Constructing ensembles for intrinsically disordered proteins

Citation: Fisher, Charles K, and Collin M Stultz. "Constructing Ensembles for Intrinsically Disordered Proteins." Current Opinion in Structural Biology 21, no. 3 (June 2011): 426–31.

As Published: http://dx.doi.org/10.1016/j.sbi.2011.04.001

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/99137

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Published in final edited form as:
Curr Opin Struct Biol. 2011 June ; 21(3): 426–431. doi:10.1016/j.sbi.2011.04.001.

Constructing Ensembles for Intrinsically Disordered Proteins

Charles K. Fisher[a] and **Collin M. Stultz**[a,b,*]

[a]Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02139-4307, United States of America

[b]Harvard-MIT Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, United States of America

Abstract

The relatively flat energy landscapes associated with intrinsically disordered proteins makes modeling these systems especially problematic. A comprehensive model for these proteins requires one to build an ensemble consisting of a finite collection of structures, and their corresponding relative stabilities, which adequately capture the range of accessible states of the protein. In this regard, methods that use computational techniques to interpret experimental data in terms of such ensembles are an essential part of the modeling process. In this review, we critically assess the advantages and limitations of current techniques and discuss new methods for the validation of these ensembles.

Introduction

Thermal fluctuations cause proteins to sample a variety of conformations during their biological lifetime, where the probability of each conformation is determined by the topography of the underlying energy landscape. Folded proteins exhibit energy landscapes that have a well-defined global energy minimum (Fig. 1A). By contrast, intrinsically disordered proteins (IDPs) correspond to a class of polypeptides with relatively flat energy landscapes (Fig. 1B) and consequently, these proteins sample a relatively large and diverse set of conformations at room temperature [1–2]. A great deal of interest in understanding the structure of IDPs has emerged due to their proposed role in neurodegenerative disorders such as Parkinson's and Alzheimer's diseases [3–11]. Therefore, a detailed characterization of these systems could pave the way to the development of new therapeutics through structure based drug design [12–13].

The earliest attempts at modeling disordered protein states were aimed at describing folded proteins under denaturing conditions[14–17]. Denatured proteins and IDPs share the characteristic that experimental observables correspond to averages over a diverse ensemble of conformations. Therefore, the typical approach to constructing an ensemble for both

*Electronic mail to: cmstultz@mit.edu.

>We critically discuss methods for modeling Intrinsically Disordered Proteins. >Both the advantages and limitations of existing methods are analyzed. >We further outline the major challenges to modeling these proteins and discuss new methods for the validation of these approaches.

folded proteins under denaturing conditions and IDPs is to generate a set of conformations that have ensemble averages that agree with experimental values. When formulated in this way, the approach is straightforward; i.e., generate a diverse set of conformations and then find a subset of structures and their relative stabilities (or weights) that agree with experiment. In other cases, the ensemble is constructed using purely theoretical methods and the predicted data are compared to experiment[18–19]. While important insights have been obtained using this latter approach, using experimental data to guide the construction of the ensemble helps to limit the space of possible solutions.

In practice, constructing an ensemble from experimental data is quite a challenging task because the amount of data that are typically available pales in comparison to the number of parameters needed to *uniquely* define the ensemble. In other words, there are typically many different ensembles that agree with any given set of experimental data. Hence the optimization problem described above leads to degenerate solutions. In light of this, how does one reliably infer a set of conformations and weights that capture the essential features of the energy landscape, from the available experimental data? In this article, we review recent advances in this area and provide discussion regarding the advantages and limitations of various techniques.

