Chapter 3
Understanding Protein Dynamics Using Conformational Ensembles

X. Salvatella

Abstract Conformational ensembles are powerful tools to represent the range of conformations that can be sampled by proteins. They can be generated by using purely theoretical methods or, as is most often the case, by fitting ensembles of conformations to experimental data that report on the amplitude of protein dynamics. Conformational ensembles have been useful instruments to study fundamental properties of proteins such as the mechanism of molecular recognition, the early stages of protein folding and the mechanism by which structural information propagates through the structures of globular proteins structures *via* correlated backbone motions. In this chapter I will review the various approaches that have been put forward in the literature to generate conformation ensembles for proteins and present a selection of examples of how such representations of the structural heterogeneity of proteins have been used to explore the fundamental properties of these macromolecules. Finally, I will look ahead at likely future developments in this area, which is important for structural and chemical biology as well as for biophysics.

Keywords Conformational ensembles • Nuclear magnetic resonance • Correlated motions • Conformational selection • Induced fit • Allostery

3.1 Protein Dynamics

The structures of proteins fluctuate in various timescales and with various amplitudes [1]. Since these fluctuations play important roles in biological function it is desirable to complement the structural information contained in protein structures

X. Salvatella (✉)
Joint BSC-IRB Research Programme in Computational Biology,
Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain
e-mail: xavier.salvatella@irbbarcelona.org

K. Han et al. (eds.), *Protein Conformational Dynamics*, Advances in Experimental Medicine and Biology 805, DOI 10.1007/978-3-319-02970-2_3,

with an account of how these fluctuate [2]. Various methodological approaches have been put forward to reach this goal, ranging from purely theoretical methods that predict the fate of protein structures from first principles [3] to experimental methods that provide very detailed equilibrium distributions of well-defined structural properties such as specific inter-atomic distances.

Protein conformational ensembles aim at representing the range of conformations that a given protein samples at equilibrium [4]. Several methods have been put forward for their generation and most of these rely on the fitting of experimental data sensitive to structural fluctuations to ensembles of conformations that model the structural heterogeneity of proteins [2]. It is important to state early on that conformational ensembles do not in general report on protein dynamics in the sense that they do not provide information about the rate of inter-conversion between conformers. They can however report on the amplitude of the dynamics and this property has found wide use in the analysis of the behavior of proteins [4].

3.2 Generating Conformational Ensembles

As previously mentioned a wide range of tools is available for the generation of conformational ensembles. Although they vary quite widely in how the ensembles are built they share one important feature, which is that they extract information on the amplitude of protein dynamics from experimental data sensitive to structural fluctuations. I will now describe the different conceptual approaches that have been used in the field to generate such representations of the structural heterogeneity of proteins.

3.2.1 Using Molecular Dynamics and Advanced Sampling Methods

Molecular dynamics (MD) [3] is a simulation technique that can in principle provide an extremely detailed description of protein dynamics. It is based on modeling interatomic interactions by using empiric potentials called force fields and in the prediction of the time evolution of experimental structures by integration of Newton's equations of motion. The accuracy of these trajectories relies of course on the quality of such force fields and on the ability of computer hardware to simulate biologically relevant timescales, which is still a challenge, especially for large proteins and multi-protein complexes.

Although the technique was developed long ago it has experienced an extraordinary surge in recent years (Fig. 3.1) thanks to the availability of hardware designed specifically to carry out MD simulations, such as the supercomputer Anton [9], built by D. E. Shaw Research, and hardware designed to carry out other tasks but

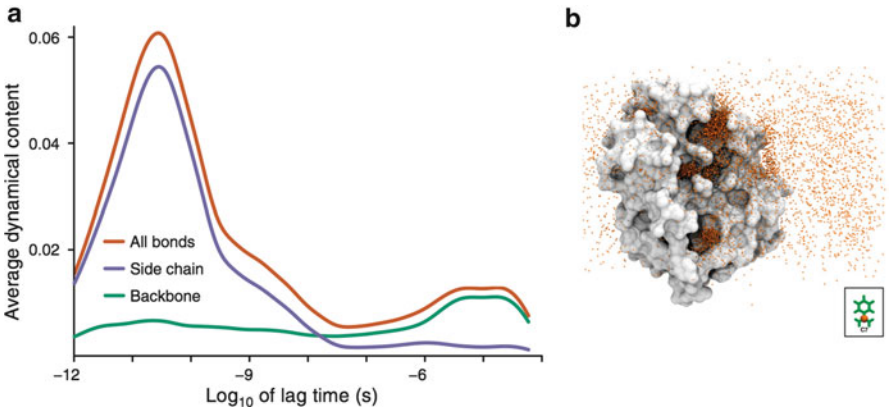

Fig. 3.1 MD is a very powerful simulation tool to characterize the structural heterogeneity and the dynamics of macromolecules [5] as well as their interactions will small molecules [6, 7]. (**a**) Analysis of a 1 ms simulation of BPTI run using the supercomputer Anton built by D. E. Shaw research [5] which illustrates how the dynamics of side chains are in general much faster than those of the backbone (**b**) snapshots of a simulation of the binding of benzamidine to trypsin carried out by using the GPUGRID distributed computing network [8], where the *orange dots* represent the various positions adopted by the C7 atom of benzamidine, shown in the *inset*, illustrating that the small molecule can bind to several sites on the surface of the enzyme before forming a stable complex

that performs MD particularly efficiently, such as graphics processing units (GPUs) [8, 10]. Mainly thanks to these improvements in hardware it has recently become possible to produce trajectories that are sufficiently long to sample the averaging time of experimental observables [11]. This is an important development because it allows the direct comparison of the simulated trajectories with experimental data, which is necessary for iteratively improving force field parameters to render the simulated behavior increasingly realistic.

