ELSEVIER

Determination of conformationally heterogeneous states of proteins

Michele Vendruscolo

Although conformationally heterogeneous states of proteins are involved in a range of important biological processes, including protein folding and misfolding, and signal transduction, detailed knowledge of their structure and dynamics is still largely missing. Proteins in many of these states are constantly changing shape, such that they are better described as ensembles of conformations rather than in terms of well-defined structures, as is normally the case for native states. Methods in which molecular simulations are combined with experimental measurements are emerging as a powerful route to the accurate determination of the conformational properties of these states of proteins.

Addresses
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Corresponding author: Vendruscolo, Michele (mv245@cam.ac.uk)

Current Opinion in Structural Biology 2007, **17**:15–20

This review comes from a themed issue on
Folding and binding
Edited by Valerie Daggett and Alan R Fersht

Available online 18th January 2007

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2007.01.002

Introduction

Non-native states of proteins play crucial roles in many aspects of molecular and cell biology. They include those that appear during biosynthesis in the ribosome and degradation in the proteasome, those populated by intrinsically unstructured peptides and proteins, the intermediates and transition states sampled during the folding of globular proteins, and the variety of pathogenic misfolded multimeric species implicated in a range of neurodegenerative and systemic disorders, such as Alzheimer's and Parkinson's disease, and type II diabetes [1,2]. States of this type pose a formidable challenge for structure determination, because, in many cases, they are inherently flexible and conformationally highly heterogeneous.

It has long been recognized that native states also constantly undergo structural fluctuations, and that these dynamics are important for enzymatic catalysis, ligand binding and the formation of biomolecular complexes [3–10]. Such motions are usually represented by structural fluctuations around a well-defined conformation, and powerful techniques are available to calculate these average structures and their related dynamics [11–14]. This description, however, is not suitable in the case of highly heterogeneous states, as the reference structure melts down and the protein populates conformations with very dissimilar structures. These states might be considered to represent the 'liquid phase' of proteins [15].

The information necessary to characterize in detail conformationally heterogeneous states is difficult to acquire experimentally for several reasons. First, the intrinsic dynamics of these states require their definition as ensembles of conformations, thus demanding more information than that required for native states. Second, such dynamics make it challenging to use many of the standard spectroscopic and diffraction techniques available for native states. Third, the translation of experimental measurements into structural information in many cases might only be possible in a semi-quantitative manner.

A view is emerging in which sparse experimental data available for protein structure determination are complemented with the use of *a priori* information [16], for example, about backbone dihedral angles, rotameric states of sidechains, and van der Waals and electrostatic interactions between atoms. This type of information can be provided through the use of force-fields in molecular dynamics [7,16,17] or through effective potentials derived from protein structure databases [18•]. An essential aspect of this type of approach, which requires the use in structure determination procedures of information that has not been specifically measured experimentally, is the development of methods to carefully assess the validity of the structural models produced.

Recent advances in the development of methods for generating ensembles of structures representing conformationally heterogeneous states of proteins are reviewed here, with an emphasis on the conceptual problems related to the translation of experimental information into structural models.

Methods of conformational sampling and structure determination

Several approaches have been proposed for characterizing non-native states that differ in the particular way in which experimental measurements are combined with the *a priori* information. The use of molecular dynamics simulations without inclusion of experimental measurements provides a range of opportunities, which have been explored in a series of recent studies that have provided,

for the first time, a description of the structure and dynamics of proteins in their non-native states [17,19–22]. In another type of strategy, the experimental information is used to filter out conformations in disagreement with observations from a previously generated ensemble of conformations [23•,24,25]. The success of this approach relies on the ability of the conformational sampling to explore states that are populated with significant probability by the protein in solution.

