (NOEs)^{6, 12, 13}, paramagnetic relaxation enhancements (PREs)^{14, 15}, residual dipolar couplings (RDCs)^{16, 17}, and hydrodynamic radii (R_h) from NMR¹⁸; and ensemble average efficiencies from Förster resonance energy transfer (FRET) experiments^{19, 20}.

Given the previous implementations of EISD for J couplings and chemical shifts⁷, we have formulated a full implementation of a simple black box-style EISD scoring approach for each new data type, demonstrating the ability of EISD to distinguish between the quality of conformational ensembles against an even wider set of experimental data. We then describe results from the expanded EISD scoring of the individual experimental restraint types, and validate the ability of the newly developed algorithms to refine a 1,700-conformation structural ensemble of the intrinsically disordered drkN Sh3 peptide²¹ from the Protein Ensemble Databank (PED)^{22, 23} using experimental data from the PED entry and the literature^{18, 24}. We close this chapter with tests to characterize the relative abilities of the different experimental modules to refine the structural ensembles, and discuss future directions for the EISD Bayesian model.

5.2 Materials and Methods

Experimental Inferential Structure Determination

The J coupling and chemical shift restraints for which EISD scoring algorithms have been developed illustrate two general ways to formulate the scoring method for any experimental observable. For J couplings, we explicitly optimize over each variable parameter involved in back-calculation, whereas for chemical shifts we treat the back-calculator as a black box and instead optimize over a generic error parameter.

For J couplings, in particular the $J_{HN-H\alpha}$ that reports on the ensemble-averaged dihedral angle ϕ , the empirically fitted Karplus equation is used for back-calculation:

$$J_{HN-H\alpha} = A(\cos(\phi - 60))^2 + B \cos(\phi - 60) + C \quad (2)$$

For every individual $J_{HN-H\alpha}$ included as a restraint, the EISD score $f(A, B, C)$ will incorporate the log likelihood of the optimal instances of the back-calculation nuisance parameters A, B , and C and an experimental error nuisance parameter ϵ_{ex} :

$$f(A, B, C) = \ln p(A; \mu_A, \sigma_A) + \ln p(B; \mu_B, \sigma_B) + \ln p(C; \mu_C, \sigma_C) + \ln p(\epsilon_{ex}(A, B, C); 0, \sigma_{\epsilon_{ex}}) \quad (3)$$

where A, B , and C are treated as Gaussian random variables for which the mean values and standard deviations are drawn from the work of Vuister and Bax²⁵. The experimental error

$$\epsilon_{ex}(A, B, C) = D - A(\cos(\phi - 60))^2 - B \cos(\phi - 60) - C \quad (4)$$

is the difference between the experimental value, D , and the back-calculated value, using the optimal instances of A , B , and C , and is treated also as a Gaussian random variable drawn from a distribution with mean 0 and standard deviation $\sigma_{\epsilon_{ex}}$ equal to the reported experimental uncertainty for the measurement. Applying the maximum likelihood estimator approach, the derivatives of Equation (3) are taken with respect to A , B , and C and each set to equal 0. The resulting set of linear equations is solved to find the optimal values of A , B , and C given the test structural ensemble, experimental value, and experimental uncertainty, which are then used in Equation (3) to calculate the score for that restraint. For a set of experimental J couplings, the total score is the sum of the score for each individual restraint. In this way, EISD incorporates both the uncertainty in the back-calculation—by weighting the score by the variances in the values of A , B , and C —and in experiment for the scoring.

The approach for chemical shifts is different because common back-calculators such as SHIFTX2⁵ and SPARTA+⁴ incorporate their own internal logic and weighting for the different components used to back-calculate chemical shifts, to the extent of precluding a simple and differentiable mathematical form. Instead, the back-calculator is treated using a black-box model for a score $f(q)$ that optimizes over a single error parameter for back-calculation q

$$f(q) = \ln p(q; 0, \sigma_q) + \ln p(\epsilon_{ex}(q); 0, \sigma_{\epsilon_{ex}}) \quad (5)$$

where q is treated as a Gaussian random variable with mean 0 and standard deviation σ_q equal to the published RMSD for Shiftx2⁵ for the relevant atom type, and the experimental nuisance parameter

$$\epsilon_{ex}(q) = D - q - \frac{1}{N} \sum_{i=1}^N o_i \quad (6)$$

is the difference between the experimental shift value D and the average of the back-calculated shifts o_i from each structure, accounting for q . Applying the maximum likelihood approach here produces a single equation to calculate the optimal value of q given the back-calculator uncertainty, experimental uncertainty, and experimental shift value.

Though it is less nuanced than the approach for J couplings, this black box-style approach still incorporates effects of variation in the experimental and back-calculation uncertainties, and is the most appropriate form for scoring when the back-calculation algorithm does not have a readily differentiable mathematical form.

New EISD Scoring Approaches for Additional Experimental Data Types

As a proof of concept of the ability of EISD to treat other data types, we start by developing similar black box scoring approaches as has been done for chemical shifts, optimizing over a generic error parameter.

Nuclear Overhauser Effects (NOEs) Homonuclear ¹H-¹H NOEs report on the distance between two spins^{6, 8, 12}, with the intensity value of the NOE cross-peak scaling with the distance to the negative sixth power and signals generally being obtainable up to time- and

ensemble-averaged distances of 5 Å. Characterization of NOEs for IDPs is more complex than for folded proteins due to the decreased ability to precisely assign peak values to specific nuclei²⁶. Other complications include the potential for signals that arise from oligomerization or aggregation, and the increased likelihood and role of chain dynamics in spin relaxation¹³. Most standard NMR spectroscopy analysis packages²⁷⁻²⁹ convert NOE intensities to distance restraints of varying tightness or strength between the pair of atoms, or possible pairs of atoms if the peak assignment is ambiguous.

Back-calculation of NOEs from simulation can be done to varying degrees of rigor, depending on whether or not dynamical information is available and incorporated. When the conformational ensemble is derived from MD, it is possible to fully incorporate the dynamical effects on NOEs with further short simulations from each conformer to characterize the relevant correlation times of each^{6, 12, 13}. These in turn are used to calculate per-conformer estimates of the spectral density functions, allowing fairly precise back-calculation of both homonuclear ¹H-¹H NOEs, and other relaxation parameters such as heteronuclear ¹H-¹⁵N NOEs and relaxation times.

When using only static structures, or structures generated with statistical coil models and other such techniques where no dynamical information is available, direct back-calculation is less rigorous. The key parameter that varies per structure for homonuclear NOEs, the hydrogen-hydrogen distance, is sufficient to estimate scaled, ensemble-averaged values of the peak intensity, albeit lacking potential variation across the length of the chain that might result from differences in dynamics. A parallel, simple back-calculation for ¹H-¹⁵N NOEs and other relaxation parameters is far less clear to define, as the analogous key distance involved, the amide-proton vector, does not vary. Rather, differences in the values of these relaxation parameters are entirely based on differences in dynamical behavior. As such, a back-calculation that could model differences between individual conformations in an ensemble, or small sets of conformations within an ensemble, would require a way to estimate differences in either per-residue or per-structure time correlation functions based off of individual static structures, or perhaps from small sets of structures by attempting to estimate correlation or transition times between structures.

We therefore focus on homonuclear NOEs, as reasonable back-calculation is possible for these regardless of the generation method of the conformational pool. The intensity of NOE cross-peaks is proportional to the distance to the inverse sixth power between the two nuclei⁸, and could therefore be the target signal for back-calculation. However, much NOE data is reported as distance restraints³⁰, after having been processed with programs such as NMRPipe²⁷, XPLOR-NIH²⁹, or CYANA²⁸. Furthermore, in many cases distance restraints are further binned into classes – e. g. strong restraints of < 2.5 Å, medium restraints < 4 Å, and weak restraints < 5 Å – rather than reporting values that are directly proportional to the intensity.

Given this common filtering into distance restraints, we take the same approach to back-calculation as ENSEMBLE^{21, 31-33}, calculating the ensemble-averaged distance D^{calc} from the set of N structures

$$D^{calc} = \left(\frac{\sum_{i=1}^N d_i^{-6}}{N} \right)^{-1/6} \quad (7)$$

incorporating inverse sixth power ensemble averaging over the per-structure distances d_i . We assign the target experimental value D to be the midpoint of the range specified by the restraint, and define the variance around the target distance $\sigma_{\epsilon_{ex}}$ as a fraction of the range, which we explore in the results section.

For every distance restraint, the EISD score is calculated using Equation (5) with the experimental error distribution set as defined above, and the experimental nuisance parameter ϵ_{ex} calculated as

$$\epsilon_{ex}(q) = D - q - D^{calc} \quad (8)$$

We obtain rather large estimates of the experimental error, such that the estimated distribution around the restraint value is fairly flat. As a result of this, since our simple back-calculation is effectively just a comparison of ensemble-averaged simulation distances to processed distance restraints, we set the back-calculation error to be very small, 0.001 Å.

Paramagnetic Relaxation Enhancements (PREs) Similar to NOEs, PREs report on ensemble- and time-averaged distances to the inverse sixth power, with strong dynamical contributions^{16, 34}. To conduct PRE experiments, a paramagnetic center must be introduced to the protein, such as through covalent bonding of a spin label, commonly MTSL for IDPs. The experiment then reports differences in the relaxation rates when the paramagnetic center is present versus absent, which are converted to estimates of distances between the paramagnetic center and, most commonly, the amide protons of each residue. PRE signals can distinguish between a much larger range of distances than NOEs, roughly 10 – 25 Å. Above and below these distances there is no longer a strong change in signal¹⁴. As with NOEs again, PREs are often converted to distance restraints in a very coarse manner based on binning of intensity ratios, with generic long distance restraints constrained to a range of 25 – 100 Å, short distance restraints to less than 10 Å, and a set of medium-range distance restraints, where the signal response is strongest with respect to distance, covering the middle³⁰. Multiple constructs with the tag at different locations on the peptide may be used to provide several sets of restraints.

One potential issue with PREs is whether the chemical modification of system induces different dynamics, or alters structural population. While folded proteins are fairly resistant to perturbation from this kind of structural modification, we have found that in simulation, tagged IDPs can sample and favor non-native conformations and contacts¹⁵, such that a set of experimentally derived PRE restraints might possibly provide conflicting data from other experiments that do not so perturb the structural ensemble. Careful selection of the tag and its location should be used to minimize this potential for error.

