# Chapter 2
# Structure and Dynamics of Intrinsically Disordered Proteins

**Biao Fu and Michele Vendruscolo**

**Abstract** Intrinsically disordered proteins (IDPs) are involved in a wide range of essential biological processes, including in particular signalling and regulation. We are only beginning, however, to develop a detailed knowledge of the structure and dynamics of these proteins. It is becoming increasingly clear that, as IDPs populate highly heterogeneous states, they should be described in terms of conformational ensembles rather than as individual structures, as is instead most often the case for the native states of globular proteins. Within this context, in this chapter we describe the conceptual tools and methodological aspects associated with the description of the structure and dynamics of IDPs in terms of conformational ensembles. A major emphasis is given to methods in which molecular simulations are used in combination with experimental nuclear magnetic resonance (NMR) measurements, as they are emerging as a powerful route to achieve an accurate determination of the conformational properties of IDPs.

**Keywords** Structure · Dynamics · Molecular dynamics · Conformational ensembles

## 1 Introduction: From Average Structures to Conformational Ensembles

Intrinsically disordered proteins (IDPs) play crucial roles in many aspects of molecular and cell biology, as these proteins are involved in a variety of signalling and regulation processes as well as being implicated in a range of neurodegenerative and systemic disorders such as Alzheimer's and Parkinson's diseases, and type II diabetes (Dyson and Wright 2005; Knowles et al. 2014; Uversky 2013). From the point of view of structural biology, IDPs pose formidable challenges since they are conformationally highly heterogeneous (Fig. 2.1) and are thus not readily amenable to the standard approaches for structure determination that have been developed for folded proteins.

---

M. Vendruscolo (✉) · B. Fu
Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK
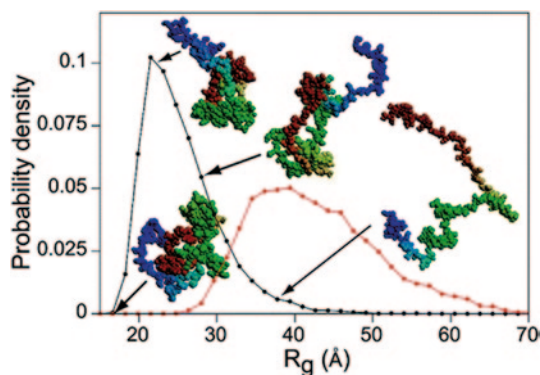e-mail: mv245@cam.ac.uk

**Fig. 2.1** IDPs are conformationally highly heterogeneous. This fundamental aspect of the nature of IDPs is illustrated here by the probability distribution of the radius of gyration, $Rg$, of α-synuclein (*black line*), an IDP associated with Parkinson's disease (Dedmon et al. 2005). The values of $Rg$ range from about 18 Å to more than 40 Å. For comparison the probability distribution of a polypeptide chain with the same length as α-synuclein in a random coil state is also shown (*red line*) (Dedmon et al. 2005)

Native states are also undergoing structural fluctuations, the dynamics of which are important for enzymatic catalysis, ligand binding and the formation of biomolecular complexes (Frauenfelder et al. 1991; Fersht 1999; Karplus and Kuriyan 2005; Mittermaier and Kay 2006; Vendruscolo and Dobson 2006; Boehr et al. 2009). The dynamics of native states are usually represented by a conformational variability around a well-defined structure, and powerful techniques are available to calculate them and their related conformational fluctuations (Brooks et al. 1983; Brunger et al. 1998; Schwieters et al. 2006). This type of description, however, is not suitable in the case of highly heterogeneous states because in such states, in the absence of a specific reference structure, an IDP populates a wide range of conformations having very dissimilar structures (Varadi et al. 2014).

The characterization of the behaviour of IDPs requires novel approaches with respect to standard protein structure determination procedures. The gold standard for the determination of the structures of native states is represented by X-ray crystallography, a technique that allows the positions of all the atoms to be identified with great accuracy through the mapping of the corresponding electron densities (Blundell and Johnson 1976). Nuclear magnetic resonance (NMR) spectroscopy can also achieve this type of accurate positioning of the atoms making up a protein molecule through the measurement of inter-proton distances by exploiting nuclear Overhauser effects (Wüthrich 1986). In this context, the problem of protein structure determination is solved by acquiring an amount of experimental information sufficient to determine essentially all the degrees of freedom of a protein molecule once its sequence and covalent bond topology are known. In the case of IDPs, by contrast, this approach is not possible, since the presence of a wide variety of different conformations prevents the definition of the structural properties of proteins by providing a single list of three-dimensional atomic coordinates.