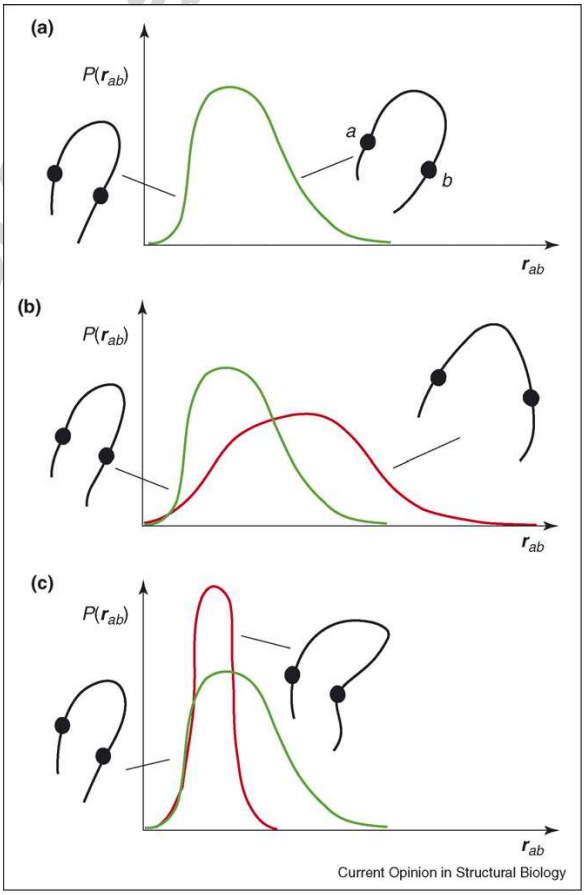
An approach that has been considered more recently involves extending to highly heterogeneous states methods of structure determination that have been developed for native states. In this approach, the experimental information is used to construct structural restraints to be used in molecular simulations. The sampling is biased to take place in regions of conformational space that are consistent with the available experimental information. In favourable cases, interproton distances derived from nuclear Overhauser effects (NOEs) in NMR spectroscopy can be used to define the structures of unfolded states [24,26,27•]. In a series of recent studies, NMR chemical shift information has been used to restrain molecular simulations to obtain structural models of the molten globule of α-lactalbumin at increasing concentrations of urea [28], of two intermediate states of mutational variants of the Fyn SH3 domain [29•] and of a partially unfolded form of the photoactive yellow protein [30•].

From average structures to conformational ensembles

The heterogeneous nature of many of the states populated by proteins makes it necessary to represent them as ensembles of conformations. Such ensembles provide an effective way to represent the fluctuations around an average structure [13]. Even more crucially, they can also represent situations in which an average structure is missing altogether and the protein constantly changes its overall shape [31•,32•]. The absence of a clearly defined reference state, which is, for example, signalled by the presence of multiple rotameric states even in the hydrophobic core of folded proteins [13,33–35], is typical of non-native states of proteins in the presence of a substantial amount of residual structure. NMR spectroscopy is uniquely suitable to monitor this type of behaviour and a variety of techniques have been exploited [2]. The analysis of transverse relaxation rates (R_2) was interpreted in terms of the formation of clusters of transient non-native clusters of hydrophobic residues in hen lysozyme [36]. In a recent study, residual dipolar couplings (RDCs) were used to suggest specific conformational preferences for the natively unfolded form of α-synuclein [23•]. The use of paramagnetic resonance enhancement (PRE) effects in combination with protein mutagenesis has also enabled the characterization of the structural propensities of a series of proteins, including the Δ131Δ fragment of staphylococcal nuclease [37•,38], acyl-coenzyme A binding protein (ACBP) [32•] and α-synuclein [31•,39].

The problem of determining ensembles of structures represents an area of intense research and many conceptual problems remain to be solved to establish computational techniques capable of providing accurate solutions. The most immediate difficulty arises from the availability of only a small number of experimental restraints. This problem, known as overfitting (or under-restraining; Figure 1), has been studied in detail in the case of native states [40,41] and is essentially caused by the lack of sufficient information. The impact of overfitting can be limited, at least in principle, either by developing new experimental techniques or by using

Figure 1



Schematic representation of the probability distribution [$P(r_{ab})$] of the distance (r_{ab}) between two atoms (*a* and *b*) in a highly heterogeneous state of a protein. The true probability distribution **(a)** (green line) is compared with the distributions (red lines) obtained in the presence of **(b)** under-restraining (overfitting) and **(c)** over-restraining (underfitting), which are broader and narrower, respectively.

the information provided by force-field molecular dynamics simulations. By contrast, the incorrect use of the available information may result in the problem of over-restraining (or underfitting) [42••] (Figure 1), which occurs, for instance, when the restraints are imposed too tightly (e.g. on backbone ω angles [43]). This problem becomes especially challenging when NMR techniques are applied to proteins in highly dynamic states, as, in this case, experimental measurements provide time- and ensemble-averaged values for structural observables that should be carefully analysed. If, for example, several average interatomic distances are imposed on a single molecule, the only conformations compatible with this type of restraint might be compact ones (Figure 2). As a consequence of the time and ensemble averaging during the acquisition of NMR spectra, not all the interatomic contacts detected experimentally need to be simultaneously present in any given conformation (Figure 2). 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 [38]. When instead the experimental distances were imposed as averages over many molecules (replicas), 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 [37•]. Approaches in which NMR observables are interpreted in terms of ensemble averages are also useful for determining ensembles of conformations representing the dynamics in the native state. By combining the information provided by NOE effects and NMR S^2 order parameters or by RDCs, it was recently shown that it is possible to simultaneously determine the structure and dynamics of proteins [13,44,45•].