With PREs, the possibilities for back-calculation and treatment are the same as for NOEs. The simplest approach, which we use as a proof-of-concept here, is to compare ensemble-averaged distances from simulation ensembles to converted distance restraints from PRE intensities, setting the mean value of the restraint to the center of the allowed range, and basing the experimental uncertainty on the restraint range. We therefore use the same EISD scoring approach as well, comparing ensemble-averaged distances o_i with experimental distance restraint values D , calculating the error between the two with Equation (8) and the corresponding EISD score with Equation (5).

Residual Dipolar Couplings (RDCs) Dipolar couplings between pairs of spins can provide useful signals for predicting local structure, upon inducing partial alignment of molecules in solution with magnetic field^{16, 17}. For IDPs, RDCs resulting from the alignment of the amide in the peptide bond are the most commonly measured and reported. RDCs are noted to be sensitive to formation of secondary structure, and useful for structure refinement in combination with chemical shifts.

Back-calculation of RDCs uses calculation of the alignment tensors of the static structure, which can be done globally for the entire peptide as in PALES³⁵, or locally using fragments of the peptide as in the local RDC calculator from Forman-Kay¹⁷. Local back-calculation of RDCs has been shown to be able to better model experimental RDCs when using smaller ensembles of structures.

Here, we use the local RDC back-calculator, with standard values of parameters, to get per-conformation RDCs o_i for the amide bond vector of each residue. For EISD scoring, we take the simple ensemble average of these RDCs, calculating the experimental nuisance parameter as with chemical shifts using equation (6) and the EISD score with Equation (5). We estimate the uncertainty in back-calculation error σ_q based on the performance of the test set of peptides in the local RDC publication¹⁷, settling on a value of 0.88 Hz. Our estimation is further discussed in the Appendix, summarized in Table S1.

Hydrodynamic radius (R_h) The hydrodynamic radius of a molecule is the radius of a sphere that would have the same diffusion coefficient as the molecule. It can be experimentally determined by calculating the translational diffusion coefficient of the molecule with a technique such as pulsed field gradient NMR¹⁸, and then using the Stokes-Einstein relationship to calculate an ensemble-averaged estimate of the R_h . Like the radius of gyration (R_g), the R_h is useful as a measure of the overall degree of expansion of the IDP, ranging from more globule-like collapsed structures to fully extended random coil.

We use the program HYDROPRO³⁶ to calculate R_h , which takes static structures and uses a bead-shell model to estimate hydrodynamic properties. HYDROPRO is estimated to typically have 4 % error in the estimation of R_h .

For EISD scoring, we calculate the ensemble-averaged back-calculated R_h over the set of candidate structures, and set the experimental error based on the values published. We estimate the back-calculation error by taking the experimental R_h value multiplied by the 4 % error estimated by HYDROPRO.

Förster Resonance Energy Transfer (FRET) Efficiency FRET^{19, 20, 24} reports on long range distances between two covalently bound dyes, the pair chosen so that the fluorescence wavelength of one dye, the donor, matches the excitation wavelength of the other, the acceptor. The excited donor dye may, rather than fluorescing, transfer energy to the acceptor dye with a dipole-dipole coupling, such that the relative amounts of fluorescence from the two dyes can be related to the efficiency of energy transfer E between the two, which in turn is dependent on the distance r_{D-A} between the donor and acceptor:

$$E = (1 + (r_{D-A}/r_0)^6)^{-1} \quad (9)$$

where r_0 is the Förster radius of the donor-acceptor pair. In an ensemble experiment, the average E can be used to generate distance restraints for the two residues to which the

dyes are attached, with multiple experiments consisting of different FRET constructs—different pairs of dyes, or dyes linked to different residues on the protein—sometimes used to produce multiple restraints. Commonly used fluorophores can be as large or larger than bulky amino acid side chains, so there is a possibility that, depending on where the dye is attached, it might perturb the structure or dynamics of the system, as has been seen for PREs¹⁵.

Ensemble FRET efficiencies can be back-calculated by simply taking the appropriate distance measurements from static structures, calculating per-structure efficiencies, and then averaging together. Except for the case of studies that include explicit representation of the dyes during dynamics or conformer generation, a model is needed to account for the difference between the distance between the two residues to which dyes would be attached, and the distance between the dye centers themselves. A common approach¹⁹ is to treat the addition of the dyes as some number of additional residues, and use a simple polymer model to scale up the C α -C α distance of the native peptide:

$$r_{D-A} = r_{C\alpha-C\alpha} \left(\frac{N+N_{linker}}{N} \right)^{\nu} \quad (10)$$

where $r_{C\alpha-C\alpha}$ is the C α -C α distance, N is the number of residues between the relevant residues, N_{linker} is the number of estimated additional amino acids, and ν is the Flory scaling exponent.

For EISD scoring, we initially back-calculate the individual FRET efficiency of every conformation assuming mean values of the involved back-calculation parameters as discussed in the Results section. Estimation of the uncertainty σ_q in back-calculation is also discussed in Results. We use a typical estimate of the error in the experimental ensemble E value of 0.02 for $\sigma_{\epsilon_{ex}}$, and calculate the score as before with Equation (5).

Test Data for the drkN Sh3 Disordered Domain

We test the new EISD modules on the unfolded state ensemble of the drkN Sh3 peptide^{18, 24, 33}, chosen because of the wide variety of experimental data that is available for it, which has made it popular as a test system for other ensemble scoring and refinement programs, most notably ENSEMBLE. For structures, we use a 1,700-conformer ensemble deposited by the Forman-Kay group²¹ on the PED³⁷, consisting of an aggregated set of best-fit ensembles from multiple runs of the program ENSEMBLE³². The PED deposition includes a set of experimental data we also use, consisting of chemical shifts, J couplings, homonuclear NOE distance restraints, PRE distance restraints, and RDCs. We additionally use experimental R_h ¹⁸ and ensemble FRET efficiency²⁴ from other work by Forman-Kay and collaborators. In all cases we use the full set of restraints available, with the exception of discarding the small number of NOEs for which there is ambiguity in the pairs of residues involved.

Sub-Ensemble Scoring and Optimization using EISD

To characterize the general agreement between the ensemble and the experimental data, we generate 1,000 random sub-ensembles of sizes $N = 2, 5, 10, 25, 50,$ and 100 , demonstrating the typical scoring behavior for each data type, as well as the overall expected convergence to improved scores and lower disagreement with experiment as the ensemble size increases. We allow the same conformation to be selected for any number of

times in any ensemble, to reflect the potential different weighting or sampling of different conformational states.

To provide a comparison against the EISD score for any data type, we also calculate a restraint RMSD

$$RMSD = \left\langle \sqrt{\frac{\sum_{i=1}^M (D_i^{calc} - D_i^{exp})^2}{M}} \right\rangle \quad (11)$$

where for any data type, we take the set of M restraints with experimental values D_i^{exp} and compare them to the ensemble-averaged back-calculated values D_i^{calc} . The exterior brackets reflect averaging over the repeated 1,000 ensembles. There is only one restraint each for two of the experimental data types, R_h and FRET efficiency E , so we will generally refer to an absolute deviation from the restraint for these two data types rather than an RMSD.

To characterize the utility of an EISD score as a target value for optimization, we attempt to improve the score of the 1,000 randomly sampled ensembles for ensemble size $N = 100$. We use a simple direct maximization, performing 10,000 exchange attempts to replace one member conformation with another from the total pool of 1,700, accepting an exchange if the new ensemble has a higher EISD score than the previous. We perform the optimization comparing against only a single experimental data type at a time, to allow for characterization of the different degrees of restraining power for each.

Given that the total pool of 1,700 conformers is derived from pre-refined ensembles²¹ against much of the experimental data used here, there is likely some limitation to the degree of improvement that is seen, relative to tests using a pool of truly random coil conformations. However, since the primary goal of this initial work is to demonstrate the applicability of EISD scoring to the various new experimental data types being tested, we are satisfied with working within this range.

5.3 Results

Full Ensemble Characterization

Having already been refined against this full set of experimental data, the 1,700-conformer ensemble in the PED demonstrates strong agreement with the full experimental dataset, as summarized in Table 1 per data type and Supplemental Figure 1 for each individual restraint.

The secondary structure content of the full ensemble is plotted in Figure 1. Secondary structure content was calculated using the implementation of the DSSP algorithm within the AmberTools program *cpptraj*³⁸. Results are not shown for β content or π helices, as the full ensemble contains only negligible amounts of these structures. Error bars represent \pm one standard deviation of per-residue structure propensities among a set of 1,000 randomly drawn sub-ensembles consisting of 100 conformers each.

The full ensemble is largely disordered, with the main structural feature being a peak in 3-10 helix propensity around 35 – 40 % in residues 16-20. The ensemble-averaged $\langle R_g \rangle$, also calculated with *cpptraj*³⁸, is 18.5 ± 0.2 Å, with the uncertainty calculated in the same way as for secondary structure propensities.

Table 1: Full ensemble restraint RMSDs (NOEs, PREs, RDCs, J couplings, and chemical shifts) or absolute deviations (R_h and FRET efficiency) from experimental data.

Experimental data type	Number of restraints	Restraint RMSD
NOEs (\AA)	93	2.21
PREs (\AA)	68	1.82
RDCs (Hz)	28	0.50
R_h (\AA)	1	0.30
FRET efficiency	1	0.07
J couplings (Hz)	47	0.31
Chemical shifts (ppm)	267	0.28

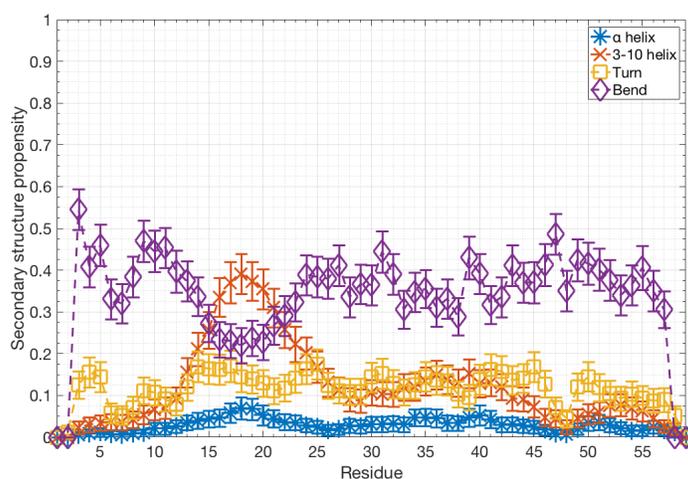


Figure 1: Secondary structure propensities per residue of the full ensemble. Error bars are shown as \pm one standard deviation for the secondary structure propensities of 1,000 random sampling ensembles of 100 conformers each.