A powerful conceptual framework in this case is that of statistical mechanics (Chandler 1987; van Kampen 1992). In this type of description the objective is to determine a range of representative conformations populated by IDPs together with their statistical weights. In other words, the aim is to characterise the Boltzmann distributions of IDPs. The reason for adopting this approach is that if one calculates the number of possible states of an IDP, one realizes that no experiment will ever be able to provide sufficient information to determine the atomic coordinates of the exceedingly large number of different conformations that it can explore. To obtain an insight into this issue, one can consider a most common textbook example, in which the velocities of the particles of an ideal gas in a box are provided in terms of a well-defined probability distribution, the Maxwell-Boltzmann distribution (Chandler 1987). The knowledge of such a distribution enables a great variety of properties of the ideal gas to be calculated, and these calculations provide accurate predictions for experimental measurements that can be performed on rarefied weakly-interacting gasses. In this view, the goal of measuring the positions of all the atoms in the myriad different conformations of a protein is not only practically impossible to achieve, but also essentially irrelevant, since one can perform accurate predictions of many aspects of its behaviour even without such knowledge.

For a given IDP, in order to generate an ensemble of structures according to their Boltzmann probabilities, or statistical weights, the availability of only sparse experimental measurements for structure determination can be complemented with the use of *a priori* information, including about covalent bond lengths, dihedral angles and rotameric states of side chains. This type of information can be provided through the use of force fields in molecular dynamics (Brooks et al. 1983; Hornak et al. 2006; Lindorff-Larsen et al. 2012a) or through effective potentials derived from protein structure databases (Das and Baker 2008). In this approach, a computational model of the conformational space populated by IDPs is combined with the information provided by the experimental measurements in order to achieve a description of the structure and dynamics of IDPs simultaneously consistent with the overall theoretical knowledge of the behaviour of these proteins and with the specific observation made about specific systems. As we will describe in the following, a range of different methods have been proposed to combine theoretical knowledge about IDPs and the experimental measurements on them. Before coming to that, however, we address the two major, and in many ways complementary, problems that should be considered in the determination of the structure and dynamics of IDPs.

## 2 The Two Fundamental Problems in the Computational Study of IDPs

The strategy in which experimental data are combined with a theoretical modelling of IDPs requires an ability to generate a relatively accurate sampling of their conformational space. A powerful approach to achieve this result is provided, for

example, by all-atom molecular dynamics simulations (Karplus and Kuriyan 2005; Shaw et al. 2010; Best 2012). In these simulations, the conformational space of a protein is sampled by integrating the equations of motion for a time interval sufficiently long to enable the relevant regions to be explored. There are, however, two major challenges in the implementation of this approach. The first is the 'force field problem' and the second is 'the conformational sampling problem'. We should also note that although we describe these two problems here in the case of all-atom molecular dynamics simulations, they are common to essentially any scheme to sample the conformational space of proteins, as one needs always to evaluate the energy of a given protein conformation and to explore the range of its available conformations.

## 2.1   The Force Field Problem

One of the most fundamental aspects of any theoretical method to describe the behaviour of proteins concerns the ability to associate an energy to a given conformation. In molecular dynamics simulations, the function that associates an energy to a given conformation is called a 'force field' (although rather than a force it is actually an energy, or more precisely, a potential energy). The most common force fields are based on molecular mechanics, in which classical mechanics is used to describe the behaviour of proteins and the interactions are provided in a classical framework, involving a combination of terms describing the covalent bond distances and angles ('bonded terms') and of terms describing other interactions, including van der Waals and Coulomb interactions, between atoms ('non-bonded terms') (Brooks et al. 1983; Hornak et al. 2006; Lindorff-Larsen et al. 2012a).