Sources of experimental data

To date, most of the experimental measurements that have been used to guide the construction of unfolded ensembles correspond to observables obtained via NMR spectroscopy. Examples of such measurements include chemical shifts, which provide information about local conformational preferences [20–22**], scalar couplings, which report on backbone dihedral angles, [23*], Residual Dipolar Couplings (RDCs), which report on the angle of a bond relative to an external frame of reference, [8,22**,24*–28] and Paramagnetic Relaxation Enhancement (PRE) effects, which provide long-range distance restraints [7,29**]. These data can be combined with radius of gyration (R_g) estimates or scattering profiles from Small Angle X-Ray Scattering (SAXS) [9,30–32]. FRET studies have also provided additional experimental data that can further limit the available space of solutions [33–35]. In most cases, the measured quantities correspond to ensemble averages over a vast set of distinct conformations that interconvert on a fairly rapid timescale. Therefore to ensure that the calculated ensemble agrees with experiment, one must compute the corresponding observable from the ensemble and compare these data to the corresponding experimental measurements.

A number of programs, such as SHIFTX [36], SPARTA [37], CamShift [38] and SHIFTS [39] among others, can be used to estimate chemical shifts from the atomic coordinates of a protein structure. PALES [40] is typically used to estimate the RDCs, although other methods based on the radius of gyration tensor or the shape of the molecular surface are also available [41–43]. It is important to keep in mind that the prediction of experimental observables from a structure may not be entirely accurate [36,42]. For example, calculated RDCs are often rescaled to better fit the data to account for uncertainty in the concentration of the alignment medium [8]. It is also possible that experimental conditions, such as the introduction of an alignment medium for measuring RDCs [44] or introducing a spin label to the protein to measure PREs, could perturb the underlying ensemble.

Validation of Ensemble Building Methods

Before discussing specific algorithms used for constructing ensembles it is useful to introduce a technique, which we will refer to as the reference ensemble method, which has become a standard tool for evaluating the performance of these methods [22**,45*–47]. The reference ensemble method is illustrated in Fig. 2. A reference ensemble is a pre-defined

"truth," i.e. a pre-specified set of conformations and their statistical weights that can be used to calculate synthetic experimental data. The synthetic experimental data, therefore, correspond to measurements that would arise if the protein in question had the conformational distribution of the reference ensemble.

Within this formalism the marker of success is clear. That is, can the ensemble building algorithm being evaluated reproduce the reference ensemble using only the synthetic experimental data? More precisely, the synthetic experimental data is fed into the ensemble building algorithm, which produces an "output" ensemble. Comparison of the reference and output ensembles allows one to assess the performance of the algorithm (Fig. 2). This method is powerful because it allows one to control the sources of uncertainty in the problem; one can use the reference ensemble method to analyze the effects of experimental error or even systematic error in the prediction algorithms on the resulting ensemble, or these sources of uncertainty can be eliminated completely to examine the performance of the algorithm under ideal circumstances.

Ensemble-restrained MD simulations

Restrained MD simulations introduce a term into the potential function that biases the simulation towards regions of conformational space that agree with experimental observations. For an IDP, the restraints should be applied to an entire ensemble rather than an individual structure [45*]. This is accomplished by simulating multiple replicas of the protein in parallel and calculating the biasing potential based on averages taken over all of the replicas [45*,48]. Ganguly and Chen [29**] used the reference ensemble method to examine the performance of ensemble-restrained MD using only PRE derived distance restraints. The study found that the method performed poorly unless a very large number of distance restraints were used - more than 4 PRE distance restraints per residue per replica. They suggested that ensembles with a large number of replicas may be under-restrained because only a small number of conformations in the ensemble need to satisfy the restraint due to the r^{-6} weighting of the PRE effect. Subsequently, Allison et al. [45*] applied ensemble-restrained MD using approximately 2 PRE derived distance restraints per residue per replica in combination with the R_g to model the ensemble of α-synuclein. The R_g provides independent information to the PRE distances, and therefore helps to mitigate the problem encountered by Ganguly and Chen. In addition, the studies were performed on different proteins and used slightly different methods for implementing the distance restraints. Allison et al. first used the reference ensemble method to show that the number and type of restraints used in the study was adequate for accurately constructing ensembles and also cross-validated their ensemble using predictions of other types of experimental data that were not incorporated in the modeling procedure.