Testing of New EISD Modules in Single-Mode Optimization

NOEs For NOEs, we set the target experimental value to be the mean distance of the allowed restraint range. Each experimental restraint for drkN Sh3 available restrains the distance between the pair of protons to less than 8 or 10 \AA ²¹. The optimal distance for each, then, will be 4 or 5 \AA . The remaining question is how to set $\sigma_{\epsilon_{ex}}$. Since the direct experimental measurement, the peak intensity, has already been thoroughly converted and scaled, we cannot simply use the \pm on the intensity value. Furthermore, the conversion from peak intensity to distance restraint incorporates much of what would typically be considered back-calculation uncertainty, as reflected in the fairly uniform and wide set of restraints.

We therefore test multiple values of $\sigma_{\epsilon_{ex}}$, based on dividing the restraint range by a series of integers. The resulting relative probabilities of an observed distance, normalized to the restraint value, are shown in Figure 2. As $\sigma_{\epsilon_{ex}}$ is further restricted, the model more

closely matches one intention of the restraint—to penalize observed distances that are outside of the restraint range. Smaller values of $\sigma_{\epsilon_{ex}}$, however, also result in a wide range of relative probabilities within the restraint range, and might result in too strong of a bias toward the middle of the range. Conversely, larger values of $\sigma_{\epsilon_{ex}}$ more closely match the expectation that all distances within the restraint range should be of roughly equal likelihood, but might not sufficiently penalize distances outside of the restraint range.

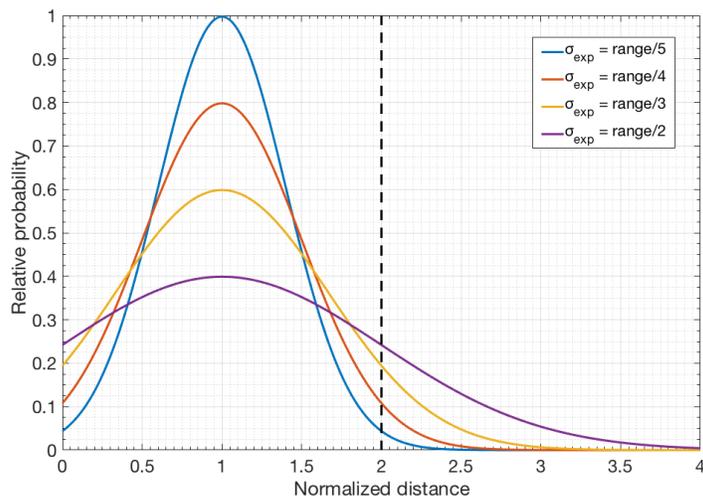


Figure 2: Variation in relative probabilities with different scaling of $\sigma_{\epsilon_{ex}}$ for NOEs. The x-axis is normalized the restraint value, 4 or 5 Å for this set of test data, such that the peak is always set to 1 in the plot, and the dashed black line represents the nominal upper bound of the restraint range.

This effect is manifested in the varying aggregate probabilities of randomly drawn ensembles of different sizes, shown in Figure 3 for $\sigma_{\epsilon_{ex}} = \text{range}/2$, and for all tested values of $\sigma_{\epsilon_{ex}}$ in Supplemental Figure 2. The value of the ensemble score increases and the range of scores among ensembles decreases as the number of conformations in the ensemble increases. This reflects and confirms the overall very good fit of the full ensemble to the experimental data, as also seen in the full ensemble fit in Table 1 and the RMSDs of random ensembles in Supplemental Figure 3. This very good overall fit results in RMSDs for randomized ensembles that, with enough conformations, are entirely within the nominal experimental restraint ranges, with restraint RMSDs less than 4 Å. We may be limited from seeing some of the possible negative effects of setting larger values of $\sigma_{\epsilon_{ex}}$ given this apparent lack of sets of conformations that significantly deviate from the restraints.

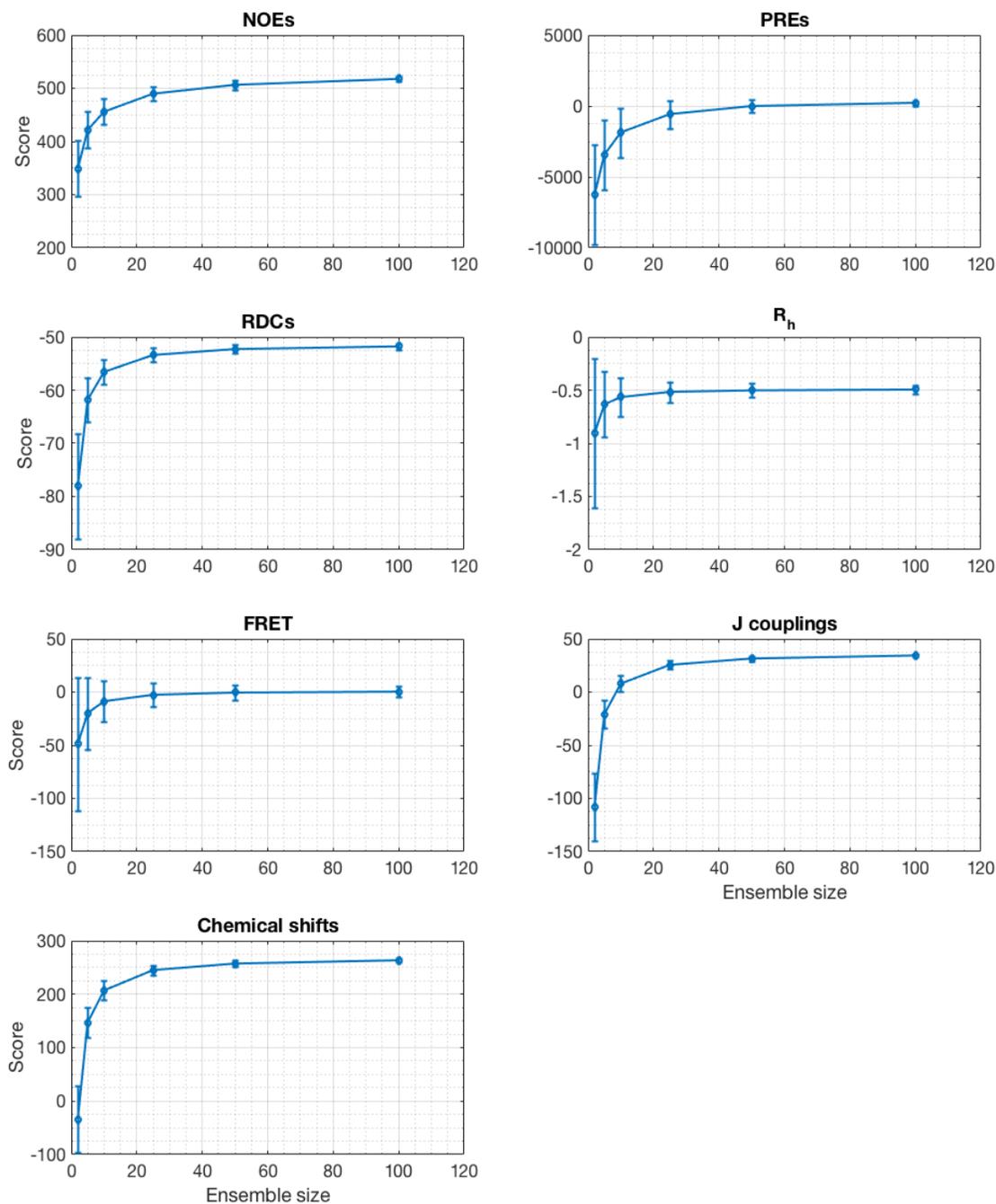


Figure 3: EISD scores for randomized ensembles of different sizes according to different experimental data types. Mean scores are calculated across 1,000 repeated random samplings of different ensembles, with error bars representing \pm one standard deviation across the 1,000 replicates.

This trend persists following optimization of the ensembles, with histograms of scores of optimized and randomized ensembles using $\sigma_{\epsilon_{ex}} = \text{range}/2$ in Figure 4, histograms of RMSDs using $\sigma_{\epsilon_{ex}} = \text{range}/2$ in Supplemental Figure 4, and a summary of the optimized scores and RMSDs in Table 2 for all tested values of $\sigma_{\epsilon_{ex}}$. In all cases there is a

clear separation between the scores and RMSDs of optimized ensembles versus randomized ensembles, reflecting consistent convergence toward minimized deviation and maximized scores. The overall tightness of the distributions of optimized scores and RMSDs reflects that maximization using the total EISD score is reproducibly capable of converging to similar results independent of the starting ensemble, even with relatively few total exchanges relative to the size of the total conformational pool.

There is no observed change in the restraint RMSDs after optimization based on the different tested values of $\sigma_{\epsilon_{ex}}$. This is likely due to the fairly extensive direct maximization procedure used relative to the overall mismatch of the full ensemble. An annealing procedure with a non-zero probability of accepting exchanges that slightly decrease the score depending on the temperature scale used would be expected to demonstrate some sensitivity to the sharpness of the $\sigma_{\epsilon_{ex}}$, and represents a potential future development in the EISD protocol.

Table 2: NOE-optimized ensemble scores and restraint RMSDs for varying values of $\sigma_{\epsilon_{ex}}$. Scores and RMSDs are averaged across 1,000 repeats for each case, with uncertainties estimated by calculating the standard deviation across the repeats in parentheses.

	$\sigma_{\epsilon_{ex}} = \text{range}/2$	$\sigma_{\epsilon_{ex}} = \text{range}/3$	$\sigma_{\epsilon_{ex}} = \text{range}/4$	$\sigma_{\epsilon_{ex}} = \text{range}/5$
Score	547.9 (0.11)	579.1 (0.25)	596.8 (0.46)	605.8 (0.68)
RMSD (Å)	1.41 (0.015)	1.41 (0.015)	1.41 (0.016)	1.41 (0.015)

PREs Though PREs are back-calculated and scored the same as NOEs in this case, we test only a single value of $\sigma_{\epsilon_{ex}}$ per restraint, dividing the experimental range by 4 to fit the experimental range to a 95 % confidence interval on the normal distribution. Scores of randomized ensembles of varying sizes are shown in Figure 3, with corresponding RMSDs in Supplemental Figure 3. We observe a similar trend as for NOEs, with scores converging to higher values and RMSDs converging to lower values as the ensemble size increases, again reflective of the prior restraining of the full ensemble. We do note the significantly larger range in score, resulting from the greater signal range of PREs generally and the far greater number of experimental restraints reporting on very distant pairs of residues, allowing for greater variation and error in ensemble averaged distances relative to the NOE dataset, which consists primarily of close contacts.