Although the introduction of molecular simulations with ensemble-averaged restraints reduces the problem of over-restraining, it also incurs the dangers of under-restraining (overfitting), because multiplying the number of molecules in the simulation increases the amount of experimental information that is needed for structure determination. A significant advance in this respect was made recently by the introduction of the MUMO (minimal under-restraining minimal over-restraining) procedure, in which both overfitting and underfitting are simultaneously minimized [42••]. The approach is based on the use of different types of ensemble averaging for different observables. In the first application, two replicas for NOEs (as they tend to give rise to the problem of overfitting [40]) and eight replicas for S^2 order parameters (as they tend to give rise to the problem of underfitting [46]) were used to show that native state ensembles can be calculated with very high accuracy [42••].

How can an ensemble of structures be accurately defined?

To define the individual conformation of a protein, one can specify the coordinates of all its atoms (e.g. by analyzing electron density maps), the distances between its atom pairs (e.g. by performing NOE or PRE measurements), or the orientations of interatomic vectors either with respect to each other (e.g. by considering J-couplings or chemical shifts) or relative to an external direction (e.g. by using RDCs). In order to define a heterogeneous conformational ensemble, one should, in principle,

Figure 2

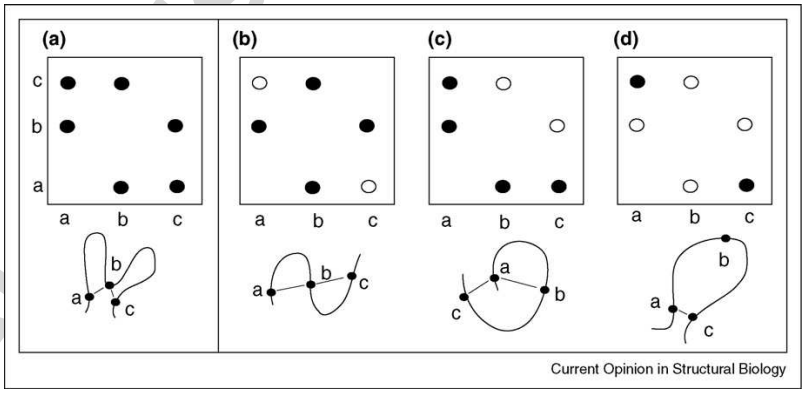


Illustration of the problem of over-restraining (underfitting). NMR measurements of the compact state represented in **(a)** result in distances similar to those observed for the equilibrium ensemble of more expanded conformations outlined in **(b–d)**. This situation arises because averaging with the inverse sixth power of the distance, as is the case, for example, in PRE measurements [38], implies that, most of the time, atom pairs are at distances greater than the average values provided by the experiments. In the example shown, this type of measurement will identify atom pair b–c as being in contact, although these two atoms are actually in contact (closed circle) only in conformation (b), but not (open circles) in conformations (c) and (d). The use of structural restraints imposed on the average values of observables taken over multiple copies (replicas) of the protein molecule in simulations is designed to alleviate the problem of over-restraining (underfitting) [42••].

specify all the conformations in the ensemble together with their statistical weights. In practice, one possibility is to characterise the ensemble by providing the distance distributions for many of its pairs of atoms. This option appears attractive, as the use of single-molecule and fluorescence resonance energy transfer (FRET) methods is proving capable of providing distance distributions [47–50]. In addition, preliminary studies suggest that such distance distributions could be reconstructed when a sufficient number of average distances are measured, for example, through PRE effects [31•,32•,37•]. At least in principle, however, this information may not be sufficient, because in the presence of correlated motions, joint distributions for pairs of interatomic distances might be different even for two ensembles with the same distance distributions for individual atom pairs. The measurement of scalar or residual dipolar couplings could lift this degeneracy by providing additional information about the distributions of the orientations of internuclear vectors. In addition, one can anticipate that methods for the determination of time and space correlation functions [51•,52,53] will provide new opportunities for achieving an accurate description of structural ensembles representing the correlated dynamics of proteins.