These studies highlight both the strengths and disadvantages of the method. When a relatively large number of constraints are known, it is possible to obtain an accurate representation of the ensemble. Nevertheless, it is difficult to know exactly how many restraints are sufficient to accurately model the protein. Also, it is likely that not all types of experimental data are created equal in that some types of data may be more informative than others. For example, while data obtained from PRE experiments are useful for estimating average inter-residue distances in an ensemble, it is important to note that in order to obtain these data one must introduce a paramagnetic probe into the protein. Furthermore, it is not clear whether such modifications alter the conformational distribution of the protein [49]. Therefore, it is likely that both the number of experimental constraints and the types of experimental data used are important factors that determine whether the method will find an accurate ensemble. While ensemble restrained MD simulations constitutes a useful and important tool in our arsenal for modeling IDPs, additional studies are needed to more fully

define the limitations of the method and under what circumstances it is expected to yield accurate results.

Ensemble construction using a pre-defined conformational library

Another method for constructing ensembles for IDPs is to first generate a library of conformations and then to select a subset of conformations from this library such that averages calculated from this subset agree with the experimental data. The initial conformational library may be generated with MD, perhaps using techniques to enhance conformational sampling (see [50] for a review) like replica exchange [51], accelerated MD [46] or quenched MD [21], by piecing together small peptide fragments that have been constructed using MD simulations [22**], or with statistical coil models such as Flexible-Meccano [24*,47] or TraDES [23*,52–53]. The statistical coil models usually involve a simplified potential function taking into account backbone dihedral angle distributions and some type of excluded volume interactions.

Once the conformational library has been constructed, a smaller sample of conformations is selected to minimize the difference between the predicted and experimental data. The selection process typically involves some form of Monte Carlo, e.g. ENSEMBLE [17,52–54] or Sample and Select (SAS) [34*,55–56], or evolutionary algorithm, e.g. ASTEROIDS [20,24*,47], that performs a stochastic search for a set of conformations that give calculated values that agree with the experimental measurements. The conformations in the ensembles generated using ASTEROIDS, ENSEMBLE (in its most recent applications) and SAS are equally weighted, so that the statistical properties of the conformational distribution are represented implicitly by the composition of the final set of conformations rather than by explicit statistical weights for each conformation. Salmon et al. [47] and Mittag et al. [52] recently used the ASTEROIDS and ENSEMBLE methods, respectively, to identify transient long-ranged contacts in two IDPs.

A related approach to the selection of a sample of structures from a conformational library is to explicitly estimate the statistical weight of each conformation in the library. In fact, the selection of conformations from a library is really just a particular way of specifying weights over the entire structural library, where a conformation is assigned a weight of 0 if it is not included in the ensemble and a weight of $1/n$ if it is included in the ensemble (where n is the number of chosen structures). Hence at its core, selecting a subset of structures from a larger structural library is a subset of the more general problem of specifying a set of weights for a given set of structures. In the Energy-minima Mapping and Weighting (EMW) algorithm, both the sample of conformations and their statistical weights are optimized simultaneously using a simulated annealing protocol [21,57]. In this instance the weight of any given structure can vary from 0 to 1, and in general the set of structural weights spans a wide range of possibilities. Application of EMW to the second microtubule binding repeat of tau protein as well as an aggregation-prone mutant suggested that the mutation lead to aggregation-initiating regions of tau protein to adopt more extended states, thereby promoting the formation of aggregates rich in extended secondary structure [21].

Degeneracy and model construction

Degeneracy of the ensembles with respect to the experimental measurements is one problem that plagues the construction of IDP ensembles. At its core, the problem of degeneracy arises because in practice the number of experimental constraints is small relative to the number of degrees of freedom that are needed to uniquely specify the ensemble. Fisher, Huang and Stultz [22**] used the reference ensemble method to show that one can often find many sets of statistical weights (for a pre-specified set of conformations) that will agree with a given dataset to within experimental uncertainty, but are all very different from the