These energy terms, however, represent only an approximate model of the actual interactions between atoms. Although better representations of these interactions are possible in principle (e.g. through the use of quantum mechanics), they become computationally more expensive and as a consequence they are more seriously affected by the conformational sampling problem (see Sect. 1.2.2) (Brooks et al. 1983; Hornak et al. 2006; Lindorff-Larsen et al. 2012a; Baker and Best 2013). The energies that can be associated with given conformations, therefore, can only be of limited accuracy, and the corresponding exploration of the conformational space is carried out with inaccurate statistical weights. Despite a range of significant recent advances in the improvement of force fields (Lindorff-Larsen et al. 2012a; Bottaro et al. 2013; Baker and Best 2013; Piana et al. 2014), one should thus bear in mind that force fields are not exact. Having said that, the use of molecular dynamics simulations provides a range of opportunities that have been explored in a series of recent studies that are beginning to provide descriptions of the structure and dynamics of IDPs and of the disordered states of other proteins (Lindorff-Larsen et al. 2012b; Camilloni and Vendruscolo 2014; Knott and Best 2012; Krzeminski et al. 2013; Varadi et al. 2014).

## 2.2  The Conformational Sampling Problem

As mentioned above, the number of possible conformations of a protein molecule is enormous. It is thus out of the question to enumerate all such possible conformations using a computer, since it would require an essentially infinite amount of time and memory. In statistical mechanics, however, it is relevant to sample the conformational space only in the regions where the statistical weights are non-negligible. For folded states, this means that only a relatively small number of conformations need to be considered, and indeed single X-ray structures represent the state of a protein quite faithfully. By contrast, many more conformations should be explored for IDPs, as the statistical weights are significantly different from zero for a wide range of different structures.

In molecular dynamics the speed at which the conformational space can be explored is inherently limited by the step of integration of the equations of motion, which is typically of 1 to 2 femtoseconds. Even with the most powerful supercomputers, trajectories can currently be followed up to the millisecond timescale—a feat that involves something like a trillion integration steps! (Shaw et al. 2010; Vendruscolo and Dobson 2011). As IDPs tend to explore their relevant conformational space on longer timescales (e.g. seconds and beyond), one should bear in mind that the sampling will necessarily be incomplete.

Several methods have been proposed to enhance the sampling efficiency. For example, one of the most common ones involves the 'coarse-graining' of the conformational degrees of freedom (Tozzini 2005; Monticelli et al. 2008). In this approach, rather than representing a protein molecule by providing a list of the three-dimensional coordinates of all its atoms, one simplifies the representation by specifying only the most relevant degrees of freedom, such as for instance only the position of the Cα atoms. In coarse-grained approaches, while the integration step becomes much less expensive, the force field becomes less accurate because in eliminating some of the atoms of a protein the corresponding interactions should be incorporated in some averaged manner in the force field, and such averaging is inherently approximate. There is therefore a trade-off between speed in the sampling and accuracy in the energy estimation.

In other approaches, the all-atom representation is maintained but the force field is modified in a controlled manner to bias the sampling towards the relevant regions of the conformational space. One of the first methods proposed for this purpose is that of 'umbrella sampling', in which a weighting function is introduced in the force field to prevent the sampling of structures outside a given region of the conformational space (Chandler 1987). This bias is then removed in order to reweight the conformations and obtain their correct statistical weights. Series of umbrella sampling simulations can be then analysed using the weighted histogram analysis method (WHAM) or its generalizations (Kumar et al. 1992; Hub et al. 2010; Zhu and Hummer 2012). A related method is that of accelerated molecular dynamics, in which the sampling of the conformational space is enhanced by reducing the energy barriers separating the different states populated by a protein (Board et al. 1992;

Markwick et al. 2007). This method modifies the potential energy landscape by raising the energy wells below a given threshold level, while leaving those above this level unaffected. As a result, the barriers between neighbouring energy basins are reduced, allowing the protein to sample regions of the conformational space that cannot be easily accessed in conventional molecular dynamics simulations.

A particularly effective method that is becoming increasingly adopted in IDP simulations is that of metadynamics (Laio and Parrinello 2002; Laio and Gervasio 2008). In this method one assumes that the behaviour of a protein can be described accurately through a small number of collective variables. The basic idea of the method is that the protein is discouraged from returning to the proximity of the conformations that it has already visited by 'remembering' their positions. This idea is implemented by calculating the position of the protein in terms of the collective variables during the simulation and by adding a Gaussian function in this position to the energy landscape of the protein itself. As the simulation progresses, the Gaussian functions accumulate preferentially in the energy minima until the free energy eventually becomes a constant as a function of the collective variables. The three main parameters that control the convergence of the simulations are the time between the addition of Gaussian functions and the height and width of the Gaussian functions themselves.