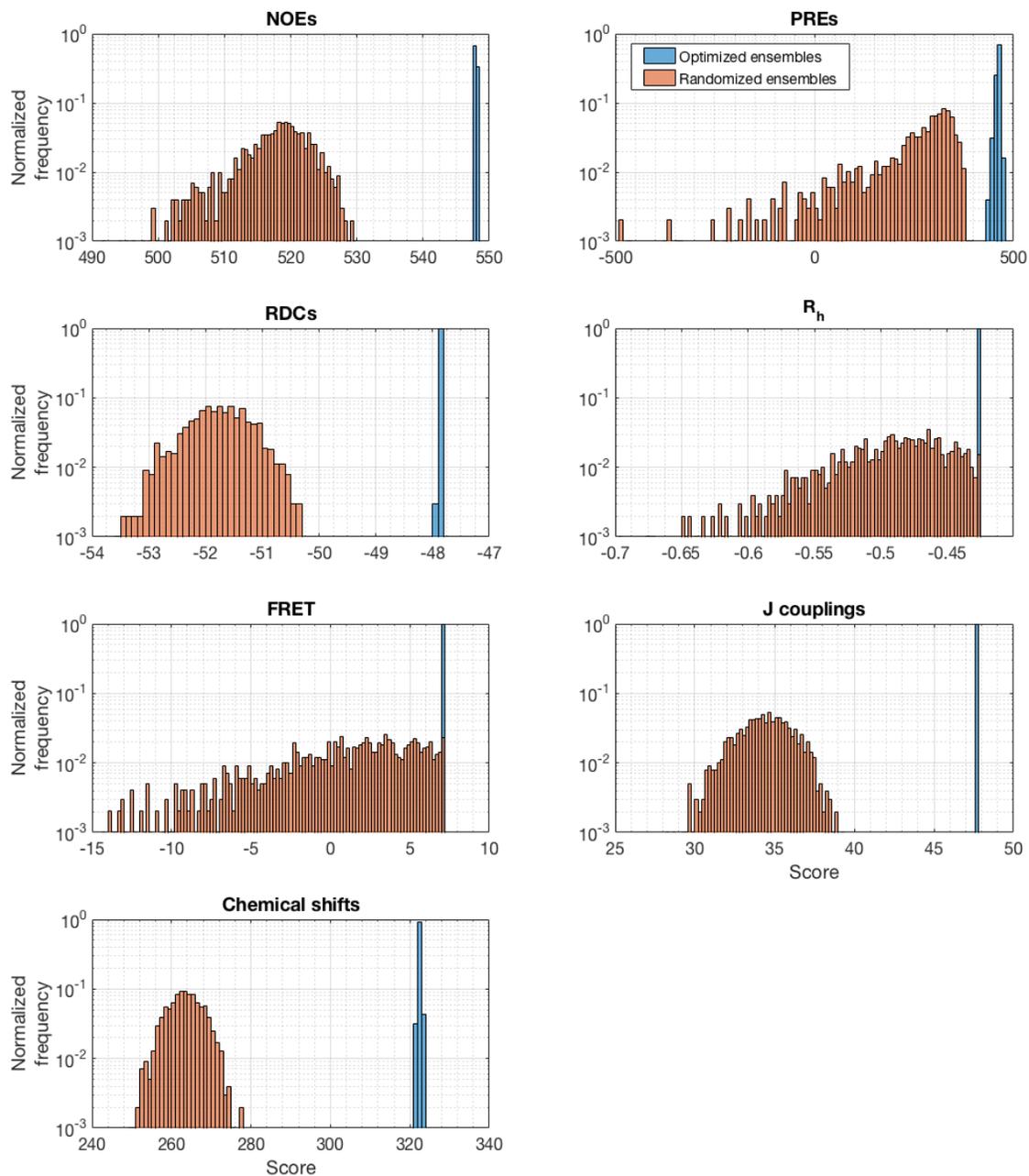


Figure 4: Histograms of scores for randomized and optimized ensembles of 100 conformers each, according to the experimental data type against which optimization and scoring are performed. Optimized ensembles are in blue, randomized ensembles are in orange. Each set represents 1,000 optimized and randomized ensembles.

Optimization against the PRE score again confirms the utility of this formulation of the EISD scoring metric by increasing the value and decreasing the variation in the score (Figure 4), and corresponding improvements in the RMSD (Supplemental Figure 4).

For both NOEs and PREs, we note that while we believe these results fully demonstrate the ability of EISD to characterize ensembles of differing quality, and its

appropriateness as a metric for ensemble comparison and optimization, the somewhat arbitrary nature of our assignment of very low values of the back-calculation error σ_q might result in larger shifts in the score for these two data types than for the others, to the point of potentially overstating the increase in true likelihood of a match between the final ensembles and the experimental data. Further testing against different systems, both different peptides and different initial pools of conformers, should be done to validate appropriate scaling of σ_q when implementing scoring against converted distance restraints. We believe these tests demonstrate that an EISD formalism for comparison to NOE and PRE data is clearly viable and useful, even if comparing to converted distance restraints rather than intensities, but that comparison to intensities would be better and more nuanced, and should be explored and implemented.

RDCs Similar performance is obtained for RDCs upon randomly sub-sampling from the full ensemble, with scores shown in Figure 3 and RMSDs in Supplemental Figure 3. As expected given the prior restraining of the full ensemble, and as demonstrated in the local RDC back-calculation paper, fairly low RMSDs are achieved with random samplings of 100 conformations. We note that the negative scores likely result from the fairly high values of $\sigma_{\epsilon_{ex}}$ (1.0 Hz) and σ_q (0.88 Hz) relative to the range of experimental values, at least compared to NOEs and PREs where, again, we have somewhat artificially set σ_q to be quite low. It is possible that the experimental errors are overestimated here, an artificial reduction of $\sigma_{\epsilon_{ex}}$ to 0.1 Hz results in scores of 7.8 ± 1.3 across 1,000 randomized ensembles of 100 conformers each. In any case, for the current set of data, the lower values of scores for the RDCs conveys that, despite the relatively good fit seen in the RMSDs, EISD assigns is lower confidence to the match between the conformational ensembles and experimental data, in a way that standard figures of merit are unable to capture.

Likely due to the relatively large values of the uncertainties, optimization of the ensembles does not lead to as large of a shift in the scores (Figure 4) despite a strong reduction in the RMSD (Supplementary Figure 4). This would again shift with lower values of $\sigma_{\epsilon_{ex}}$ and σ_q , but for now reflects that, if the uncertainties are large, even a near perfect match between back-calculated and experimental data does not provide much confidence that the simulated ensemble is necessarily correct. Alternatively, given that the conformational pool has been refined against an extensive set of experimental data, the relatively low score for RDCs suggests that this set of experimental data may contain slightly less useful information for structure refinement than the others, though we note that other work has demonstrated their utility in improving modeling of tertiary structure³².

R_h Fitting to R_h is simplified compared to the other previously discussed restraints as there is only one restraint being fit. As before, randomized ensemble scores (Figure 3) and deviation from the restraint value (Supplementary Figure 3) approach convergence as the ensemble size increases, appearing to converge at smaller sizes than for previous models due to the presence of only one restraint. The negative score values, while not truly significant, convey again that despite reasonable agreement between the ensembles and experiment, there is fairly significant uncertainty.

As a result, while optimization is able to generate ensembles that perfectly match the experimental value of the R_h with deviations on the order of 10^{-5} Å (Supplementary Figure 4), the scores remain relatively low (Figure 4). Furthermore, that there is only one

restraint to fit results in overall lower changes in the magnitude of the score upon optimization, as it is overall more likely that a random ensemble will converge to a similar level of agreement as an optimized ensemble for the single data point.

FRET Efficiency For FRET efficiency, we first require an estimate of σ_q . As with R_H , because there is only a single restraint we are able to generate ensembles that perfectly match the experimental efficiency, assuming mean values for the back-calculation and experimental nuisance parameters. To estimate σ_q , we recalculate what the FRET efficiency would be for such a set of perfect-fit, 100-conformation ensembles while varying these parameters within expected ranges, and assign σ_q to be equal to the standard deviation among the 10^6 resulting recalculated ensemble FRET efficiencies (see Supplementary Figure 5) = 0.0074.

Using this value of σ_q , we again calculate scores (Figure 3) and average deviations (Supplementary Figure 3) for randomly drawn ensembles of various sizes, finding similar behavior as for R_H , given the single restraint available for comparison, albeit with higher scores, likely due to the lower relative uncertainty values. These lower uncertainties likely drive the stronger shift in score upon optimization (Figure 4) that corresponds to a similar reduction in deviation from experiment (Supplementary Figure 4).

Comparisons Between Optimized Ensembles

We compare the different effects of the optimization according to each experimental data type by observing shifts in the secondary structure populations from the full ensemble, and changes in the variation in secondary structure among the 1,000 repeats of each optimization as shown in Figure 5. As expected given the prior refinement of the ensemble, none of the optimizations shift secondary structure populations entirely away from that of the full ensemble, there are merely small localized shifts.

Two of the modules, RH and FRET efficiency, produce secondary structure distributions and errors that are nearly identical to that of the full ensemble, as expected given that both consist of one fairly global restraint that would not be expected to be particularly sensitive to secondary structure. The greatest shift is seen for chemical shifts, where the very narrow error bars suggest that nearly identical sets of structures were selected for in final ensembles, indicative of the strong expected dependence of chemical shifts on local structure. Beyond that, chemical shift optimization produces reduced amounts of helical content in residues 1-10, slight enrichment of helical content in residues 15-20, and enrichment in the turn population at the C-terminus. The enrichment of α -helical content in residues 15-20 is a fairly common feature, with optimization by J couplings, NOEs, PREs, and RDCs also resulting in that shift. Optimization with RDCs and J couplings produce other secondary structure shifts as well, though the size of the error bars remains generally similar to that of the full ensemble, suggesting that—while they do shift certain structural populations—they are similarly permissive around those mean values. Contrastingly, NOE refinement results in lower variance in secondary structure compared to random sampling of the full ensemble, perhaps due to the preponderance of relatively short range $\{i, i + (3, 4, 5)\}$ restraints in the NOE dataset. Optimization with PREs produces in some places larger error bars than the random sampling, suggesting that some of the optimized ensembles are quite distinct from each other. This corresponds generally to the somewhat larger range in final RMSD values for the PRE-optimized ensembles,

perhaps as a result of greater difficulty of converging within the attempted number of exchanges.