"true" weights. Furthermore, it was shown that not all reference ensembles are created equal, some are easy to reconstruct while others are very hard, and that uncertainty in the predicted values of the experimental observables for each structure can have a detrimental effect on algorithm performance. The take home message is that agreement with experimental data alone does not guarantee that an ensemble is accurate. While this is known to be a significant problem, few methods have attempted to explicitly deal with the degeneracy issue. In prior work, Huang and Stultz [21] and Marsh and Forman-Kay [54] utilized similar approaches to mitigate the problem of degenerate ensembles by constructing multiple ensembles and making inferences based on characteristics they had in common. The idea is that although one cannot be sure that the final ensemble is accurate, if one generates many different ensembles and identifies structural features that are recur in the different ensembles, it is likely that inferences made on the preserved structural elements are accurate. The drawback here is that what defines a preserved structural feature is somewhat subjective. Moreover, it is not clear how many ensembles one needs to analyze to have confidence that the identified preserved structural features are not present because of chance alone.

An alternate approach is to calculate quantitative estimates of our uncertainty of a constructed ensemble, given a pre-specified set of experimental measurements. Bayesian inference provides a statistical framework for quantifying this uncertainty and for propagating it to parameters estimated from the ensemble in the form of confidence intervals [22**,30,58]. A recently described Bayesian Weighting (BW) algorithm [22**] uses a Bayesian framework for estimating the weights of a pre-defined set of conformations. In fact, the BW algorithm computes a probability distribution over all possible ways of weighting the conformations in the library. The global uncertainty of the statistical weights can be calculated from this probability distribution in terms of a posterior divergence or uncertainty parameter – a metric which is akin to the standard deviation of a Gaussian distribution. The posterior divergence ranges from 0 to 1, corresponding to a high and low confidence in the estimated population weights, respectively. An empirical study using reference ensembles suggested that when the uncertainty parameter is 0, one can be confident that the ensemble is accurate, assuming that the pre-specified set of structures is diverse enough to capture a wide range of energetically favorable conformations. Nonetheless, when the uncertainty parameter is non-zero we can still express calculated values from the ensemble with rigorous confidence intervals. The BW algorithm was applied to the K18 construct of Tau protein to construct confidence intervals examining the presence of long-ranged contacts [22**]. It was found that mutations known to affect the K18 aggregation propensity preferentially occur near regions that were identified to be involved in long-ranged contacts with high confidence.

Conclusions and future directions

Any comprehensive description of an IDP necessitates the construction of an ensemble – a finite collection of conformations and weights – that capture the essence of the conformational distribution of the protein. A variety of different approaches have been developed for constructing ensembles for IDPs, each of which has its own advantages and limitations. In the past few years, a number of advances have been made in our ability to model the conformational ensembles of IDPs. Many of these advances, including the use of reference ensembles, cross validation and Bayesian statistics, have focused on developing methods for assessing the validity of structural models.

There remain plenty of open methodological and conceptual problems in the computational modeling of IDP ensembles. Because each of the techniques discussed above is based on a specific set of assumptions (e.g. that the experimental conditions do not perturb the

ensemble) it is crucial to continue to develop methods for validating ensembles and for estimating their associated uncertainty. As more IDP ensembles are constructed, it will also become important to develop methods for comparing different ensembles [59] or generating consensus estimates from multiple ensembles.

We learn a great deal more about the IDP structural universe with each model of an IDP conformational ensemble that is constructed. In the last few years alone, we have learned that IDPs are not well described as pure random coils [23*], that some IDP ensembles may contain long-ranged contacts [7,22**,47] and that these contacts may be associated with important functional properties [22**]. Due, in part, to progress in computational modeling our understanding of IDPs is growing at an increasingly fast rate and we look forward to the surprising results that are yet to be discovered. Progress in the field would be greatly enhanced if all new ensemble generation algorithms and experimental data on IDPs (e.g., chemical shift values, RDCs, etc), were deposited in publically accessible repositories like the BioMagResBank [60] or the Database of Protein Disorder [61].