# 3   Combining Experiments with Simulations Using the Maximum Entropy Principle

As mentioned above, several approaches have been proposed for characterizing non-native states. These approaches differ in the particular way in which the system-dependent experimental measurements are combined with the system-independent theoretical information provided by the force field. A general framework for carrying out this plan is provided by the maximum entropy principle (Pitera and Chodera 2012; Cavalli et al. 2013; Roux and Weare 2013; Boomsma et al. 2014). According to this principle, the conformational space populated by a protein should be the largest possible one compatible with the information available. In the context of molecular dynamics simulations, the incorporation of the information provided by a given set of experimental data to a force field should be carried out in a manner that maximizes the number of conformations that are sampled, with the only requirement that they be compatible with the experimental data. In this sense, the maximum entropy principle provides the opposite prescription to the 'Occam's razor', according to which the minimal number of structures should be determined to generate a set consistent with the available experimental data.

The maximum entropy principle only provides a guideline about how to combine experiments with simulations, and there are many possible alternatives for its practical implementation. For example, in an approach often used for the characterization of the behaviour of IDPs, the experimental information is used to filter out conformations in disagreement with the observations from a previously generated

ensemble of conformations (Choy and Forman-Kay 2001; Bernadó et al. 2005; Heise et al. 2005). The success of this approach relies on the ability of the conformational sampling to explore regions that are populated with significant probability by IDPs, as otherwise it becomes impossible to select conformations consistent with the experimental data. When this condition is met, the maximum entropy principle framework offers a highly effective way to carry out the selection.

An approach that has been investigated extensively in recent years consists of extending the methods of structure determination that have been developed for native states to highly heterogeneous states (Bonvin et al. 1994; Bonvin and Brunger 1995; Burgi et al. 2001; Constantine et al. 1995; Fennen et al. 1995; Kemmink and Scheek 1995; Kessler et al. 1988; Torda et al. 1989; Clore and Schwieters 2004; Lindorff-Larsen et al. 2005). In this approach the experimental information is used to construct structural restraints to be used in molecular simulations. In this case the sampling is biased to take place in regions of conformational space that satisfy the available experimental information. It has been shown that the addition of the bias can be carried out in a manner compatible with the maximum entropy principle (Pitera and Chodera 2012; Cavalli et al. 2013; Roux and Weare 2013; Boomsma et al. 2014). In this context, if the experimental restraints are imposed as averages over a number $N$ of replicas of the protein molecule, the sampling is carried out according to the maximum entropy principle in the limit of large values of $N$ and large values of the force constant in front of the energy restraint term. In practice, it has also been shown that the number $N$ of replicas can be relatively small, ranging from 2 to 16 (Cavalli et al. 2013; Roux and Weare 2013; Boomsma et al. 2014).

By building on these advances, the recently proposed replica-averaged metadynamics (RAM) method (Camilloni et al. 2013; Camilloni and Vendruscolo 2014) combines the advantages of advanced sampling techniques (in this case metadynamics) to improve the conformational sampling problem with the use of experimentally-driven energy biases in the molecular dynamics simulations to improve on the force field problem. In RAM simulations, the replicas needed for the maximum entropy principle implementation of the experimental restraints are also exploited opportunistically to speed up the sampling of conformational space as they are used as collective variables.

## 4 Free Energy Representations of Conformational Ensembles

In order to define the specific conformation of a protein one can provide a list of the coordinates of all its atoms (e.g. by the analysis of electron density maps obtained by X-ray crystallography). In NMR spectroscopy, alternative options include the specification of a large set of distances between atom pairs (e.g. by using NOE-derived interproton distance information), or of orientations of interatomic vectors either with respect to each other (e.g. by considering J couplings) or relative to an external direction (e.g. by the use of residual dipolar couplings (RDCs)). This

information is then readily translated, usually in an unequivocal manner, into atomic coordinates using standard computational methods. This approach, however, as noted above, is unsuitable for IDPs, as these proteins populate a vast number of different conformations, so that conformational ensembles, which include a variety of structures together with their statistical weights, should be specified.