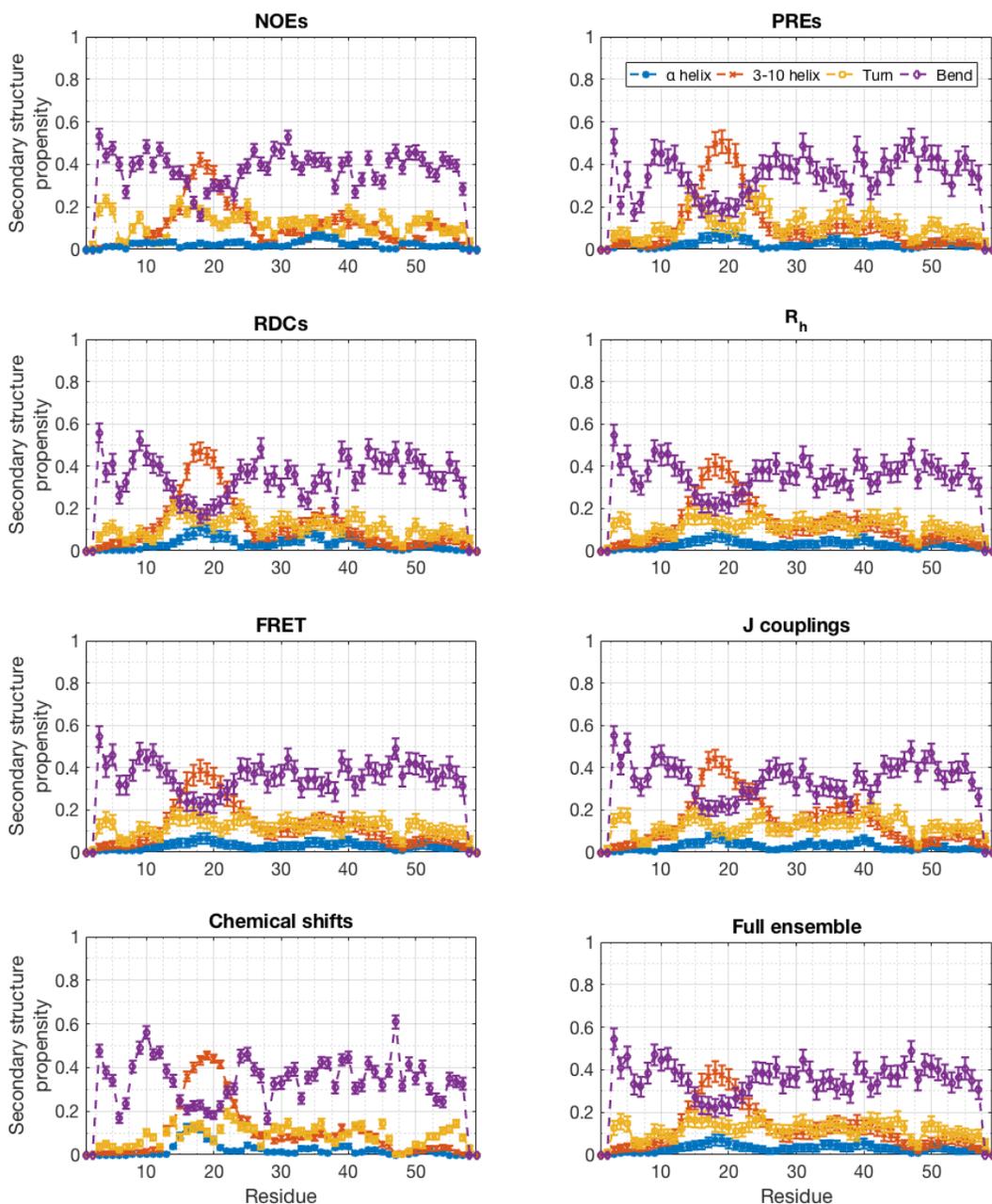


Figure 5: Secondary structure propensities of optimized ensembles. Mean secondary structure propensities per residue from 1,000 optimized ensembles for each experimental data type. Error bars are \pm one standard deviation from the 1,000 replicates. Propensities for α helix are in blue, 3-10 helix in orange, turn in gold, and bend in purple. Propensities for β content and π helices are not shown due to negligible sampling. Full ensemble secondary structure propensities are reproduced at bottom right for convenience of comparison, with error bars estimated from 1,000 random samplings of 100 conformers each.

Ensemble-averaged values of the R_g , a global measure of the degree of extension of the peptide, are presented in Table 3. Again there are no huge shifts in R_g , doubly expected given that the full ensemble was additionally refined against SAXS data, but the minor perturbations in both mean values and variances point to differences in preferences upon optimizing to the near-limit for each data type. Here again optimization with the R_h produces no shift in the mean value, but has lower variance than random sampling and optimization with each other experimental restraint, suggesting that R_h is powerfully specific for overall extension. Refinement with NOEs produces the largest shift, potentially suggesting that the optimization performed was too strict, and might be better tempered by incorporating some temperature scaling or Metropolis criterion. We may also here be overfitting to a set of restraints that do not sufficiently account for the effects of chain dynamics on NOEs, given the uniformity of the restraint values. Development of an EISD scoring formalism for NOEs that accounts for dynamical effects, as well as for other observables with similar dynamical dependence, is important for truly capturing the physics represented by the measurement, and is underway in the group. Interestingly again PRE refinement produces larger variance in R_g than random sampling, highlighting the potential, at least within these already strongly refined ensembles, for disparate final ensembles to satisfy the PRE data, or just much greater difficulty in optimizing to match the set of PRE restraints.

Table 3: Ensemble-averaged $\langle R_g \rangle$ values upon optimization with different experimental data types. Mean values represent an average across 1,000 repeated optimizations for each restraint type, and uncertainty values in parentheses represent the standard deviation across the repeated optimizations. Uncertainty for the full ensemble is estimated by calculating the standard deviation across 1,000 repeated random samplings.

Ensembles Restrained with:	$\langle R_g \rangle$ (Å)
NOEs	17.67 (0.16)
PREs	18.15 (0.35)
RDCs	19.18 (0.19)
R_h	18.52 (0.09)
FRET	18.17 (0.17)
J couplings	18.78 (0.19)
Chemical shifts	18.82 (0.13)
Full ensemble	18.50 (0.20)

Lastly, we analyze the agreement of each set of ensembles optimized with a single experimental data type with the all of the other experimental data, with scores presented in Table 4 and RMSDs presented in Supplementary Table 2. This first summarizes the effect that the number of restraints and the level of uncertainty has on the total value and variation of the score for an individual experimental data type. Data types for which there are more restraints are expected to have a greater magnitude of score. Data types with high

amounts of uncertainty will generally have lower scores even despite potentially having many restraints, as in the case of the RDCs, and will similarly be expected to exhibit lower variability in scores.

Table 4: *Cross-module scores of ensembles optimized against a single experimental data type.* Each row is a different ensemble generation type, starting with randomized ensembles and then below ensembles generated by optimizing against the single data type listed at the left. Each column is the score of that ensemble for the data type listed at the top, so that variations in the values within a column reflect the different effects of different optimization on the score of the data type for the column. Mean values are generated by averaging across the 1,000 repeated ensembles in every case, and uncertainties in parentheses are generated by calculating the standard deviation across the 1,000 ensembles. The rightmost column is the sum of the per-module scores for each ensemble.

		Scores:						Sum	
		NOE	PRE	RDC	FRET	R_h	JC		CS
Number of restraints:		93	68	28	1	1	47	267	
Random ensemble score:		517.7 (5.52)	225.2 (191.2)	-51.8 (0.70)	0.35 (5.09)	-0.37 (1.3e-2)	34.4 (1.75)	263.5 (4.37)	989.0
Ensembles optimized using:	NOE	548.0 (0.11)	292.2 (26.5)	-52.7 (0.38)	6.02 (1.22)	-0.44 (3.6e-2)	30.9 (1.52)	255.5 (2.89)	1,079.5
	PRE	518.1 (4.84)	462.3 (5.57)	-51.0 (0.59)	4.28 (4.34)	-0.39 (4.4e-2)	26.3 (2.71)	263.1 (6.22)	1,222.7
	RDC	511.2 (6.98)	-102.9 (456.1)	-47.9 (2.5e-3)	-8.93 (6.41)	-0.47 (5.4e-2)	31.6 (1.87)	269.8 (3.67)	652.4
	FRET	519.2 (4.52)	310.1 (57.4)	-51.9 (0.58)	7.16 (3.8e-10)	-0.39 (2.1e-2)	34.0 (1.90)	263.0 (4.46)	1081.2
	R_h	517.9 (5.32)	248.3 (127.7)	-51.8 (0.56)	0.74 (3.90)	-0.36 (1.4e-9)	34.5 (1.83)	263.5 (4.28)	1012.7
	JC	512.4 (6.17)	101.9 (261.8)	-51.9 (0.50)	-9.95 (7.18)	-0.37 (1.5e-2)	47.7 (1.2e-2)	248.0 (3.88)	847.8
	CS	506.6 (3.38)	58.2 (145.8)	-51.8 (0.29)	6.75 (0.57)	-0.45 (3.2e-2)	20.3 (1.26)	322.5 (0.28)	862.2

This testing also allows for evaluation of the potential effects that optimization with one data type has on the score for another data type. One such example where the effect is strong is the improvement in the PRE score, relative to randomized ensembles, upon optimizing with NOEs, with the mean PRE score shifting from 225.2 (random) to 292.2 (NOE-optimized). This suggests that there is sufficient shared or correlated information from the NOE restraints that they are able to produce better modeling of the PRE restraints. Interestingly, the inverse does not turn out to be true, with PRE optimization barely shifting the NOE score from the random result.

As expected given the application of a straightforward maximization procedure, the highest score for any individual experimental data type is achieved by optimizing with that

specific data. This is true by a large margin in all cases except for the data types that exhibit relatively low variability in scores at all, particularly the RDCs and R_h . This is potentially a result of the relatively high uncertainties assigned to these two data types relative to both the values of the experimental restraints, and relative to the overall disagreement between the full ensemble and the restraints.

Conversely, the PRE and FRET modules exhibit relatively huge variability in scores, both in the mean values upon optimization with different data types, and in the variance of random ensembles. Both exhibit strong decreases in scores relative to the random set upon optimization with RDCs and J couplings, perhaps representing disagreement between those more locally sensitive data types and the longer-range PRE and FRET experiments. The PRE and FRET modules interestingly show opposite responses upon optimization with chemical shifts, with the PRE score decreasing and the FRET score increasing to nearly the optimal FRET score achieved after optimization using FRET data. This could potentially point to some unseen disagreement between the PRE and FRET data about long-range behavior that is captured by chemical shifts but not the RDCs or J couplings.

