

Conformational distributions of unfolded polypeptides from novel NMR techniques

Sebastian Meier, Martin Blackledge¹, and Stephan Grzesiek¹

Citation: *The Journal of Chemical Physics* **128**, 052204 (2008); doi: 10.1063/1.2838167

View online: <http://dx.doi.org/10.1063/1.2838167>

View Table of Contents: <http://aip.scitation.org/toc/jcp/128/5>

Published by the *American Institute of Physics*

COMPLETELY

REDESIGNED!



**PHYSICS
TODAY**

Physics Today Buyer's Guide
Search with a purpose.

Conformational distributions of unfolded polypeptides from novel NMR techniques

Sebastian Meier,¹ Martin Blackledge,^{2,a)} and Stephan Grzesiek^{3,b)}

¹*Carlsberg Laboratory, Gamle Carlsberg Vej 10, 2500 Valby, Denmark*

²*Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France*

³*Biozentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland*

(Received 28 September 2007; accepted 8 November 2007; published online 4 February 2008)

How the information content of an unfolded polypeptide sequence directs a protein towards a well-formed three-dimensional structure during protein folding remains one of the fundamental questions in structural biology. Unfolded proteins have recently attracted further interest due to their surprising prevalence in the cellular milieu, where they fulfill not only central regulatory functions, but also are implicated in diseases involving protein aggregation. The understanding of both the protein folding transition and these often natively unfolded proteins hinges on a more detailed experimental characterization of the conformations and conformational transitions in the unfolded state. This description is intrinsically very difficult due to the very large size of the conformational space. In principle, solution NMR can monitor unfolded polypeptide conformations and their transitions at atomic resolution. However, traditional NMR parameters such as chemical shifts, J couplings, and nuclear Overhauser enhancements yield only rather limited and often qualitative descriptions. This situation has changed in recent years by the introduction of residual dipolar couplings and paramagnetic relaxation enhancements, which yield a high number of well-defined, quantitative parameters reporting on the averages of local conformations and long-range interactions even under strongly denaturing conditions. This information has been used to obtain plausible all-atom models of the unfolded state at increasing accuracy. Currently, the best working model is the coil model, which derives amino acid specific local conformations from the distribution of amino acid torsion angles in the nonsecondary structure conformations of the protein data bank. Deviations from the predictions of such models can often be interpreted as increased order resulting from long-range contacts within the unfolded ensemble. © 2008 American Institute of Physics. [DOI: 10.1063/1.2838167]

INTRODUCTION

One of the most fundamental and enigmatic biological processes is the transition from the structural ensemble of an unfolded polypeptide chain to a folded protein structure. Despite the detailed knowledge on the dynamical behavior, thermodynamics, and structure of thousands of proteins, we are still lacking both a comprehensive theory and accurate practical tools that could predict the folding of a particular amino acid sequence into a particular three-dimensional structure.¹

An essential ingredient to such a theory would be the detailed quantitative description of the unfolded state ensemble. The latter remains a largely unresolved task since the immense number of degrees of freedom leads to enormous experimental as well as theoretical problems when structural probability distributions are to be determined from a finite number of measurable parameters.

Besides the fundamental importance of this “other half of the folding reaction”² for the protein folding problem, recent interest in unfolded states has been sparked by the

biomedical relevance of natively unfolded proteins, i.e., proteins which are predominantly unfolded in the cellular milieu in the absence of a suitable interaction partner. In particular, the key role that natively unfolded proteins play in cancer-related signalling and neurodegenerative diseases like Alzheimer’s and Parkinson’s has prompted a large number of biophysical studies on the characterization of the respective unfolded states.

A wealth of highly relevant experimental information on the unfolded state has been provided in recent years by NMR spectroscopy. This technique is particularly suited for a detailed description, because single atomic nuclei can be resolved in the spectra of partially folded or unfolded proteins and a number of NMR observables can be used to obtain structural and dynamical information. For example, while early views considered the unfolded polypeptide chain as featureless,³ NMR spectroscopic studies have now accumulated overwhelming evidence that, whenever studied at high enough detail, unfolded proteins have residual structure, even under strongly denaturing conditions.^{2,4,5} These findings agree with the conclusion that a certain bias towards the native state must exist within the unfolded ensemble⁶ in order to overcome Levinthal’s paradox, which states that a non-directed search of the entire conformational states acces-

^{a)}Electronic mail: martin.blackledge@ibs.fr.

^{b)}Electronic mail: stephan.grzesiek@unibas.ch.

sible to a polypeptide would take astronomical times to lead to a folded structure for any reasonable time estimate (pico-second to nanosecond) of a single search step.⁷

The most widely used NMR observables to characterize the unfolded state are chemical shifts^{5,8} that depend on backbone torsion angles, hydrogen exchange rates^{9,10} that report protection of hydrogen bonds against exchange with the solvent, relaxation parameters^{11–14} for dynamical information, three-bond J -couplings^{11,15,16} for torsion angle, and nuclear Overhauser effects (NOEs) for distance information. Of these classical parameters, NOE contacts provide the most distinctive long-range information. For example, contacts in similar arrangement as in the native state were detected by NOEs in a classical study of the urea-denatured 434 repressor.⁴ However, the reduced packing density of denatured states makes NOE detection rather insensitive. Furthermore, the quantitative interpretation of NOEs is very difficult, because they depend both on distance and local correlation times, which are usually not accessible by separate experiments.