A very effective way to represent conformational ensembles is through the use of free energy landscapes (Boehr et al. 2009; Frauenfelder et al. 1991; Vendruscolo and Dobson 2006). A free energy landscape represents the probability of observing a given value of a given parameter of a system (Fig. 2.2). For example, the free energy landscape $F(R_g)$ as a function of the radius of gyration $R_g$ can be calculated as

$$F(R_g) = -\alpha \log P(R_g) \tag{2.1}$$

where $\alpha$ is a proportionality constant. This free energy landscape is thus proportional to the negative of the logarithm of the probability distribution $P(R_g)$ of the radius of gyration $R_g$. This probability distribution can be calculated from a simulation as

$$P(R_g) = \frac{N(R_g)}{N} \tag{2.2}$$

where $N(R_g)$ is the number of times that the trajectory has visited a conformation with the value $R_g$ and $N$ is the total number of conformations generated during the trajectory. In many cases, the free energy landscape can be calculated as a function of multiple parameters, although when more than two parameters are used the graphical representation becomes less intuitive.

The major advantage of working with free energy landscapes is that they readily give insights about a number of essential properties of IDPs, including: (1) the list
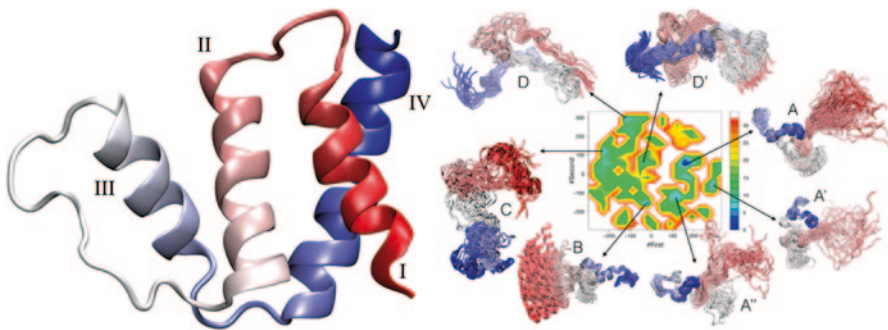


**Fig. 2.2** The free energy landscape of a disordered protein is characterized by the presence of a large number of local minima. This feature is illustrated for the case of the low pH state of acyl-CoA-binding protein (*ACBP*) (Camilloni and Vendruscolo 2014), a four helix bundle protein (*left panel*) that populates a variety of conformationally distinct substates under acidic conditions (*right panel*). The characterization of highly heterogeneous conformational states of proteins in terms of free energy landscapes provides a concise and comprehensive overview of the nature and properties of such states

of their conformational states, (2) the structural features of these states, (3) the extent of their dynamics (in the sense of their equilibrium structural fluctuations), (4) their populations (i.e. their statistical weights), and (5) their mechanisms of function. More specifically, one can use the free energy landscape of a protein to find the number of its states by counting the number of minima, as such minima correspond to regions of high occupation probability, as specified by Eq. (2.1), even if sometimes in disordered states, such as those populated by IDPs, the number of minima can be very large and their populations very small. Furthermore, the extension of the basin around a given free energy minimum provides information about the overall size of the conformational ensemble corresponding to that state, as wide basins will correspond to conformational fluctuations of larger amplitude and hence to larger conformational ensembles.

Most importantly, the knowledge of the different states accessible to a protein is crucial in providing insights into the molecular basis of its function. A very common example is that of the description of the molecular recognition process between two proteins in terms of the 'conformational selection' model (Lange et al. 2008; Boehr et al. 2009). In this model, bound-like conformations are explored by the unbound protein, which then recognises its partner preferentially by binding it in one of these bound-like structures. The characterisation of the free energy landscapes of IDPs can provide a compelling demonstration of this principle (Knott and Best 2012).