One illustrative example of how increases in simulation time can lead to improvements in the quality of force fields is provided by a recent report of the D. E. Shaw Research team [12]. To improve the force field parameters that describe the conformational preferences of amino acid side chains the authors first compared the rotamer distributions obtained in very long (720 ns) MD simulations of model α-helices using a state of the art force field (Amber ff99SB) [13] with those derived from a statistical analysis of experimental structures deposited in the protein data bank (PDB).

After identifying four side amino acid types for which the agreement was poor (Ile, Leu, Asp, Asn) they optimized the force field parameters that govern the conformational properties of their side chains against quantum chemical calculations. Finally, and crucially, they validated the force field thus obtained (Amber ff99SB – ILDN) by predicting the NMR parameters (scalar and dipolar couplings, see Table 3.1) of a number of globular proteins that have been well studied using

this technique such as hen egg white lysozyme, bovine pancreatic trypsin inhibitor, ubiquitin, and the B3 domain of Protein G. That the authors were able to generate trajectories with a length (1.2 μs) that matches the averaging time of the NMR parameters was key for proving that the new force field parameters, that provide a better validation that the old ones, represent a substantial improvement [12].

In cases where the size of the system and the timescale of the dynamics of interest allow investigation by MD this is undoubtedly the most informative technique that is currently available for characterizing the fluctuations of the structure of proteins. Even in cases where MD can be used it is nevertheless necessary to validate the resulting trajectories either by predicting experimental data sensitive to dynamics such as nuclear magnetic resonance (NMR) parameters (see below) or by predicting the outcome of perturbations of the system such as point mutations. Only in cases when these validations are successful is it advisable to consider the trajectories provided by MD a realistic model of the behavior of the protein [14].

An illustrative example of how it is possible to use MD trajectories validated by experiments to analyze very subtle but important dynamical properties of proteins is provided by the work of the Bruschweiler group on the protein ubiquitin. Ubiquitin is small protein that is used as a model system for this type of studies because it is of a size that allows the simulation of relative long timescales, is stable in most force fields and because its spectroscopic properties render its characterization using NMR relatively straightforward.

In an important study published in 2007 Showalter and Bruschweiler showed that simulating the dynamics of this protein using MD and a state of the art force field (Amber ff99SB) for 50 ns lead to a trajectory that agreed with NMR data sensitive to dynamics (RDCs) better than the X-ray structure (1UBQ) and only slightly worse than the NMR structure (1D3Z) refined against the NMR data used for validation [15]. These results indicate that, at least for ubiquitin, proper consideration of the contribution of motional averaging to the measured NMR data by using MD can lead to much improved representations of the structural properties of proteins (Fig. 3.2).

Armed with this validation Bruschweiler and co-workers analyzed the degree of correlation of the motions of the backbone torsions in this protein [16]. They recently found that there is a weak but certain degree of correlation of the motions of the ϕ and ψ torsion angles of residues facing one another across the β-sheet of ubiquitin, especially when they are hydrogen bonded, but that this decays very quickly as the distance between two given residues increases. It is important to emphasize that this type of analysis, that relies on a very accurate representation of the dynamics of the protein, is warranted due to the notable ability of the trajectory obtained by these researchers to validate against experiments [15].

Although MD has made spectacular progress in the last few years there are still many biological processes that cannot yet be routinely simulated using this technique. Processes that fall under this category include those involving intrinsically disordered proteins (IDPs), that have important biological functions but that fail to fold into conventional structures that can be characterized using the tools of structural biology such as X-ray crystallography, conventional NMR and cryo-

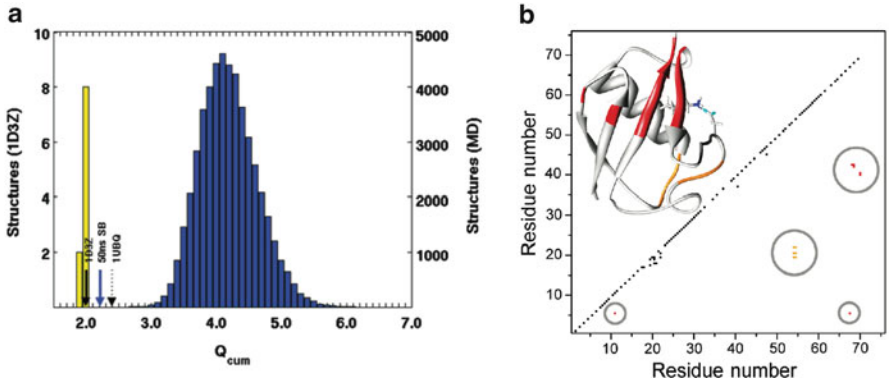

Fig. 3.2 MD trajectories validated against experimental data that reports on the amplitude of protein dynamics can be used to understand fundamental properties of proteins such as the presence of correlated backbone motions. (**a**) Histogram of the quality factors, with $Q = \text{rms}(D^{exp} - D^{calc})/\text{rms}(D^{exp})$, of the various conformations sampled during a 50 ns trajectory of ubiquitin (in *blue*) and of the conformations of the conventional NMR ensemble (in *yellow*, pdb code 1D3Z); quality factors of the various structural representations (with 1UBQ representing the highest resolution X-structure) (**b**) matrix representation of the pairwise correlation coefficient between backbone torsion angles of ubiquitin with R^2 larger than 0.1, which shows the presence of correlated motions across the β-sheet of the protein as indicated in *red* in the structure as well as in other hydrogen bonded residues

electron microscopy [17]. These proteins present an extreme degree of structural heterogeneity and play important roles in molecular mechanisms of enormous importance for biology, such as transcription [18, 19], and biomedicine [20].

Due to the challenges involved in sampling the particularly vast conformational space explored by IDPs and to the possibility that current force fields are not optimized to accurately describe the weak inter-atomic interactions that dominate the behavior of this type of proteins the use of conventional MD to study such systems is still in its infancy [21]. For this reason there is substantial interest in the development of approaches that allow determining conformational ensembles for IDPs by combining molecular simulations with the information contained in experimental observables reporting on the amplitude of protein dynamics such as SAXS and NMR [22] (Sect. 3.2.2).