References

1. Huang A, Stultz CM. Finding Order within Disorder: Elucidating the Structure of Proteins Associated with Neurodegenerative Disease. Future Medicinal Chemistry. 2009; 1:467–482. [PubMed: 21426127]
2. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics. 2008; 9 Suppl 2:S1.
3. von Bergen M, Friedhoff P, Biernat J, Heberle J, Mandelkow EM, Mandelkow E. Assembly of Tau protein into Alzheimer paired helical filaments depends on local sequence motif (306 VQIVYK 311) forming beta-structure. PNAS. 2000; 97:5129–5134. [PubMed: 10805776]
4. Barghorn S, Zheng-Fischhofer Q, Ackmann M, Biernat J, Bergen Mv, Mandelkow EM, Mandelkow E. Structure, Microtubule Interactions, and Paired Helical Filament Aggregation by Tau Mutants of Frontotemporal Dementias. Biochemistry. 2000; 39:11714–11721. [PubMed: 10995239]
5. Fischer D, Mukrasch MD, Biernat J, Bibow S, Blackledge M, Griesinger C, Mandelkow E, Zweckstetter M. Conformational Changes Specific for Pseudophosphorylation at Serine 262 Selectively Impare Binding of Tau to Microtubules. Biochemistry. 2009; 48:10047–10055. [PubMed: 19769346]
6. Jeganathan S, von Bergen M, Brutlach H, Steinhoff HJ, Mandelkow E. Global hairpin folding of tau in solution. Biochemistry. 2006; 45:2283–2293. [PubMed: 16475817]
7. Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M. Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. PLOS Biology. 2009; 7:399–414.
8. Mukrasch MD, Markwick P, Biernat J, von Bergen M, Bernardo P, Griesinger C, Mandelkow E, Zweckstetter M, Blackledge M. Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. Journal of the American Chemical Society. 2006; 129:5235–5243. [PubMed: 17385861]
9. Mylonas E, Hacher A, Bernardo P, Blackledge M, Mandelkow E, Svergun DI. Domain Conformation of Tau Protein Studied by Solution Small-Angle X-ray Scattering. Biochemistry. 2008; 47:10345–10353. [PubMed: 18771286]
10. von Bergen M, Barghorn S, Li L, Marx A, Biernat J, Mandelkow EM, Mandelkow E. Mutations of Tau Protein in Frontotemporal Dementia Promote Aggregation of Paired Helical Filaments by Enhancing Local beta-structure. Journal of Biological Chemistry. 2001; 276:48165–48174. [PubMed: 11606569]
11. Yao T-M, Tomoo K, Ishida T, Hasegawa H, Sasaki M, Taniguchi T. Aggregation Analysis of the Microtubule Binding Domain in Tau Protein by Spectroscopic Methods. The Journal of Biochemistry. 2003; 134:91–99.

12. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins. 2007; 66:399–421. [PubMed: 17096427]
13. Metallo SJ. Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol. 2010; 14:481–488. [PubMed: 20598937]
14. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol. 1997; 268:170–184. [PubMed: 9149150]
15. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol. 1997; 268:158–169. [PubMed: 9149149]
16. Neri D, Billeter M, Wider G, Wuthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science. 1992; 257:1559–1563. [PubMed: 1523410]
17. Choy W-Y, Forman-Kay JD. Calculation of Ensembles of Structures Representing the Unfolded State of an SH3 Domain. Journal of Molecular Biology. 2001; 308:1011–1032. [PubMed: 11352588]
18. Mathieson SI, Penkett CJ, Smith LJ. Characterisation of side-chain conformational preferences in a biologically active but unfolded protein. Pac Symp Biocomput. 1999:542–553. [PubMed: 10380226]
19. Fawzi NL, Phillips AH, Ruscio JZ, Doucleff M, Wemmer DE, Head-Gordon T. Structure and dynamics of the Abeta(21–30) peptide from the interplay of NMR experiments and molecular simulations. J Am Chem Soc. 2008; 130:6145–6158. [PubMed: 18412346]
20. Jensen MR, Salmon L, Nodet G, Blackledge M. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. J Am Chem Soc. 2010; 132:1270–1272. [PubMed: 20063887]
21. Huang A, Stultz CM. The Effect of a delta-K280 Mutation on the Unfolded State of a Microtubule-Binding Repeat in Tau. PLOS Computational Biology. 2008; 4:1–12.
22. Fisher CK, Huang A, Stultz CM. Modeling intrinsically disordered proteins with bayesian statistics. J Am Chem Soc. 2010; 132:14919–14927. [PubMed: 20925316] The authors use reference ensembles to study the uncertainty in conformational ensembles derived from experimental data and introduce the Bayesian Weighting (BW) method to determine ensembles while providing quantitative error metrics. An ensemble for the K18 isoform of intrinsically disordered Tau protein is constructed providing evidence for long-ranged contacts.
23. Marsh JA, Forman-Kay JD. 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; 391:359–374. [PubMed: 19501099] The authors provide a detailed discussion of the ENSEMBLE program for generating conformational ensembles with a variety of types of experimental observables. An ensemble is constructed for the drk N-terminal SH3 domain, which posesses significant amounts of residual structure.
24. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. J Am Chem Soc. 2009; 131:17908–17918. [PubMed: 19908838] The ASTEROIDS method for constructing ensembles using residual dipolar couplings and structures generated using the *flexible-meccano* statistical coil model is introduced and is validated using a reference ensemble. Using their approach, the authors study local backbone conformational preferences in urea denatured Ubiquitin.
25. Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernado P, Blackledge M. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure. 2009; 17:1169–1185. [PubMed: 19748338]
26. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci U S A. 2005; 102:17002–17007. [PubMed: 16284250]
27. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc. 2007; 129:5656–5664. [PubMed: 17411046]