Validation methods for conformational ensembles

As both the translation of the experimental information into structural restraints and their use in computational schemes require a range of assumptions, the structures that have been determined should be critically assessed to establish whether they are in conflict with other types of independent information that may be available. Several methods of validation have been considered. The internal consistency of a structure determination procedure can be verified by using only a subset of restraints and testing whether the remaining ones are reproduced (cross-validation) [43]. The use of cross-validation, however, although reducing the danger of overfitting, might not protect against over-restraining, especially in the case of highly heterogeneous ensembles of structures [31•,32•,37•,38]. In alternative validation methods, the statistical properties of the conformations obtained can be compared with those in structural databases. These methods have become highly sophisticated for native states [18•,43], but it may be too early to apply them to non-native states, as large repositories of high-resolution structures are not yet available.

A stringent method of validation involves using the calculated structures to make predictions about experimental measurements that were not employed as restraints. This approach has only begun to be applied to non-native states of proteins. The intermediate state for the folding of Im7, which was determined using protection factors from hydrogen exchange as restraints in molecular dynamics simulations, was validated by using NMR chemical shifts and protein engineering Φ-value measurements [54•]. The transition states for the folding of barnase, determined from Φ-value restraints, were validated by predicting the results of double-mutant cycle experiments [55•]. Suitable types of experimental parameters for validating (or indeed generating) non-native states include FRET-derived distances [47–50] and several NMR observables, such as RDCs [23•], PRE-derived distances [31•,32•,37•,38], J-couplings [56], chemical shifts [29•], R_2 values [36] and protection factors from hydrogen exchange [54•]. The future exploitation of these techniques will direct current efforts towards increasing the resolution of structures of non-native states of proteins, including complex cases such as the oligomeric assemblies associated with amyloid formation and the nascent chains in the ribosome.

Conclusions

The characterization at high resolution of conformationally heterogeneous states of proteins is challenging primarily 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 to describe native states (for which a reference structure is well defined) or highly denatured states (for which random-coil models are often well suited), and they require the development of computational methods for determining ensembles of structures. The use of sparse experimental information as ensemble-averaged restraints in molecular simulations represents a promising approach that should enable progress to be made from the current outline representations to highly accurate structural ensembles.

Acknowledgements

The research of MV is supported by the Royal Society, the Leverhulme Trust and the European Community.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Chiti F, Dobson CM: **Protein misfolding, functional amyloid, and human disease**. *Annu Rev Biochem* 2006, **75**:333-366.

2. Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions**. *Nat Rev Mol Cell Biol* 2005, **6**:197-208.

3. Boehr DD, McElheny D, Dyson HJ, Wright PE: **The dynamic energy landscape of dihydrofolate reductase catalysis**. *Science* 2006, **313**:1638-1642.

4. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D: **Intrinsic dynamics of an enzyme underlies catalysis**. *Nature* 2005, **438**:117-121.

5. Fersht AR: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* New York: WH Freeman; 1999.

6. Frauenfelder H, Sligar SG, Wolynes PG: **The energy landscapes and motions of proteins**. *Science* 1991, **254**:1598-1603.

7. Karplus M, Kuriyan J: **Molecular dynamics and protein function**. *Proc Natl Acad Sci USA* 2005, **102**:6679-6685.

8. Mittermaier A, Kay LE: **New tools provide new insights in NMR studies of protein dynamics**. *Science* 2006, **312**:224-228.

9. Vendruscolo M, Dobson CM: **Dynamic visions of enzymatic reactions**. *Science* 2006, **313**:1586-1587.

10. Volkman BF, Lipson D, Wemmer DE, Kern D: **Two-state allosteric behavior in a single-domain signaling protein**. *Science* 2001, **291**:2429-2433.

11. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **Charmm - a program for macromolecular energy, minimization, and dynamics calculations**. *J Comp Chem* 1983, **4**:187-217.

12. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.*: **Crystallography & NMR system: a new software suite for macromolecular structure determination**. *Acta Crystallogr D Biol Crystallogr* 1998, **54**:905-921.

13. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M: **Simultaneous determination of protein structure and dynamics**. *Nature* 2005, **433**:128-132.

14. Schwieters CD, Kuszewski JJ, Clore GM: **Using Xplor-NIH for NMR molecular structure determination**. *Prog Nucl Magn Reson Spectrosc* 2006, **48**:47-62.