## 5   Validation Methods for Conformational Ensembles

As the translation of the experimental measurements into structural restraints and their use in computational methods require a range of assumptions, the resulting structures should be critically assessed in order to establish whether they are correct or not. Ultimately, a powerful guiding principle is that a given conformational ensemble should enable successful predictions to be made about the outcome of the measurements of a variety of different properties of an IDP. In this case such an ensemble represents a comprehensive description of this protein within a statistical mechanics framework. When this happens, one should conclude that the conformations that have been determined, together with their statistical weights, provide a satisfactory representation of the state of a protein as their validity can be tested extensively. Suitable types of experimental parameters available for validating non-native states include fluorescence resonance energy transfer (FRET) derived distances (Haas 2005; Schuler et al. 2002; Sherman and Haran 2006; Moglich et al. 2006) and several NMR observables such as RDCs (Bernadó et al. 2005), paramagnetic relaxation enhancement (PRE) derived distances (Francis et al. 2006; Dedmon et al. 2005; Lindorff-Larsen et al. 2004), J-couplings (Smith et al. 1996), chemical shifts (Korzhnev et al. 2004; Camilloni and Vendruscolo 2014), $R_2$ values (Klein-Seetharaman et al. 2002) and protection factors from hydrogen exchange (Gsponer et al. 2006). The exploitation of these techniques will undoubtedly direct future efforts for increasing the resolution of IDP structures.

Several other methods of validation have been considered in the context of protein structure determination, many of which can be extended readily to IDPs. The internal consistency of a structural determination procedure can be verified by using only a subset of restraints and by testing whether the remaining ones are reproduced (cross-validation) (Spronk et al. 2004). The use of cross-validation, however, is potentially prone to error, especially in the case of highly heterogeneous ensembles of structures (Francis et al. 2006; Dedmon et al. 2005; Lindorff-Larsen et al. 2004). If for example several average inter-atomic distances are imposed on a single molecule, the only conformations compatible with this type of restraint may be compact ones. As a consequence of the time and ensemble averaging during the acquisition of NMR spectra, however, not all of the inter-atomic contacts detected experimentally need to be simultaneously present in any given conformation. For instance, the Δ131Δ fragment of staphylococcal nuclease was represented as a rather compact and native-like ensemble by imposing PRE-derived distances on a single molecule in the simulations (Gillespie and Shortle 1997). When instead the experimental distances were imposed on the average over many molecules, a much more expanded ensemble of conformations was obtained, in which states with an overall native-like topology were present but with very low statistical weights (Francis et al. 2006, Vendruscolo 2007).

In alternative validation methods, the statistical properties of the conformations obtained can be compared with those in the protein structure databases. These methods have become highly sophisticated for native states (Grishaev and Bax 2004; Spronk et al. 2004), and it will become increasingly possible to apply them to non-native states, since large repositories of high resolution structures are beginning to be available (Varadi et al. 2014).

Another highly effective strategy to assess the quality and performance of different structure determination methods is to directly compare their performances on a set of common targets. In the case of structure prediction, this community-wide strategy has been implemented and optimised in the series of Critical Assessment of Protein Structure Prediction (CASP)[1] exercises, which have run every 2 years since 1994 (Moult et al. 2014). In these assessments, experimental groups provide a set of sequences for which they have predicted the structures, and the various computational groups submit their predicted structures within a given deadline. The structures are then released publicly after the completion of the exercise and the performance of the various prediction methods is assessed. For protein structure determination methods this strategy has recently been extended to NMR spectroscopy methods with the Critical Assessment of Automated Structure Determination by NMR (CASD-NMR) assessment (Rosato et al. 2009; Rosato et al. 2012). Within the IDPbyNMR[2] initiative, there is now a plan to extend this assessment to IDPs

---

[1] http://predictioncenter.org/.

[2] IDPbyNMR (High resolution tools to understand the functional role of protein intrinsic disorder) is a Marie Curie activity funded under the FP7 people programme, project number 264257; http://www.idpbynmr.eu/home/.

within the broader Protein Ensemble Database (pE-DB)[3] initiative (Varadi et al. 2014), which is described in more detail in Chap. 11.

# 6   Conclusions

The characterization at high resolution of the conformationally heterogeneous states of IDPs is challenging because the dynamics of these states make it difficult both to obtain accurate experimental measurements and to translate them into a source of structural information. These states have structural features that are difficult to extract by extending experimental and theoretical techniques developed for describing native states (for which an average structure is well-defined) or highly denatured states (for which random coil models are often well-suited), and they require the development of novel computational methods for determining ensembles of structures. In this chapter we have discussed methods to incorporate experimental measurements into molecular simulations, as these methods represent a promising approach that is beginning to provide highly accurate conformational ensembles of IDPs.