In cases where the size of the protein is too large, the conformational space too vast [23] and the dynamics often too slow to be sampled by conventional MD it is possible to use advanced sampling methods such as replica exchange MD [24], umbrella sampling [25], accelerated MD [26], local elevation [27] and metadynamics [28], among others available, to explore the conformational space sampled by proteins. In these methods various strategies are used to overcome the free energy barriers present in the energy landscape, that prevent efficient sampling in conventional MD. A detailed theoretical and technical description of these techniques is beyond the scope of this chapter and can be found elsewhere [29, 30]. Advanced sampling tools have been important to show that appropriate

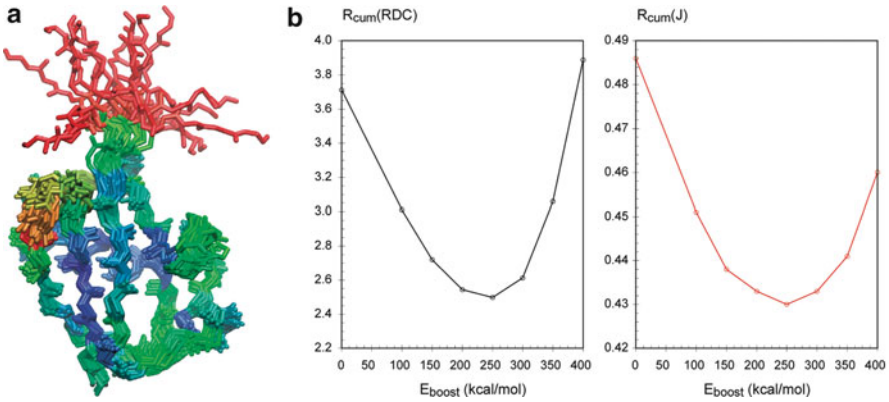

Fig. 3.3 Advanced sampling methods can lead to accurate representations of the structural heterogeneity of proteins. (**a**) Ubiquitin ensemble obtained by the McCammon and Blackledge groups by AMD with a degree of boost optimized by validation against NMR data [31], where the residues are colored according to their flexibility (*blue*: rigid, *red*: flexible). (**b**) Level of agreement of AMD trajectories, expressed as $R_{cum} = Q_{cum}/\sqrt{2}$, obtained with increasing degrees of boost with an indication that both the scalar and the dipolar couplings, that average on the same timescale, validate best when $E_{boost} = 250$ kcal.mol^{-1}, strongly indicating that this level of boost allows sampling all conformations that contribute to the average, experimental NMR parameter [31]

consideration of the averaging implicit in the measurement of nuclear magnetic resonance (NMR) parameters can lead to very accurate descriptions of the structural heterogeneity of proteins and peptides [31].

In a particularly powerful illustration of how such methods can be used to report on the amplitude of protein dynamics McCammon, Blackledge and co-workers use accelerated molecular dynamics (AMD) [26] to analyze the dynamics of the protein ubiquitin [31]. In AMD sampling is accelerated by adding an energy term to the potential energy, that depends on the difference between the potential energy and a reference energy called the boost energy (E_{boost}), that effectively decreases the height of free energy barriers encountered by the simulation [26]. The correct value of the E_{boost} that needs to be used to reach a particular timescale is however, in principle, not known *a priori*.

To generate a trajectory describing all conformations contributing to the average NMR parameters measured experimentally for the small model protein ubiquitin (scalar and residual dipolar couplings, that average in the ms timescale) these authors carried out AMD simulations of this protein with increasing values of E_{boost}. An analysis of how well the simulated trajectories agreed with the NMR experiments showed that conventional MD simulations did not validate well, that moderate degrees of acceleration lead to very significant improvements in the agreement against experiment and that it was possible to use the experimental data to determine the optimal value of E_{boost}, one that samples all conformations that contribute to the time-averaged NMR parameters (Fig. 3.3).

3.2.2 From Experimental Data

In the methods discussed in the previous section experimental data plays a modest role. It is merely used as a validation tool to reassure the researcher that the trajectory provided by MD or by an advanced sampling method is a realistic representation of the dynamics of the system. There are however scenarios where it is desirable that experimental data plays a much more important role because the conformational space sampled by the system is too vast to be sampled by MD, for example, when there are reasons to think that a particular property of the system may not we well-described by force fields or when, for a variety of possible reasons, it is desirable to minimize the role played by the force field in determining the structural and dynamical properties of the system.

There are a number of experimental methods that can provide structural and dynamical information to determine conformational ensembles for proteins. These include Förster resonance energy transfers (FRET) measured using fluorescence, that provide information about r^{-6} averaged inter-dye distances [32] and, in single molecule mode, distributions of distances as well as small angle x-ray scattering (SAXS), that provides information about the hydrodynamic properties of proteins [33]. The most powerful experimental method is however undoubtedly NMR [34] because it provides information at atomic resolution, unlike SAXS, and because it does not require, unlike FRET, labeling the protein with fluorescent groups that can alter the structure, the dynamics and the interactions of the protein and therefore require performing extensive experimental controls.

It is possible to measure a number of parameters by NMR (chemical shifts, scalar couplings, nuclear Overhauser effects, residual dipolar couplings and chemical shift anisotropy in aligned samples, cross-correlated relaxation rates) and these can be related to quantities that can be in principle computed from structures, trajectories and ensembles (distances, angles, torsion angles) as shown in Table 3.1. For the purposes of validating and, especially, generating conformational ensembles it is of course important to take into appropriate consideration the range of validity of the equations used to back-calculate NMR parameters, the accuracy with which NMR parameters can be measured experimentally [35] and the accuracy with which they can be back-calculated [36]. Another factor to take into account is the way these equations were parameterized i.e. whether the parameters were determined from first principles or whether they were instead determined by fitting to known crystallographic structures, a procedure that leads to equations that under-estimate the contribution of conformational averaging [37].