In all cases, it is important to note that the changes in RMSD, summarized in Supplementary Table 2, are restricted to fairly low values given the pre-refinement of this full ensemble. We can generally expect that random ensembles drawn from a true random-coil ensemble would likely score much more poorly, which might result in cases such as the above producing similar responses relative to the random case rather than fully diverging. In another light, however, the range of scores possible, and the very different preferences of certain modules within that fairly tight range of conformations, is demonstrative of the sensitivity of EISD to subtle changes in the ensemble.

The rightmost column of Table 2 contains the summed scores for each ensemble type across the full set of experimental data. The PRE, NOE, and FRET-optimized ensembles have the highest scores, while the J coupling, chemical shift, and RDC-optimized have the lowest, with the R_h -optimized ensembles barely shifting the score from random sampling. This suggests that the available PRE, NOE, and FRET data might carry the most positive mutual information with the other modules, with the least amount of information that contradicts other modules. This is particularly noteworthy in the case of the FRET data, where optimization occurs against only one long-range pairwise distance. Almost the entirety of the improvement in the FRET module score comes from improvement in modeling of the PRE data however, and while that does reflect a likely agreement in the type of information the two modules provide and refine against, there is also certainly some affect of the potential over-sensitivity of the PRE module, given the low assigned σ_q .

5.4 Discussion and Conclusions

We have developed Bayesian scoring formalisms for comparing simulated ensembles to a variety of experimental data types, spanning those that report on very local to very global structure. The very good scaling of EISD calculations with ensemble sizes facilitates the high number of replicates we perform for each test, demonstrating the cost-effectiveness of the algorithm and providing high levels of confidence in the significance of the observed responses.

The current implementations presented here all follow a simple black-box approach in which a single generic error parameter q is optimized for every tested experimental restraint. This approach is able to account for varying levels of uncertainty in both experiment and back-calculation, and has been shown to be fully capable to serving as a characterization metric and target for optimization for ensembles of IDP conformations.

There do, however, remain significant areas for improvement in EISD scoring. For certain modules, such as NOEs and PREs, this implementation targets a forward-calculated distance restraint from experiment, and asserts near-zero variance in back-calculation uncertainty as a result. A scoring method that targets the actual peak intensity could likely identify differences between NOEs of different intensities, though that would require sufficiently well resolved experiments to be able to confidently assign and resolve different peaks, which remains difficult for many IDPs. This could furthermore aid in development of more confidently representative estimates of $\sigma_{\epsilon_{ex}}$ and σ_q . Further improvement could also likely be gained in implementation of scoring algorithms that, as has been developed for J couplings, explicitly model the math involved in back-calculation. Whenever it is possible, particularly for NOEs and other such observables where there is a strong dynamical effect, the back-calculation method and corresponding scoring should as well, either through calculation of correlation times from standard MD as has been done before⁶ or a hypothetical new method for estimation of correlation times from static structures that are not dynamically linked.

Despite these limitations of the current model, we believe these results demonstrate the appropriateness and efficiency of the EISD formulation for a wide variety of experimental data types. Even within a restricted pool of total conformers, optimization using EISD scores for a single experimental data type produces ensembles with distinct scores, aiding in the identification of conformational ensembles that best model the available experimental data. We can further identify the relative utility of different experimental comparisons in the context of current uncertainty in back-calculation and experiment for each, useful for learning which types of comparisons should be prioritized as well as identifying areas of need for more precise back-calculators.

5.5 References

1. Wright, P. E.; Dyson, H. J., Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat Rev Mol Cell Biol* **2015**, 16 (1), 18-29.
2. Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T., Finding Our Way in the Dark Proteome. *J Am Chem Soc* **2016**, 138 (31), 9730-42.
3. Conicella, A. E.; Fawzi, N. L., The C-Terminal Threonine of Abeta43 Nucleates Toxic Aggregation Via Structural and Dynamical Changes in Monomers and Protofibrils. *Biochemistry* **2014**, 53 (19), 3095-105.
4. Shen, Y.; Bax, A., Sparta+: A Modest Improvement in Empirical Nmr Chemical Shift Prediction by Means of an Artificial Neural Network. *J Biomol NMR* **2010**, 48 (1), 13-22.
5. Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S., Shiftx2: Significantly Improved Protein Chemical Shift Prediction. *J Biomol NMR* **2011**, 50 (1), 43-57.

6. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-Beta Peptides. *J Phys Chem B* **2014**, 118 (24), 6405-16.
7. Brookes, D. H.; Head-Gordon, T., Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. *J Am Chem Soc* **2016**, 138 (13), 4530-8.
8. Rieping, W.; Habeck, M.; Nilges, M., Inferential Structure Determination. *Science* **2005**, 309 (5732), 303-6.
9. Fisher, C. K.; Huang, A.; Stultz, C. M., Modeling Intrinsically Disordered Proteins with Bayesian Statistics. *J Am Chem Soc* **2010**, 132 (42), 14919-27.
10. Fisher, C. K.; Ullman, O.; Stultz, C. M., Comparative Studies of Disordered Proteins with Similar Sequences: Application to Abeta40 and Abeta42. *Biophys J* **2013**, 104 (7), 1546-55.
11. Fisher, C. K.; Stultz, C. M., Protein Structure Along the Order-Disorder Continuum. *J Am Chem Soc* **2011**, 133 (26), 10022-5.
12. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-Beta Peptides. *Biochemistry* **2011**, 50 (35), 7612-28.
13. Peter, C.; Daura, X.; van Gunsteren, W. F., Calculation of Nmr-Relaxation Parameters for Flexible Molecules from Molecular Dynamics Simulations. *J Biomol NMR* **2001**, 20 (4), 297-310.
14. Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M. R., Characterization of Intrinsically Disordered Proteins and Their Dynamic Complexes: From in Vitro to Cell-Like Environments. *Prog Nucl Magn Reson Spectrosc* **2018**, 109, 79-100.
15. Sasmal, S.; Lincoff, J.; Head-Gordon, T., Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-Beta. *Biophys J* **2017**, 113 (5), 1002-1011.
16. Newby, F. N.; De Simone, A.; Yagi-Utsumi, M.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M., Structure-Free Validation of Residual Dipolar Coupling and Paramagnetic Relaxation Enhancement Measurements of Disordered Proteins. *Biochemistry* **2015**, 54 (46), 6876-86.
17. Marsh, J. A.; Baker, J. M.; Tollinger, M.; Forman-Kay, J. D., Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment. *J Am Chem Soc* **2008**, 130 (25), 7804-5.
18. Choy, W. Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E., Distribution of Molecular Size within an Unfolded State Ensemble Using Small-Angle X-Ray Scattering and Pulse Field Gradient Nmr Techniques. *J Mol Biol* **2002**, 316 (1), 101-12.
19. Meng, F.; Bellaiche, M. M. J.; Kim, J. Y.; Zerze, G. H.; Best, R. B.; Chung, H. S., Highly Disordered Amyloid-Beta Monomer Probed by Single-Molecule Fret and Md Simulation. *Biophys J* **2018**, 114 (4), 870-884.
20. Song, J.; Gomes, G. N.; Shi, T.; Gradinaru, C. C.; Chan, H. S., Conformational Heterogeneity and Fret Data Interpretation for Dimensions of Unfolded Proteins. *Biophys J* **2017**, 113 (5), 1012-1024.
21. Marsh, J. A.; Forman-Kay, J. D., Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J Mol Biol* **2009**, 391 (2), 359-74.

22. Penkett, C. J.; van Ginkel, G.; Velankar, S.; Swaminathan, J.; Ulrich, E. L.; Mading, S.; Stevens, T. J.; Fogh, R. H.; Gutmanas, A.; Kleywegt, G. J.; Henrick, K.; Vranken, W. F., Straightforward and Complete Deposition of Nmr Data to the Pdbe. *J Biomol NMR* **2010**, *48* (2), 85-92.
23. Mir, S.; Alhroub, Y.; Anyango, S.; Armstrong, D. R.; Berrisford, J. M.; Clark, A. R.; Conroy, M. J.; Dana, J. M.; Deshpande, M.; Gupta, D.; Gutmanas, A.; Haslam, P.; Mak, L.; Mukhopadhyay, A.; Nadzirin, N.; Paysan-Lafosse, T.; Sehnal, D.; Sen, S.; Smart, O. S.; Varadi, M.; Kleywegt, G. J.; Velankar, S., Pdbe: Towards Reusable Data Delivery Infrastructure at Protein Data Bank in Europe. *Nucleic Acids Res* **2018**, *46* (D1), D486-D492.
24. Mazouchi, A.; Zhang, Z.; Bahram, A.; Gomes, G. N.; Lin, H.; Song, J.; Chan, H. S.; Forman-Kay, J. D.; Gradinaru, C. C., Conformations of a Metastable Sh3 Domain Characterized by Smfret and an Excluded-Volume Polymer Model. *Biophys J* **2016**, *110* (7), 1510-1522.
25. Vuister, G. W.; Delaglio, F.; Bax, A., The Use of 1jc Alpha H Alpha Coupling Constants as a Probe for Protein Backbone Conformation. *J Biomol NMR* **1993**, *3* (1), 67-80.
26. Novacek, J.; Zidek, L.; Sklenar, V., Toward Optimal-Resolution Nmr of Intrinsically Disordered Proteins. *J Magn Reson* **2014**, *241*, 41-52.
27. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., Nmrpipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J Biomol NMR* **1995**, *6* (3), 277-93.
28. Guntert, P., Automated Nmr Structure Calculation with Cyana. *Methods Mol Biol* **2004**, *278*, 353-78.
29. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M., The Xplor-Nih Nmr Molecular Structure Determination Package. *J Magn Reson* **2003**, *160* (1), 65-73.
30. Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L., Biomagresbank. *Nucleic Acids Res* **2008**, *36* (Database issue), D402-8.
31. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W. Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinformatics* **2013**, *29* (3), 398-9.
32. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins* **2012**, *80* (2), 556-72.
33. Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W. Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D., Improved Structural Characterizations of the Drkn Sh3 Domain Unfolded State Suggest a Compact Ensemble with Native-Like and Non-Native Structure. *J Mol Biol* **2007**, *367* (5), 1494-510.
34. Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M., Nmr Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J Am Chem Soc* **2010**, *132* (24), 8407-18.
35. Zweckstetter, M.; Bax, A., Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid to Structural Genomics. *J Am Chem Soc* **2001**, *123* (38), 9490-1.
36. Ortega, A.; Amoros, D.; Garcia de la Torre, J., Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys J* **2011**, *101* (4), 892-8.