# References

Baker CM, Best RB (2013) Matching of additive and polarizable force fields for multiscale condensed phase simulations. J Chem Theor Comp 9(6):2826–2837

Bernadó P, Bertoncini CW, Griesinger C et al (2005) Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. J Am Chem Soc 127(51):17968–17969

Best RB (2012) Atomistic molecular simulations of protein folding. Curr Op Struct Biol 22(1):52–61

Blundell TL, Johnson LN (1976) Protein crystallography. Academic Press, New York

Board JA Jr, Causey JW, Leathrum JF Jr et al (1992) Accelerated molecular dynamics simulation with the parallel fast multipole algorithm. Chem Phys Lett 198(1):89–94

Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5(11):789–796

Bonvin A, Brunger AT (1995) Conformational variability of solution nuclear-magnetic-resonance structures. J Mol Biol 250(1):80–93. doi:10.1006/jmbi.1995.0360

Bonvin A, Boelens R, Kaptein R (1994) Time-averaged and ensemble-averaged direct NOE restraints. J Biomol NMR 4(1):143–149

Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K (2014) Combining experiments and simulations using the maximum entropy principle. PLoS Comp Biol 10(2):e1003406

Bottaro S, Lindorff-Larsen K, Best RB (2013) Variational optimization of an all-atom implicit solvent force field to match explicit solvent simulation data. J Chem Theor Comp 9(12):5641–5652

Brooks BR, Bruccoleri RE, Olafson BD et al (1983) CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 4(2):187–217

Brunger AT, Adams PD, Clore GM et al (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 54:905–921

---

[3] http://pedb.vib.be/.

Burgi R, Pitera J, van Gunsteren WF (2001) Assessing the effect of conformational averaging on the measured values of observables. J Biomol NMR 19(4):305–320. doi:10.1023/a:1011295422203

Camilloni C, Vendruscolo M (2014) Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. J Am Chem Soc 136:8982–8991

Camilloni C, Cavalli A, Vendruscolo M (2013) Replica-averaged metadynamics. J Chem Theor Comp 9(12):5610–5617

Cavalli A, Camilloni C, Vendruscolo M (2013) Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. J Chem Phys 138(9):094112

Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, New York

Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 308(5):1011–1032

Clore GM, Schwieters CD (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc 126(9):2923–2938

Constantine KL, Mueller L, Andersen NH et al (1995) Structural and dynamic properties of a beta-hairpin-forming linear peptide.1. Modeling using ensemble-averaged constraints. J Am Chem Soc 117(44):10841–10854. doi:10.1021/ja00149a007

Das R, Baker D (2008) Macromolecular modeling with Rosetta. Annu Rev Biochem 77:363–382

Dedmon MM, Lindorff-Larsen K, Christodoulou J et al (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127(2):476–477

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6(3):197–208

Fennen J, Torda AE, van Gunsteren WF (1995) Structure refinement with molecular-dynamics and a Boltzmann-weighted ensemble. J Biomol NMR 6(2):163–170

Fersht AR (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W. H. Freeman, New York

Francis CJ, Lindorff-Larsen K, Best RB et al (2006) Characterization of the residual structure in the unfolded state of the $\Delta131\Delta$ fragment of staphylococcal nuclease. Proteins 65(1):145–152

Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254(5038):1598–1603

Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. 2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 268(1):170–184

Grishaev A, Bax A (2004) An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. J Am Chem Soc 126(23):7281–7292

Gsponer J, Hopearuoho H, Whittaker SBM et al (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. Proc Natl Acad Sci U S A 103(1):99–104

Haas E (2005) The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. ChemPhysChem 6(5):858–870

Heise H, Luca S, de Groot BL et al (2005) Probing conformational disorder in neurotensin by two-dimensional solid-state NMR and comparison to molecular dynamics simulations. Bioph J 89(3):2113–2120

Hornak V, Abel R, Okur A et al (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins 65(3):712–725

Hub JS, De Groot BL, Van Der Spoel D (2010) g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. J Chem Theor Comp 6(12):3713–3720

Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci U S A 102(19):6679–6685

Kemmink J, Scheek RM (1995) Dynamic modeling of a helical peptide in solution using NMR data—multiple conformations and multi-spin effects. J Biomol NMR 6(1):33–40. doi:10.1007/bf00417489

Kessler H, Griesinger C, Lautz J et al (1988) Conformational dynamics detected by nuclear magnetic-resonance NOE values and J-coupling constants. J Am Chem Soc 110(11):3393–3396. doi:10.1021/ja00219a008

Klein-Seetharaman J, Oikawa M, Grimshaw SB et al (2002) Long-range interactions within a nonnative protein. Science 295(5560):1719–1722