3.2.2.1 Selection Methods

Various approaches are available for the generation of conformational ensembles from experimental but I will start will selection methods because they these conceptually related to the use of conventional MD and advanced sampling methods

Table 3.1 NMR parameters that can be used for generating conformational ensembles and their relationship to protein structure

Symbol	Description	Structural interpretation
NOE_{ij}	Nuclear Overhauser effect	Distance between ^{1}H nuclei i and j (<6 Å)
PRE_{ij}	Paramagnetic relaxation enhancement	Distance between an unpaired electron attached to site i and nucleus j (<30 Å)
$^{3}J_{ij}$	Three-bond scalar coupling	Dihedral angle between bond vectors i and j
$^{3h}J_{ij}$	Trans-hydrogen bond scalar coupling	Geometry of hydrogen bond linking heavy atoms i and j
RDC or D_{ij}	Residual dipolar coupling	Angle between bond vectors i and j and the molecular frame defined by macromolecular alignment
CS_i	Chemical shifts	Convolution of a large number of structural properties in the vicinity of nucleus i
S_i^2	Order parameter	Rigidity of bond vector i in the molecular frame defined by macromolecular tumbling ($0 \leq S_i^2 \leq 1$)

described in Sect. 3.2.1. Selection methods use experimental data to, as their name suggests, select conformations from a pre-defined pool of conformations generated *a priori*. The pool is meant to contain all physically possible conformations that the protein can sample in a defined timescale with some probability but that these are not present with their correct statistical weights. Since any experimental (NMR or otherwise) parameter contains information about the distribution of conformations contributing to the average, it can be in principle be used to optimize the statistical weights.

Several algorithms have been used to select the conformations from the pool. These range from Monte Carlo algorithms, such as the ENSEMBLE method developed by the Forman-Kay laboratory to generate ensembles for IDPs [38], to genetic algorithms, such as the OED method of the Svergun laboratory to generate ensembles from SAXS data [39] and the ASTEROIDS method (Fig. 3.4a) developed by the Blackledge laboratory to generate ensembles for IDPs from NMR data [41]. These methods differ in the nature of the pool and in the technical details of the selection method but are all based on the same idea and make the same key assumption, which is that all possible conformations are present in the pool.

Of course whether a selection method performs well depends fundamentally on the properties of the pool. This must be an accurate representation of the range of conformations that can be sampled by the protein and its size must be representative of the size of the conformational space available. If the quality of the pool is low, that is if the conformations present in the pool do not represent the conformations that the protein can adopt, the selection method will generate a conformational