37. Velankar, S.; Best, C.; Beuth, B.; Boutselakis, C. H.; Cobley, N.; Sousa Da Silva, A. W.; Dimitropoulos, D.; Golovin, A.; Hirshberg, M.; John, M.; Krissinel, E. B.; Newman, R.; Oldfield, T.; Pajon, A.; Penkett, C. J.; Pineda-Castillo, J.; Sahni, G.; Sen, S.; Slowley, R.; Suarez-Uruena, A.; Swaminathan, J.; van Ginkel, G.; Vranken, W. F.; Henrick, K.; Kleywegt, G. J., Pdbe: Protein Data Bank in Europe. *Nucleic Acids Res* **2010**, 38 (Database issue), D308-17.
38. Roe, D. R.; Cheatham, T. E., 3rd, Ptraaj and Cpptraaj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, 9 (7), 3084-95.
39. McCarney, E. R.; Werner, J. H.; Bernstein, S. L.; Ruczinski, I.; Makarov, D. E.; Goodwin, P. M.; Plaxco, K. W., Site-Specific Dimensions across a Highly Denatured Protein; a Single Molecule Study. *J Mol Biol* **2005**, 352 (3), 672-82.

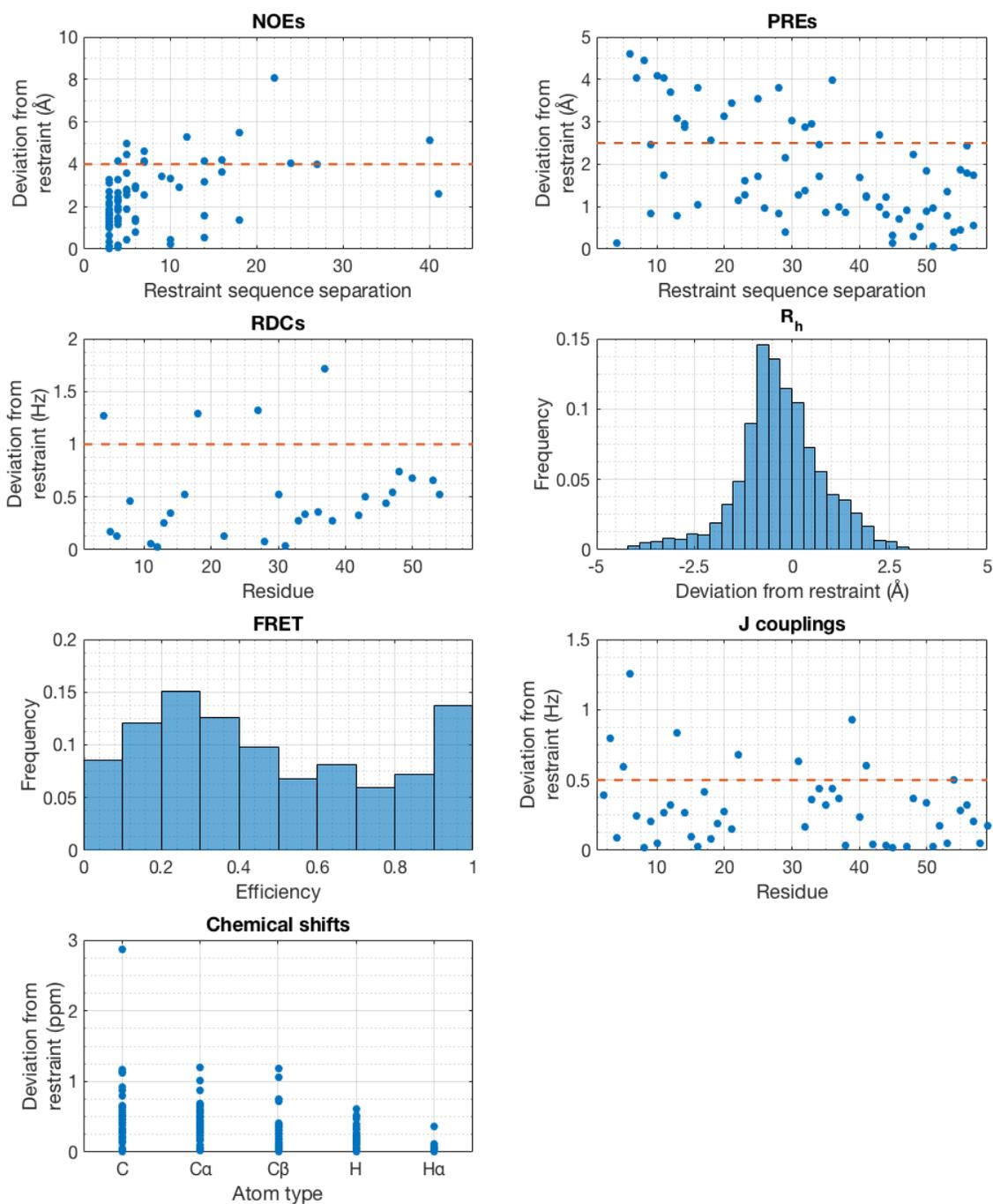
5.6 Appendix

Supplementary Table 1: *Normalized RMSDs of test systems for local RDC back-calculation for estimating back-calculator uncertainty of RDCs.* By digitizing the plots from the original work, the RMSD per peptide of the local RDC back-calculated results from large ensembles of simulated conformations against the experimental data is calculated, to develop an estimated mean RMSD normalized by the experimental range per peptide = 0.191, which is scaled back up by the experimental signal range for drkN Sh3 to estimate the value of $\sigma_q = 0.88$ Hz used in this work.

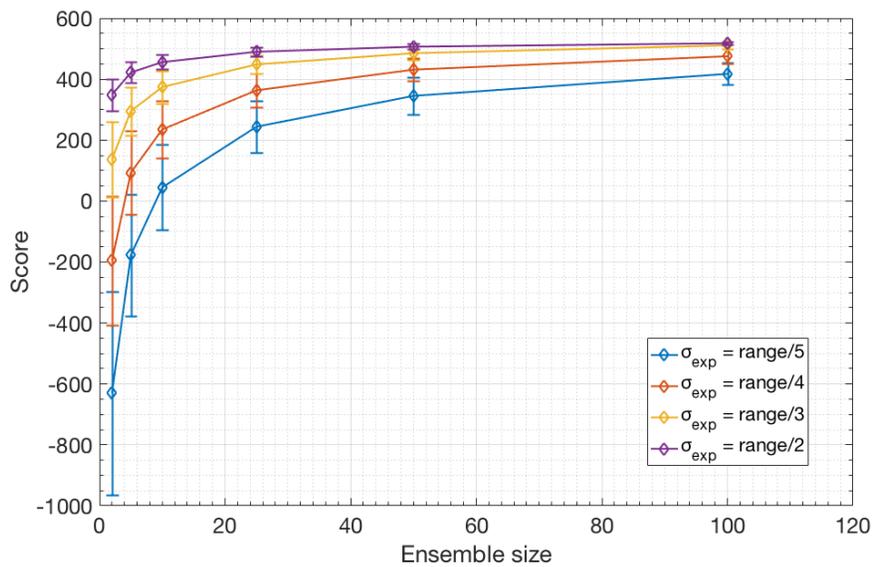
System	# RDCs	Exp. range (Hz)	RMSD normalized by exp. range
Drk	28	4.61	0.2263
ACBP	76	21.43	0.1879
CFTR	129	17.54	0.1666
StaphN	38	10.51	0.1822

Supplementary Table 2: *Cross-module RMSDs of ensembles optimized against a single experimental data type.* Each row is a different ensemble generation type, starting with randomized ensembles and then below ensembles generated by optimizing against the single data type listed at the left. Each column is the RMSD (NOE, PRE, RDC, J couplings, chemical shifts) or absolute deviation (R_h , FRET) of that ensemble for the data type listed at the top, so that variations in the values within a column reflect the different effects of different optimization on the RMSD or deviation of the data type for the column. Mean values are generated by averaging across the 1,000 repeated ensembles in every case, and uncertainties in parentheses are generated by calculating the standard deviation across the 1,000 ensembles.