28. Bernardo P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M. Defining Long-Range Order and Local Disorder in Native alpha-Synuclein Using Residual Dipolar Couplings. Journal of the American Chemical Society. 2005; 127:17968–17969. [PubMed: 16366524]
29. Ganguly D, Chen J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. J Mol Biol. 2009; 390:467–477. [PubMed: 19447112] The reference ensemble method is used to show that PRE distance restraints may not be sufficient for modeling disordered protein ensembles with ensemble-restrained MD simulations, even under idealized conditions.
30. Yang S, Blachowicz L, Makowski L, Roux B. Multidomain assembled states of Hck tyrosine kinase in solution. Proc Natl Acad Sci U S A. 2010; 107:15757–15762. [PubMed: 20798061]
31. Rozycki B, Kim YC, Hummer G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure. 2011; 19:109–116. [PubMed: 21220121]
32. Petsko GA, Ringe D. Fluctuations in protein structure from X-ray diffraction. Annu Rev Biophys Bioeng. 1984; 13:331–371. [PubMed: 6331286]
33. Ferreon AC, Moran CR, Gambin Y, Deniz AA. Single-molecule fluorescence studies of intrinsically disordered proteins. Methods Enzymol. 2010; 472:179–204. [PubMed: 20580965]
34. Chen Y, Campbell SL, Dokholyan NV. Decipering Protein Dynamics from NMR Data Using Explicit Structure Sampling and Selection. Biophysical Journal. 2007; 93:2300–2306. [PubMed: 17557784] The authors introduce the Sample and Select method, in which they select a subset of conformations from a structural library based on NMR S^2 order parameters.
35. Huang F, Rajagopalan S, Settanni G, Marsh RJ, Armoogum DA, Nicolaou N, Bain AJ, Lerner E, Haas E, Ying L, et al. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. Proc Natl Acad Sci U S A. 2009; 106:20758–20763. [PubMed: 19933326]
36. Neal S, Nip AM, Zhang H, Wishart DS. Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. Journal of Biomolecular NMR. 2003; 26:215–240. [PubMed: 12766419]
37. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR. 2007; 38:289–302. [PubMed: 17610132]
38. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. J Am Chem Soc. 2009; 131:13894–13895. [PubMed: 19739624]
39. Xu XP, Case DA. Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database. J Biomol NMR. 2001; 21:321–333. [PubMed: 11824752]
40. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. Journal of the American Chemical Society. 2000; 122:3791–3792.
41. Almond A, Axelsen JB. Physical interpretation of residual dipolar couplings in neutral aligned media. J Am Chem Soc. 2002; 124:9986–9987. [PubMed: 12188652]
42. Berlin K, O'Leary DP, Fushman D. Improvement and analysis of computational methods for prediction of residual dipolar couplings. J Magn Reson. 2009; 201:25–33. [PubMed: 19700353]
43. Fernandes MX, Bernado P, Pons M, Garcia de la Torre J. An analytical solution to the problem of the orientation of rigid particles by planar obstacles. Application to membrane systems and to the calculation of dipolar couplings in protein NMR spectroscopy. J Am Chem Soc. 2001; 123:12037–12047. [PubMed: 11724612]
44. Dames SA, Aregger R, Vajpai N, Bernado P, Blackledge M, Grzesiek S. Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids. J Am Chem Soc. 2006; 128:13508–13514. [PubMed: 17031964]
45. Allison JR, Varnai P, Dobson CM, Vendruscolo M. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. J Am Chem Soc. 2009; 131:18314–18326. [PubMed: 20028147] The authors use the reference ensemble method to show that ensemble-restrained MD simulations with a large number of PRE restraints combined with the radius of gyration can provide an accurate model of an ensemble. An ensemble for the IDP α-synuclein is constructed and analyzed using a variety of tools developed by the authors.