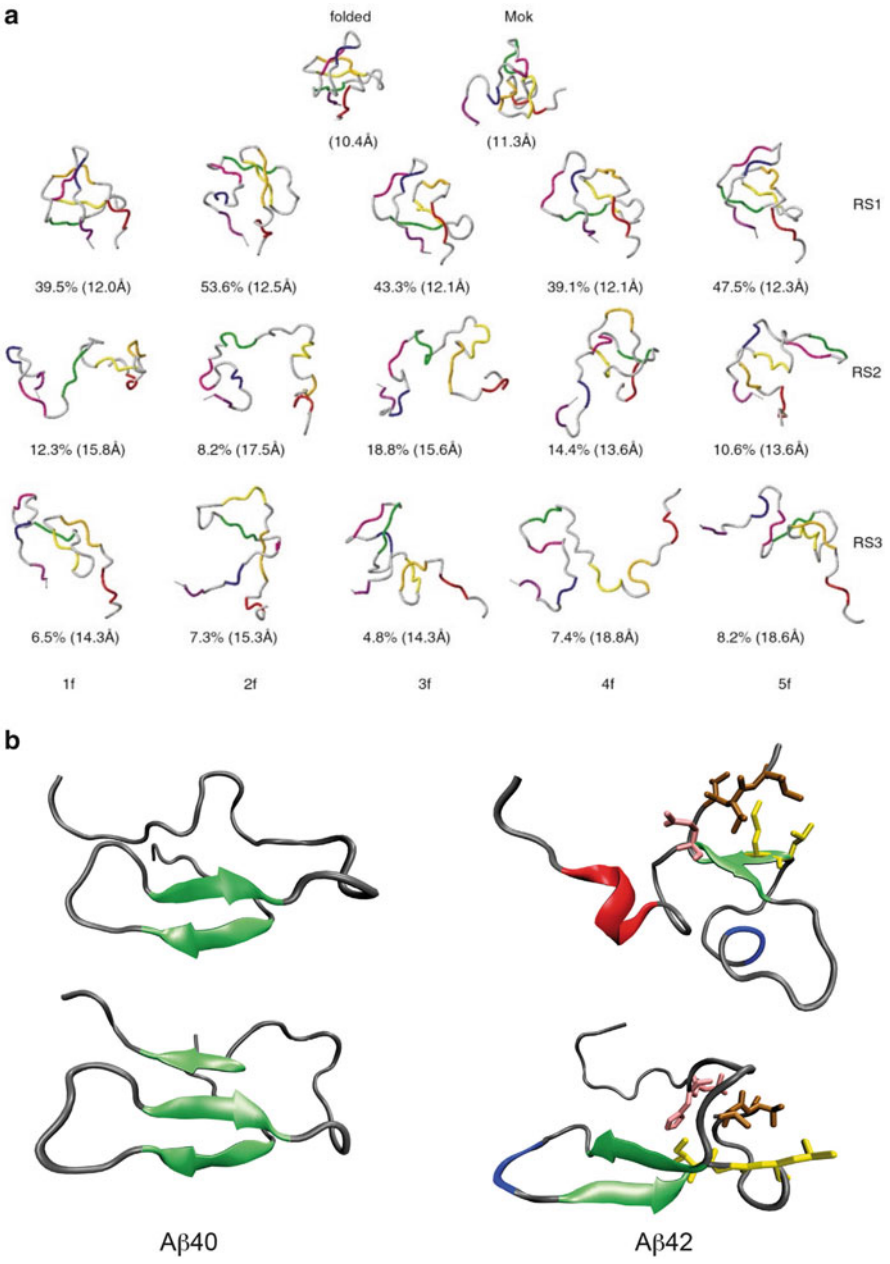

Fig. 3.4 (**a**) Representative structures belonging to the three main clusters (RS1, RS2, RS3) defining the denatured state of a drkN SH3 domain by selecting conformations from a pool generated by a simulation of the thermal denaturation of the native state of this protein [38] (**b**) Comparison of the dominant conformations of Aβ40 and Aβ42 as determined by selecting structures from a pool derived from trajectories of these peptides computed using MD or REMD. It can be observed that the C-terminal residues of Aβ42, that are not present in Aβ40, form long-range transient contacts [40]

ensemble that will be in agreement with experiment but that will be not be a realistic description of the structural properties of the protein. If the pool is of good quality but it contains too few conformations, that is if conformations actually sampled by the protein are absent from the pool, the selection algorithm will also fail to produce a useful result.

From Statistical Coils

As already mentioned selection methods have used quite extensively for the generation of conformational ensembles describing the properties of chemically denatured proteins and IDPs. The pioneering work of Dobson and co-workers [42], recently followed up by the Sosnick [23] and Blackledge [43] groups, showed that it is possible to produce reasonable representations for the range of structures sampled by these systems by generating conformational ensembles where the distributions of backbone torsion angles of the different residues of the protein match those of the same residues in loops and termini of structures deposited in the PDB.

These conformational ensembles aim at representing the structural properties of polypeptides where these are dominated by the local structural preferences [42], that is in the absence of the long-range interactions that play an important role in stabilizing the tertiary structure of globular proteins. These statistical coils have, in spite of the simplicity of the approach used to generate them, structural properties that match those determined experimentally for disordered proteins and that validate reasonably well with SAXS and NMR experimental parameters such as backbone scalar and residual dipolar couplings [23, 42, 43].

There is considerable experimental evidence, mainly from FRET, chemical shifts and paramagnetic relaxation enhancement (PRE) experiments measured using NMR (Table 3.1) that IDPs can form transient long range interactions that are important for their physiological and physiopathological roles [44–47]. Since these long range interactions cannot be definition be described by statistical coils [42] efforts have been made in improving the description of IDPs by using selection methods where statistical coils are used as pools.

From Ensembles Determined Using Simulations

Since the quality of the pool is key for the performance of selection methods efforts have been directed at using molecular simulations for constructing pools that better reproduce the structural properties of IDPs. These were pioneered by the Forman-Kay group, that used thermal unfolding trajectories to generate the pool in a study of the denatured state of an SH3 domain [38] in which they used ^{1}H-^{1}H NOEs, scalar couplings, ^{13}C chemical shifts, among other parameters, to optimize the statistical weights of the conformations of the pool (Fig. 3.4b). This [38], as well as other studies by the same group [48, 49], in which they also explored the use of statistical

coils as pools, indicated that the denatured state is substantially collapsed, with a fraction of the secondary structure of the native state, and that it is stabilized by native as well as non-native long range interactions.

This pioneering work has been followed up by the Head-Gordon group, that has used instead ensembles derived from MD or ERMD trajectories to generate ensembles for the peptides Aβ40 and Aβ42 [40]. Although these two peptides have very similar sequences the latter is much more prone than the former to form neurotoxic oligomeric species thought to trigger Alzheimer's disease [50, 51]. There has been much discussion about how can the addition of only two residues have such a profound effect on the structural properties of an IDP. The recent study of the Head-Gordon group, in which the authors used ^{1}H-^{1}H NOEs as well as scalar and residual dipolar couplings to bias a selection algorithm, is an important contribution to this topic. It shows that Aβ42 has a substantially different contact map due to the propensity of the two additional residues to form long-range contacts with hydrophobic residues in the rest of the sequence (Fig. 3.4b).

3.2.2.2 Restrained Simulations

Both selection methods and, especially, MD rely heavily on an accurate description of the conformational space available to proteins either by using motional models, such as statistical coils, or molecular simulations force fields. These approaches are therefore unsuitable when these descriptions are not available or when it is thought that the conformations that they provide are not correct. In these cases it is possible to carry out restrained molecular simulations in which empirical potentials are added to the potential energy of the protein provided by the force field to penalize configurations with back-calculated experimental parameters that are in disagreement with those measured experimentally.

Since these methods bias the sampling they have the potential to generate conformations that would otherwise not be sampled in an unrestrained MD simulation *i.e.* they use the experimental data as a protein-specific force field correction [52]. Restrained simulations have of course a long history in structure determination and in fact lie at the heart of the ability of NMR to produce average structures for proteins by generating configurations with structural properties (bond lengths, angles, inter-atomic distances, etc.) that do not deviate too much from those considered optimal for molecular simulation force fields and are in addition in agreement with NMR parameters reporting on protein structure (Table 3.1) [53, 54].

Given that the NMR parameters cannot be measured or back-calculated from structures with infinite accuracy they usually do not define a unique conformation and it is customary to represent NMR structures as ensembles of conformations that fit the NMR data. The spread of these ensembles depends on the ability of the experimental data to define the average structure and can be considered equivalent to the resolution of crystallographic structures, with significant heterogeneity reflecting poor resolution. Even though the presence of significant dynamics can

lead, like in X-ray crystallography, to poor resolution it is important to emphasize that the spread of conventional NMR ensembles does not directly report on protein dynamics [52].