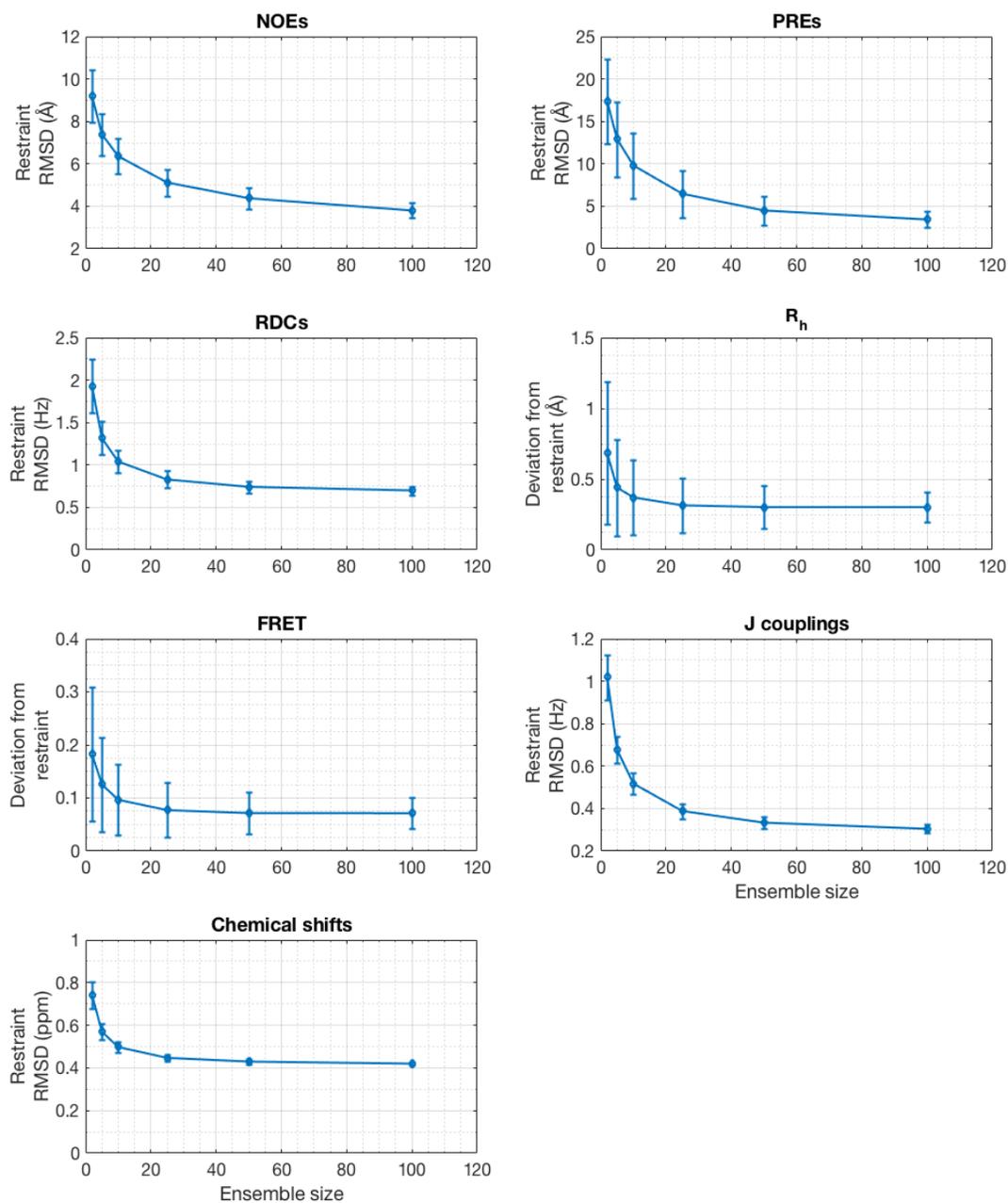
		RMSDs:						
		NOE (Å)	PRE (Å)	RDC (Hz)	FRET (Eff)	R_h (Å)	JC (Hz)	CS (ppm)
Number of restraints:		93	68	28	1	1	47	267
Random ensemble RMSD:		3.80 (0.35)	3.44 (0.94)	0.70 (5.1e-2)	0.07 (3.0e-2)	9.3e-2 (6.7e-2)	0.30 (2.0e-2)	0.42 (7.9e-3)
Ensembles optimized using:	NOE	1.41 (1.5e-2)	3.07 (0.20)	0.78 (3.1e-2)	2.7e-2 (1.6e-2)	0.32 (7.5e-2)	0.34 (1.5e-2)	0.43 (4.8e-3)
	PRE	3.72 (0.30)	1.08 (0.13)	0.63 (5.9e-2)	3.9e-2 (3.1e-2)	0.16 (0.12)	0.38 (2.4e-2)	0.41 (8.2e-3)
	RDC	4.12 (0.40)	5.02 (1.66)	3.6e-2 (4.3e-3)	0.12 (2.4e-2)	0.37 (9.3e-2)	0.33 (1.9e-2)	0.40 (5.1e-3)
	FRET	3.70 (0.28)	2.94 (0.39)	0.71 (5.2e-2)	1.9e-7 (2.5e-7)	0.17 (7.9e-2)	0.31 (2.1e-2)	0.42 (7.8e-3)
	R_h	3.78 (0.33)	3.33 (0.73)	0.70 (5.1e-2)	7.1e-2 (2.3e-2)	3.3e-5 (2.3e-5)	0.30 (2.1e-2)	0.42 (7.7e-3)
	JC	4.09 (0.34)	4.10 (1.16)	0.71 (4.4e-2)	0.12 (2.6e-2)	0.11 (7.2e-2)	2.4e-2 (1.8e-3)	0.45 (5.8e-1)
	CS	4.33 (0.19)	4.57 (0.70)	0.70 (2.6e-3)	1.5e-2 (1.1e-2)	0.35 (6.0e-2)	0.42 (9.6e-3)	0.34 (1.5e-3)



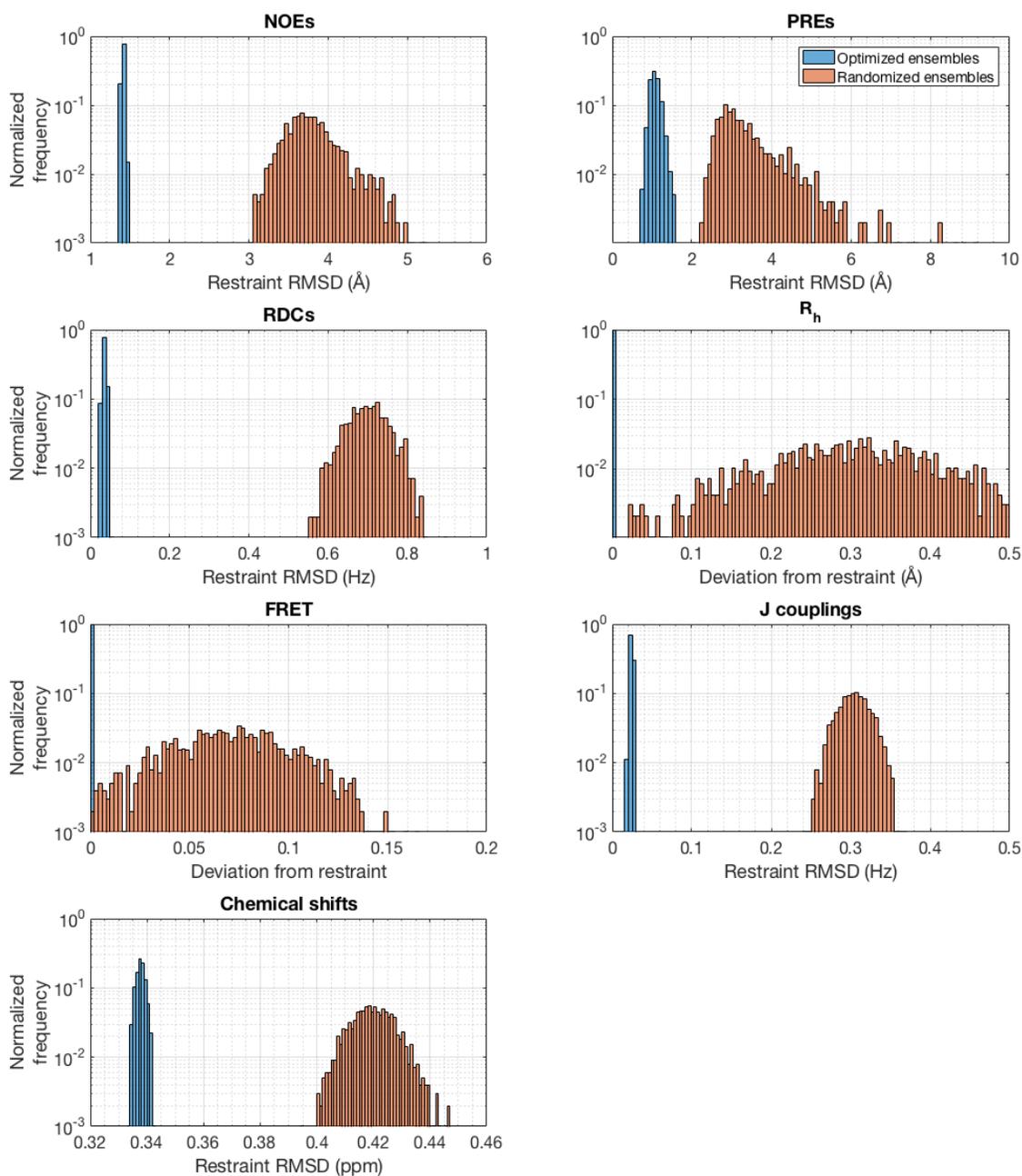
Supplementary Figure 1: Full ensemble agreement with experimental data. Shown as deviation from restraint based on ensemble-averaged properties for NOEs, PREs, RDCs, J couplings, and chemical shifts. Shown as per-structure deviation from restraint for R_h . Shown as per-structure FRET efficiency for FRET, experimental value = 0.55 ± 0.02 .



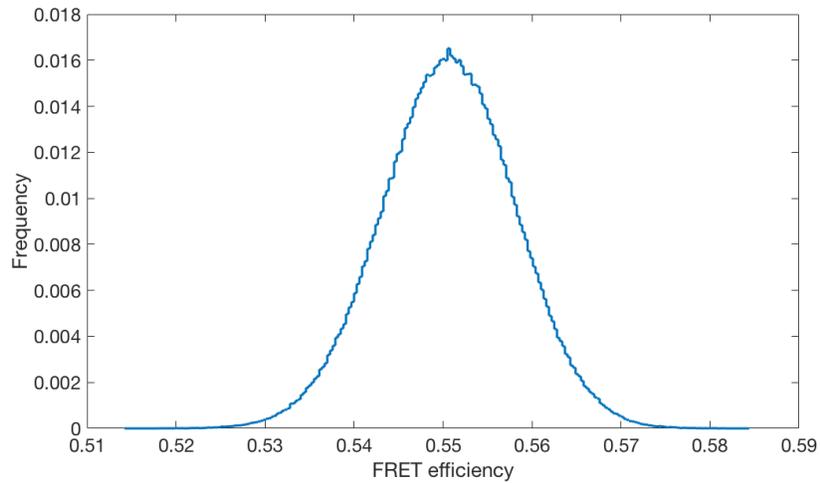
Supplementary Figure 2: NOE scores for randomized ensembles of different sizes, using different values of experimental uncertainty. Error bars are calculated as \pm one standard deviation across 1,000 replicates.



Supplementary Figure 3: Restraint RMSDs (NOEs, PREs, RDCs, J couplings, Chemical shifts) or deviation from restraint (R_h , FRET) for random ensembles of different sizes.



Supplementary Figure 4: Histograms of RMSDs for randomized and optimized ensembles of 100 conformers each, according to the experimental data type against which optimization and scoring are performed. Optimized ensembles are in blue, randomized ensembles are in orange. Each set represents 1,000 optimized and randomized ensembles.



Supplementary Figure 5: Histogram of resampled FRET efficiencies for 100-conformation best-fit ensembles upon varying values of involved nuisance parameters for estimation of back-calculation uncertainty. Each parameter is modeled with a normal distribution: N_{linker} with $\mu = 7$ based on the experimental work and $\sigma = 3$ based on the range of estimated values from simulated work³⁹, r_0 with $\mu = 4.4$ nm and $\sigma = 0.2$ nm from the experimental work, and v with $\mu = 0.5$ and $\sigma = 0.05$ to model slight uncertainty around the assumption of an ideal chain. The 10^6 resampled efficiencies have a standard deviation = 0.0074, which is used as the value of σ_q .