Knott M, Best RB (2012) A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. PLoS Comp Biol 8(7):e1002605

Knowles TP, Vendruscolo M, Dobson CM (2014) The amyloid state and its association with protein misfolding diseases. Nat Rev Mol Cell Biol 15(6):384–396

Korzhnev DM, Salvatella X, Vendruscolo M et al (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. Nature 430(6999):586–590

Krzeminski M, Marsh JA, Neale C et al (2013) Characterization of disordered proteins with ensemble. Bioinformatics 29(3):398–399

Kumar S, Rosenberg JM, Bouzida D et al (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comp Chem 13(8):1011–1021

Laio A, Gervasio FL (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys 71(12):126601

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci U S A 99(20):12562–12566

Lange OF, Lakomek N-A, Farès C et al (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320(5882):1471–1475

Lindorff-Larsen K, Kristjansdottir S, Teilum K et al (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. J Am Chem Soc 126(10):3291–3299

Lindorff-Larsen K, Best RB, DePristo MA et al (2005) Simultaneous determination of protein structure and dynamics. Nature 433(7022):128–132

Lindorff-Larsen K, Maragakis P, Piana S et al (2012a) Systematic validation of protein force fields against experimental data. PLoS ONE 7(2):e32131

Lindorff-Larsen K, Trbovic N, Maragakis P et al (2012b) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. J Am Chem Soc 134(8):3787–3791

Markwick PR, Bouvignies G, Blackledge M (2007) Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. J Am Chem Soc 129(15):4724–4730

Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. Science 312(5771):224–228

Moglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. Proc Natl Acad Sci U S A 103(33):12394–12399

Monticelli L, Kandasamy SK, Periole X et al (2008) The Martini coarse-grained force field: extension to proteins. J Chem Theor Comp 4(5):819–834

Moult J, Fidelis K, Kryshtafovych A et al (2014) Critical assessment of methods of protein structure prediction (CASP)—round X. Proteins 82(S2):1–6

Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Op Struct Biol 24:98–105

Pitera JW, Chodera JD (2012) On the use of experimental observations to bias simulated ensembles. J Chem Theor Comp 8(10):3445–3451

Rosato A, Bagaria A, Baker D et al (2009) CASD-NMR: critical assessment of automated structure determination by NMR. Nat Methods 6(9):625–626

Rosato A, Aramini JM, Arrowsmith C et al (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20(2):227–236

Roux B, Weare J (2013) On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. J Chem Phys 138(8):084107

Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419(6908):743–747

Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. Prog Nucl Mag Res Spectrosc 48(1):47–62

Shaw DE, Maragakis P, Lindorff-Larsen K et al (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330(6002):341–346

Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci U S A 103(31):11539–11543

Smith LJ, Bolin KA, Schwalbe H et al (1996) Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. J Mol Biol 255(3):494–506

Spronk C, Nabuurs SB, Krieger E et al (2004) Validation of protein structures derived by NMR spectroscopy. Prog Nucl Mag Res Spectrosc 45(3–4):315–337

Torda AE, Scheek RM, van Gunsteren WF (1989) Time-dependent distance restraints in molecular-dynamics simulations. Chem Phys Lett 157(4):289–294. doi:10.1016/0009-2614(89)87249-5

Tozzini V (2005) Coarse-grained models for proteins. Curr Op Struct Biol 15(2):144–150

Uversky VN (2013) A decade and a half of protein intrinsic disorder: Biology still waits for Physics. Protein Sci 22(6):693–724

van Kampen NG (1992) Stochastic processes in physics and chemistry. North-Holland, Amsterdam, New York

Varadi M, Kosol S, Lebrun P et al (2014) pE-DB: A database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucl Acids Res 42(D1):D326–D335

Vendruscolo M, Dobson CM (2006) Dynamic visions of enzymatic reactions. Science 313(5793):1586–1587

Vendruscolo M (2007) Structure determination of highly heterogenous states of proteins. Curr Op Struct Biol 17:15–20

Vendruscolo M, Dobson CM (2011) Protein dynamics: Moore's law in molecular biology. Curr Biol 21(2):R68–R70

Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York

Zhu F, Hummer G (2012) Convergence and error estimation in free energy calculations using the weighted histogram analysis method. J Comp Chem 33(4):453–465