For NMR ensembles to report on protein dynamics it is necessary to use restrained simulation protocols that fit the NMR parameters to an ensemble of conformations rather to a single average conformation [2, 55] In these protocols an energy penalty is applied when the ensemble–averaged back-calculated NMR parameters are in disagreement with experiment.

Ensemble Averaged Restrained Simulations

In ensemble-averaged restrained simulations an ensemble of conformations of the protein of interest is simulated simultaneously and the ensemble-averaged back-calculated NMR parameters are restrained by an empirical quadratic potential to be in agreement with the experimental value [2]. When there is a violation the energy penalty generates a force in all conformations that contribute to the average so that the NMR parameter is fulfilled. This leads to conformational ensembles where individual conformations may have NMR parameters that deviate from experiment but where the ensemble collectively does not and is, therefore, a representation of the range of conformations that is sampled by the protein at equilibrium.

As early as the 1990s there was a general awareness of the importance of conformational averaging in protein structure determination by NMR and, as a consequence, attempts at using this type of simulation protocols to generate conformational ensembles from the NMR parameters commonly used for determining structures [55, 56]. This was however a challenging task because NOEs, the main source of structural information available at the time, are not suitable restraints for ensemble simulations because they average non-linearly [57] and could only be measured semi-quantitatively due to the low signal to noise ratio of NMR and the presence of spin-diffusion [58, 59]. As a consequence it was not possible to cross-validate the resulting conformational ensembles, which were significantly under-restrained and therefore presented artifactual structural heterogeneity. It is worth mentioning that the measurement of very exact NOEs in per-deuterated proteins is to a large extent alleviating the problems of this NMR observable as restraint in ensemble simulations as illustrated by recent work of Riek and co-workers [58, 59].

This situation, however, changed quite significantly when Tjandra and Bax showed that it was possible to induce a small degree of anisotropy in the rotational diffusion on protein samples by using an external alignment medium [62]. This lead to the possibility of measuring residual dipolar couplings (RDCs) for pairs of nuclei, which otherwise average to zero when the inter-nuclear vector rotates isotropically around the magnetic field of the NMR apparatus. For proteins which do not experience important changes in alignment when their structures fluctuate [63] the value of the RDC of a given conformation is given by the degree of alignment, which depends on the overall shape of the protein, and on the orientation

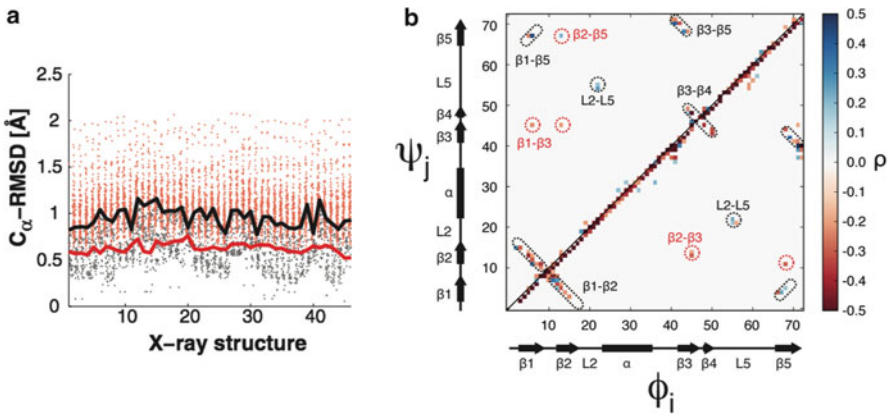

Fig. 3.5 Restrained ensemble simulations can be used to generate detailed descriptions of the structural heterogeneity of proteins that can be of great use to understand fundamental properties of their structures (**a**) A conformational ensemble determined for the protein ubiquitin using RDCs in its free state contains the structures that the protein adopts upon binding other proteins, indicating that the molecular recognition of this protein occurs by conformational selection. Plot of the Calpha-RMSD between each bound structure of ubiquitin (*x-axis*) and the ensemble members of free (*red dots*) and bound (*black dots*) ubiquitin [60] (**b**) Using a similar approach we showed that the motions of residues in the β-sheet of this protein are correlated, which suggests that the collective motions that these correlations underlie could play role in molecular recognition. The matrix represents the circular correlation coefficient between two backbone torsion angles of ubiquitin [61]

of the inter-nuclear vector in the molecular structure [62]. Unlike NOEs RDCs can be measured with quite high accuracy and, in addition, average linearly; they are therefore much better suited to ensemble simulation protocols that aim at generating conformational ensembles and have been used quite extensively in the last few years [60, 61, 64, 65].

Parallel to these developments methods were also set up to use S^2 order parameters, that are derived from ^{15}N relaxation rates and that report on the amplitude of the motions of individual backbone NH bond vectors, as restraints for ensemble restrained simulations [66]. Vendruscolo, Dobson an co-workers showed that using these in combination with NOEs allowed the generation of conformational ensembles that validated against ensemble-averaged experimental RDCs better than single structures [67]. In further developments Vendruscolo and co-workers showed how implementing changes to the simulation protocol, particularly to the way in which the averaging of the various observables was carried out, lead to very substantial increases in accuracy of the ensembles [68].

The use of RDCs as restraints in ensemble simulations was pioneered by Clore and co-workers, which setup up the basic simulation protocol and put forward an ingenious strategy to simultaneously fit the coordinates of the ensemble members with the 5 independent elements of the alignment tensor [64]. This mathematical object describes the degree and direction of alignment or the protein in the molecular frame and can be represented by a symmetric traceless matrix that is in generally

not known but can be obtained by single value decomposition from a set of more than 5 accurately measured RDCs if a reasonable structural model for the protein is available [69]. The setup was initially implemented to explore whether restrained ensemble simulations lead to ensembles that agreed with alternative ways of treating the information about dynamics contained in RDCs [64] but has found use in the analysis of protein dynamics since then [61, 65].