15. Dobson CM, Sali A, Karplus M: **Protein folding: a perspective from theory and experiment**. *Angew Chem Int Ed Engl* 1998, **37**:868-893.

16. Vendruscolo M, Dobson CM: **Towards complete descriptions of the free-energy landscapes of proteins**. *Philos Transact A Math Phys Eng Sci* 2005, **363**:433-450.

17. Fersht AR, Daggett V: **Protein folding and unfolding at atomic resolution**. *Cell* 2002, **108**:573-582.

18. • Grishaev A, Bax A: **An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation**. *J Am Chem Soc* 2004, **126**:7281-7292.
The authors provide a lucid description of an approach whereby effective potentials for assessing the quality of structures are constructed from the analysis of protein structure repositories.

19. Paci E, Smith LJ, Dobson CM, Karplus M: **Exploration of partially unfolded states of human alpha-lactalbumin by molecular dynamics simulation**. *J Mol Biol* 2001, **306**:329-347.

20. Prompers JJ, Bruschweiler R: **General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation**. *J Am Chem Soc* 2002, **124**:4522-4534.

21. Shea JE, Brooks CL: **From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding**. *Annu Rev Phys Chem* 2001, **52**:499-535.

22. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS: **Native-like mean structure in the unfolded ensemble of small proteins**. *J Mol Biol* 2002, **323**:153-164.

23. • Bernado P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M: **Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings**. *J Am Chem Soc* 2005, **127**:17968-17969.
By analyzing the natively unfolded state of α-synuclein, this paper proposes how NMR RDCs can be used to select structures that represent the conformational preferences of highly heterogeneous states of proteins.

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

25. Heise H, Luca S, de Groot BL, Grubmuller H, Baldus M: **Probing conformational disorder in neurotensin by two-dimensional solid-state NMR and comparison to molecular dynamics simulations**. *Biophys J* 2005, **89**:2113-2120.

26. Neri D, Billeter M, Wider G, Wuthrich K: **NMR determination of residual structure in a urea-denatured protein, the 434-repressor**. *Science* 1992, **257**:1559-1563.

27. • Religa TL, Markson JS, Mayor U, Freund SMV, Fersht AR: **Solution structure of a protein denatured state and folding intermediate**. *Nature* 2005, **437**:1053-1056.
High-resolution structures of a denatured state and folding intermediate of the engrailed homeodomain protein are obtained using interproton distances measured by NOE effects as structural restraints in molecular dynamics simulations.

28. Vendruscolo M, Paci E, Karplus M, Dobson CM: **Structures and relative free energies of partially folded states of proteins**. *Proc Natl Acad Sci USA* 2003, **100**:14817-14821.

29. • Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE: **Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR**. *Nature* 2004, **430**:586-590.
An outline description of the partially folded intermediate states of two mutational variants of the Fyn SH3 domain is provided by the use of chemical shift information obtained through relaxation dispersion NMR experiments.

30. • Fuentes G, Nederveen AJ, Kaptein R, Boelens R, Bonvin A: **Describing partially unfolded states of proteins from sparse NMR data**. *J Biomol NMR* 2005, **33**:175-186.
A model of a partially unfolded state of the photoactive yellow protein is obtained by using NMR chemical shift information to guide the sampling of conformational space by molecular dynamics simulations.

31. • Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM: **Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations**. *J Am Chem Soc* 2005, **127**:476-477.
The authors use PRE-derived distance information to calculate an ensemble of structures representing the natively unfolded state of α-synuclein. A weak but statistically significant conformational propensity was found for the C-terminal region to fold back onto the NAC region. The possibility that this effect could decrease the tendency of α-synuclein to aggregate was discussed.

32. • Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M: **Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein**. *J Am Chem Soc* 2004, **126**:3291-3299.
The use of PRE-derived distances as ensemble-averaged restraints is proposed. The technique is applied to provide a model of the state of bovine ACBP at high concentrations of guanidine hydrochloride.

33. Wuthrich K, Wagner G: **Internal motion in globular proteins**. *Trends Biochem Sci* 1978, **3**:227-230.

34. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M: **Relation between native ensembles and experimental structures of proteins**. *Proc Natl Acad Sci USA* 2006, **103**:10901-10906.

35. Chou JJ, Case DA, Bax A: **Insights into the mobility of methyl-bearing side chains in proteins from (3)J(CC) and (3)J(CN) couplings**. *J Am Chem Soc* 2003, **125**:8959-8966.

36. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H: **Long-range interactions within a nonnative protein**. *Science* 2002, **295**:1719-1722.

37. • Francis CJ, Lindorff-Larsen K, Best RB, Vendruscolo M: **Characterization of the residual structure in the unfolded state of the Delta131Delta fragment of staphylococcal nuclease**. *Proteins* 2006, **65**:145-152.
In this work, an ensemble structure representing the conformation of the $\Delta131\Delta$ fragment of staphylococcal nuclease is calculated. This work illustrates clearly the possible problems introduced by underfitting for highly heterogeneous states of proteins.

38. Gillespie JR, Shortle D: **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* 1997, **268**:170-184.

39. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M: **Release of long-range tertiary**

interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Natl Acad Sci USA* 2005, **102**:1430-1435.

40. Bonvin AM, Brunger AT: **Conformational variability of solution nuclear-magnetic-resonance structures**. *J Mol Biol* 1995, **250**:80-93.

41. Burgi R, Pitera J, van Gunsteren WF: **Assessing the effect of conformational averaging on the measured values of observables**. *J Biomol NMR* 2001, **19**:305-320.

42. •• Richter B, Gsponer J, Varnai P, Salvatella X, Vendruscolo M: **Critical assessment of molecular dynamics simulations of proteins with ensemble-averaged NMR restraints.** *J Biomol NMR* 2006, in press.
This work proposes a simultaneous solution to the problems of underfitting and overfitting involving a technique in which different types of ensemble averaging are implemented for different NMR observables.

43. Spronk C, Nabuurs SB, Krieger E, Vriend G, Vuister GW: **Validation of protein structures derived by NMR spectroscopy**. *Prog Nucl Magn Reson Spectrosc* 2004, **45**:315-337.

44. Bouvignies G, Markvick P, Bruschweiler R, Blackledge M: **Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings**. *J Am Chem Soc* 2006, **128**:15100-15101.

45. • Clore GM, Schwieters CD: **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* 2004, **126**:2923-2938.
The authors provide evidence that an equivalent description of backbone dynamics can be obtained by using information derived from NMR RDCs or S^2 order parameters, or from crystallographic B factors.

46. Best RB, Vendruscolo M: **Determination of protein structures consistent with NMR order parameters**. *J Am Chem Soc* 2004, **126**:8090-8091.

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

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

49. Sherman E, Haran G: **Coil-globule transition in the denatured state of a small protein**. *Proc Natl Acad Sci USA* 2006, **103**:11539-11543.

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

51. • Bouvignies G, Bernado P, Meier S, Cho K, Grzesiek S, Bruschweiler R, Blackledge M: **Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings**. *Proc Natl Acad Sci USA* 2005, **102**:13885-13890.
A very extensive set of RDC measurements is used to identify correlated motions of a β sheet in the native state of the immunoglobulin-binding B1 domain of streptococcal protein G.

52. Lundstrom P, Mulder FAA, Akke M: **Correlated dynamics of consecutive residues reveal transient and cooperative unfolding of secondary structure in proteins**. *Proc Natl Acad Sci USA* 2005, **102**:16984-16989.

53. Wang T, Frederick KK, Igumenova TI, Wand AJ, Zuiderweg ERP: **Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements**. *J Am Chem Soc* 2005, **127**:828-829.

54. • Gsponer J, Hopearuoho H, Whittaker SBM, Spence GR, Moore GR, Paci E, Radford SE, Vendruscolo M: **Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7**. *Proc Natl Acad Sci USA* 2006, **103**:99-104.
A partially folded intermediate of the bacterial immunity protein Im7 with a substantial degree of non-native interactions is derived using hydrogen exchange protection factors as restraints in molecular dynamics simulations. The structure was validated by the prediction of chemical shift and Φ-value measurements.

55. • Salvatella X, Dobson CM, Fersht AR, Vendruscolo M: **Determination of the folding transition states of barnase by using Phi(I)-value-restrained simulations validated by double mutant Phi(IJ)-values**. *Proc Natl Acad Sci USA* 2005, **102**:12389-12394.
Structural models of the two successive transition states in the folding of the protein barnase are calculated using protein engineering Φ-values as restraints in molecular dynamics simulations. These models are then validated by using them to predict the results of double-mutant cycle experiments that were not involved in the structure determination procedure.

56. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM: **Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations**. *J Mol Biol* 1996, **255**:494-506.