46. Chen LY, Horing NJ. An exact formulation of hyperdynamics simulations. J Chem Phys. 2007; 126:224103. [PubMed: 17581040]
47. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M. NMR characterization of long-range order in intrinsically disordered proteins. J Am Chem Soc. 2010; 132:8407–8418. [PubMed: 20499903]
48. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature. 2005; 433:128–132. [PubMed: 15650731]
49. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol. 2007; 17:3–14. [PubMed: 17250999]
50. Lei H, Duan Y. Improved sampling methods for molecular simulation. Current Opinion in Structural Biology. 2007; 17:187–191. [PubMed: 17382533]
51. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. 1999; 314:141–151.
52. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure. 2010; 18:494–506. [PubMed: 20399186]
53. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure. 2010; 18:1094–1103. [PubMed: 20826336]
54. Marsh JA, Forman-Kay JD. Structure and Disorder of in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. Journal of Molecular Biology. 2009; 391:359–374. [PubMed: 19501099]
55. Stelzer AC, Frank AT, Bailor MH, Andricioaei I, Al-Hashimi HM. Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. Methods. 2009; 49:167–173. [PubMed: 19699798]
56. Frank AT, Stelzer AC, Al-Hashimi HM, Andricioaei I. Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. Nucleic Acids Res. 2009; 37:3670–3679. [PubMed: 19369218]
57. Yoon MK, Venkatachalam V, Huang A, Choi BS, Stultz CM, Chou JJ. Residual structure within the disordered C-terminal segment of p21(Waf1/Cip1/Sdi1) and its implications for molecular recognition. Protein Science. 2009; 18:337–347. [PubMed: 19165719]
58. Rieping W, Habeck M, Nilges M. Inferential Structure Determination. Science. 2005; 309:303–306. [PubMed: 16002620]
59. Lindorff-Larsen K, Ferkinghoff-Borg J. Similarity Measures for Protein Ensembles. PLoS ONE. 2009; 4
60. Doreleijers JF, Mading S, Maziuk D, Sojourner K, Yin L, Zhu J, Markley JL, Ulrich EL. BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. J Biomol NMR. 2003; 26:139–146. [PubMed: 12766409]
61. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, et al. DisProt: the database of disordered proteins. Nucleic Acids Research. 2007; 35:D786–D793. [PubMed: 17145717]

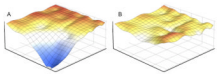

Figure 1.
The energy landscape of a folded protein (A) exhibits a well-defined minimum energy state corresponding to the folded conformation. In comparison, the energy landscape of an IDP (B) lacks a deep energy minimum.

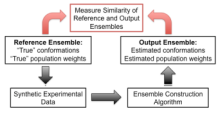

Figure 2.
A schematic representation of the Reference Ensemble Method for validating ensemble construction algorithms.