In a recent example we used the simulation setup proposed by Clore and co-workers to analyze the dynamics of ubiquitin from a very large set of RDCs in collaboration with Griesinger and co-workers [61]. One important property of the ensemble that we generated is that it is in very good agreement with NMR parameters that we did not use to restrain the ensemble simulation such as trans-hydrogen bond scalar couplings and cross-correlated relaxation rates. An analysis of the correlated motions present in this ensemble lead to the observation of weak but statistically significant correlated motions that connect the dynamics of residues that can be quite far apart in the structure of ubiquitin.

As previously mentioned the degree and directions of alignment of a protein structure expressed in the alignment tensor are generally not known and need to be fit to the experimental RDCs [69]. This is an important drawback of using RDCs for characterizing the structural heterogeneity of proteins because it decreases the information content of these NMR parameters and because it narrows the range of systems that can be studied to those for which the alignment is assumed not to change significantly during the dynamics. One possible way to alleviate this problem is to calculate the alignment tensor of the various conformations that contribute to the average RDC *on the fly* [70, 71], which is possible for mechanisms of external that are well understood such as steric [69] and electrostatic [72] alignment. This new approach to determining conformational ensembles is still under development [73, 74] but it is likely to be an important development because it will allow the characterization of the structural heterogeneity of proteins that experience large shape changes such as intrinsically disordered and multi-domain proteins.

In parallel to these developments in the characterization of the structural heterogeneity of globular proteins restrained ensemble simulations have been used extensively to generate conformational ensembles for chemically denatured and intrinsically disordered proteins. The main source of structural information for these proteins are paramagnetic relaxation enhancements (PREs). PREs are increases in the relaxation of NMR resonances caused by transient interactions of the corresponding nuclei with paramagnetic functional groups introduced by using protein engineering and can be used to probe long-range (up to 25 Å) transient interactions [77]. PREs can be used as restraints in ensemble simulations of disordered proteins [78] but suffer from the same averaging problems of NOEs *i.e.* they do not average linearly.

In order to clarify to what extent PREs are useful restraints for ensemble simulations of disordered proteins our laboratory recently carried out a detailed characterization of their information content. The conclusion that we reached is that PREs are indeed very useful probes of transient long-range interactions

but that their averaging properties render them quite inadequate restraints for ensemble simulations because their average is to a large insensitive to the shape of the distribution of distances. We find that ensemble-averaging does not provide significant advantages for obtaining an accurate characterization of transient long-range interactions and that fitting the PREs to a small number of structures, that can be as small as one, provides the most accurate map for a disordered protein [79].

Time Averaged Restrained Simulations

An alternative to ensemble restrained simulations is time averaged restrained simulations. In this approach, developed by Van Gunsteren and co-workers, a single conformation of the protein is simulated and a quadratic empirical potential ensures that the time-averaged value of a given NMR parameter is equivalent to its experimental counterpart [75]. The key parameter of this simulation protocol is the averaging time *i.e.* the time after which the trajectory is expected to satisfy the experimental values, which needs to be determined *a priori*. Although this approach was an important conceptual development when it was proposed [75, 76] it seems that carrying our ensemble-averaged restrained simulations is a more common approach to the problem of determining conformational ensembles from NMR data.

3.3 Looking Ahead

Conformational ensembles represent an exciting new development in biophysics because they allow for an explicit represent of the dynamics of proteins. Although they do not contain information about the timescale of the dynamics these ensembles provide quite accurate representations of the amplitude of the motions. It is however the case that the determination of such ensembles from experiment is not a routine endeavor because it requires the measurement of a substantial number of NMR parameters such as RDCs. From this point of view it seems that an important priority should be to extract as much information about the amplitude of dynamics from NMR chemical shifts because this NMR parameter is easy to measure. It is therefore likely that we will see developments in this area soon [80].

References

1. Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci USA 102:6679–6685
2. Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 17:15–20
3. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267: 585–590

4. Fenwick RB, Esteban-Martín S, Salvatella X (2011) Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. Eur Biophys J 40:1339–1355
5. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO et al (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330:341–346
6. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA et al (2011) How does a drug molecule find its target binding site? J Am Chem Soc 133:9181–9183
7. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc Natl Acad Sci USA 108: 10184–10189
8. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. J Chem Inf Model 50:397–403
9. Shaw DE, Chao JC, Eastwood MP, Gagliardo J, Grossman JP et al (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51:91
10. Harvey MJ, Giupponi G, Fabritiis GD (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput 5:1632–1639
11. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL et al (2008) Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. J Phys Chem B 112:6155–6158
12. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958
13. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A et al (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
14. van Gunsteren WF, Dolenc J, Mark A (2008) Molecular simulation as an aid to experimentalists. Curr Opin Struct Biol 18:149–153
15. Showalter SA, Brüschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. J Am Chem Soc 129:4158–4159
16. Li D-W, Meng D, Brüschweiler R (2009) Short-range coherence of internal protein dynamics revealed by high-precision in silico study. J Am Chem Soc 131:14610–14611
17. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Curr Opin Struct Biol 18:756–764
18. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC et al (2008) Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol 4:728–737
19. Garcia-Pino A, Balasubramanian S, Wyns L, Gazit E, De Greve H et al (2010) Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell 142: 101–111
20. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573–584
21. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE (2012) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. J Am Chem Soc 134:3787–3791
22. Jensen MR, Ruigrok RW, Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 23:426–435
23. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci USA 102:13099–13104
24. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151
25. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. J Comput Phys 23:187–199
26. Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 120:11919
27. Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Des 8:695–708

28. Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99: 12562–12566
29. Berne BJ, Straub JE (1997) Novel methods of sampling phase space in the simulation of biological systems. Curr Opin Struct Biol 7:181–189
30. Zuckerman DM (2011) Equilibrium sampling in biomolecular simulations. Annu Rev Biophys 40:41–62
31. Markwick PRL, Bouvignies G, Salmon L, McCammon JA, Nilges M et al (2009) Toward a unified representation of protein structural dynamics in solution. J Am Chem Soc 131: 16968–16975
32. Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CAM (2010) Accurate single-molecule FRET studies using multiparameter fluorescence detection. Methods Enzymol 475:455–514
33. Svergun D, Barberato C, Malfois M, Volkov V, Konarev P et al (1995) Evaluation of the solution scattering from macromolecules with known atomic structure and fitting to experimental data. J Appl Crystallogr 28:768–773
34. Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622
35. Fenwick RB, Esteban-Martín S, Salvatella X (2010) Influence of experimental uncertainties on the properties of ensembles derived from NMR residual dipolar couplings. J Phys Chem Lett 1:3438–3441
36. Robustelli P, Cavalli A, Dobson CM, Vendruscolo M, Salvatella X (2009) Folding of small proteins by Monte Carlo simulations with chemical shift restraints without the use of molecular fragment replacement or structural homology. J Phys Chem B 113:7890–7896
37. Case DA (2013) Chemical shifts in biomolecules. Curr Opin Struct Biol 23:172–176
38. Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 308:1011–1032
39. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664
40. Ball KA, Phillips AH, Wemmer DE, Head-Gordon T (2013) Differences in β-strand populations of monomeric Aβ40 and Aβ42. Biophys J 104:2714–2724
41. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR et al (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. J Am Chem Soc 131:17908–17918
42. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM 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 255:494–506
43. Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH et al (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102:17002–17007
44. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 338:1015–1026
45. Song J, Guo L-W, Muradov H, Artemyev NO, Ruoho AE et al (2008) Intrinsically disordered gamma-subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure. Proc Natl Acad Sci USA 105:1505–1510
46. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH et al (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol 14:738–745
47. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025
48. Marsh JA, Neale C, Jack FE, Choy W-Y, Lee AY et al (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. J Mol Biol 367:1494–1510
49. Marsh JA, Forman-Kay JD (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. J Mol Biol 391:359–374

50. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366
51. Roychaudhuri R, Yang M, Hoshi MM, Teplow DB (2009) Amyloid beta-protein assembly and Alzheimer disease. J Biol Chem 284:4749–4753
52. Esteban-Martín S, Bryn Fenwick R, Salvatella X (2012) Synergistic use of NMR and MD simulations to study the structural heterogeneity of proteins. WIREs Comput Mol Sci 2:466–478
53. Kaptein R, Zuiderweg ER, Scheek RM, Boelens R, van Gunsteren WF (1985) A protein structure from nuclear magnetic resonance data. lac repressor headpiece. J Mol Biol 182:179–182
54. Wüthrich K (2003) NMR studies of structure and function of biological macromolecules (Nobel Lecture). J Biomol NMR 27:13–39
55. Bonvin AMJJ, Brünger AT (1995) Conformational variability of solution nuclear magnetic resonance structures. J Mol Biol 250:80–93
56. Mierke DF, Kurz M, Kessler H (1994) Peptide flexibility and calculations of an ensemble of molecules. J Am Chem Soc 116:1042–1049
57. Bürgi R, Pitera J, van Gunsteren WF (2001) Assessing the effect of conformational averaging on the measured values of observables. J Biomol NMR 19:305–320
58. Vögeli B, Segawa TF, Leitz D, Sobol A, Choutko A et al (2009) Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. J Am Chem Soc 131:17215–17225
59. Vögeli B, Kazemi S, Güntert P, Riek R (2012) Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. Nat Struct Mol Biol 19:1-53-1057
60. Lange OF, Lakomek N-A, Farès C, Schröder GF, Walter KFA et al (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320:1471–1475
61. Fenwick RB, Esteban-Martín S, Richter B, Lee D, Walter KFA et al (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. J Am Chem Soc 133:10336–10339
62. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114
63. Salvatella X, Richter B, Vendruscolo M (2008) Influence of the fluctuations of the alignment tensor on the analysis of the structure and dynamics of proteins using residual dipolar couplings. J Biomol NMR 40:71–81
64. 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 126:2923–2938
65. Clore GM, Schwieters CD (2004) Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 43:10678–10691
66. Best RB, Vendruscolo M (2004) Determination of protein structures consistent with NMR order parameters. J Am Chem Soc 126:8090–8091
67. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433:128–132
68. Richter B, Gsponer J, Várnai P, Salvatella X, Vendruscolo M (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J Biomol NMR 37:117–135
69. Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein .... J Am Chem Soc 122:3791–3792
70. Huang J-R, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. J Am Chem Soc 132:694–705
71. Esteban-Martín S, Fenwick RB, Salvatella X (2010) Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings. J Am Chem Soc 132:4626–4632
72. Zweckstetter M, Hummer G, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86:3444–3460

73. De Simone A, Montalvao RW, Vendruscolo M (2011) Determination of conformational equilibria in proteins using residual dipolar couplings. J Chem Theory Comput 7:4189–4195
74. Montalvao RW, De Simone A, Vendruscolo M (2012) Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. J Biomol NMR 53:281–292
75. Torda AE, Scheek RM, van Gunsteren WF (1989) Time-dependent distance restraints in molecular dynamics simulations. Chem Phys Lett 157:289–294
76. Torda AE, Scheek RM, van Gunsteren WF (1990) Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. J Mol Biol 214:223–235
77. Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109:4108–4139
78. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127:476–477
79. Silvestre-Ryan J, Bertoncini CW, Fenwick RB, Esteban-Martín S, Salvatella X (2013) Average conformations determined from PRE data provide high-resolution maps of transient tertiary interactions in disordered proteins. Biophys J 104:1740–1751
80. Camilloni C, Robustelli P, De Simone A, Cavalli A, Vendruscolo M (2012) Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. J Am Chem Soc 134:3968–3971