More recently, paramagnetic relaxation enhancements¹⁷ (PREs) and residual dipolar couplings¹⁸ (RDCs) have been added to this arsenal of observables. In contrast to the more conventional parameters, the latter offer various advantages: first, they have a simple analytical dependence on the average over the internuclear (RDC) or electron-nucleus (PRE) distance vector. Thus, interpretation is much more straightforward than for NOEs. Second, they provide long-range information, and thus, also nonlocal order can be detected in the unfolded state. Third, they can be measured with high sensitivity and accuracy for a very large number of sites. For example, in a recent study, up to seven RDCs per peptide unit could be determined for urea-unfolded ubiquitin.¹⁹

These recent developments have made it possible to describe the angular and distance fluctuations of the unfolded state at unprecedented detail. The ultimate goal of such characterizations would be the calculation of complete conformational distributions of the unfolded or partially folded polypeptide chain from the experimental data. Due to the huge number of degrees of freedom, this mathematical problem is largely underdetermined, and a complete experimental determination of these distributions may remain elusive. Nevertheless, models of the unfolded state can easily be tested against the new experimental data. In particular, the use of coil libraries^{15,16,20} derived from the protein data bank (PDB) in conjunction with steric exclusion to simulate experimental RDC data yields spectacular agreement.^{21,22}

In the following, we will review these recent approaches that combine experimental NMR data with explicit model building *in silico* to achieve a meaningful characterization of unfolded proteins at atomic resolution.

STATISTICAL INTERPRETATION OF ${}^3J_{\text{HNH}_\alpha}$ SCALAR COUPLINGS AND CHEMICAL SHIFTS IN THE UNFOLDED STATE

In the past, three-bond scalar couplings and chemical shifts have been used to detect amino acid specific conformational biases in unfolded proteins. A number of three-bond

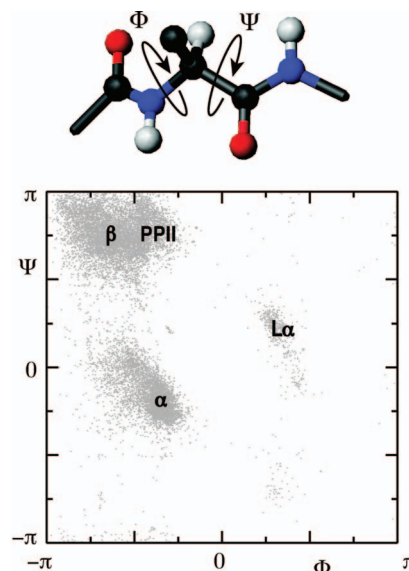


FIG. 1. (Color) Backbone conformations in proteins. Top: Due to the planarity of the peptide bond, only two angles Φ and Ψ determine the backbone conformation of amino acids in proteins. Atoms are shown in CPK colors with hydrogen in white, nitrogen in blue, carbon in gray, and oxygen in red. Bottom: Ramachandran plot of Φ and Ψ , where each point represents one amino acid conformation. 15 000 conformations are shown. The data points were randomly drawn from the protein coil library (Ref. 20) and correspond to nonglycine residues in the coil regions of protein crystal structures. The highest populated amino acid conformational basins are labeled as β -extended (β), polyproline II (PPII), right-handed α -helix (α), and left-handed α -helix ($L\alpha$).

scalar couplings give information on the backbone ϕ and ψ angles by various Karplus relations. However, only the homonuclear ${}^3J_{\text{HNH}_\alpha}$ coupling, reporting on ϕ , has found more widespread application for the characterization of unfolded states. Significantly different ${}^3J_{\text{HNH}_\alpha}$ values are detected for different amino acid types both in model peptides and denatured proteins.^{16,23,24} While alanine and glycine give the smallest ${}^3J_{\text{HNH}_\alpha}$ coupling constants around 5.8 Hz, maximal ${}^3J_{\text{HNH}_\alpha}$ values around 7.3 Hz are observed for the β -branched amino acids valine, isoleucine, and threonine. According to the Karplus relation, the higher ${}^3J_{\text{HNH}_\alpha}$ values for the β -branched amino acids imply more extended conformations. Presumably, these conformational preferences are caused by steric clashes of the branched sidechains with the backbone and neighboring sidechains, which exclude less extended conformations. Notably, such preferences are also observed in the folded state, where β -branched amino acids have a high occurrence in extended β -sheets, while alanine is among the amino acids with the highest propensity for α -helices.²⁵ Conversely, glycine can adopt a wealth of conformations in folded and unfolded states due to steric freedom from the absence of a sidechain, and as a consequence, the averaging over the ϕ space leads to small ${}^3J_{\text{HNH}_\alpha}$ values.

In order to explain this behavior of ${}^3J_{\text{HNH}_\alpha}$ couplings of unfolded polypeptides in a more quantitative way, Serrano and Schwalbe, Dobson and co-workers assumed that the conformations sampled by the unfolded amino acids would correspond to the distribution of conformations found for the respective amino acids in folded proteins (Fig. 1) in the

PDB. Indeed, averaging the Karplus relation over these distributions gave quite good agreement with the experimental ${}^3J_{\text{HNH}\alpha}$ values.^{11,15,16,23} In particular, a coil distribution derived from only nonsecondary structure conformations reproduced well the residue-specific NMR data.¹⁶ Good agreement was also found for predicted NOE values. This success of the coil model may indicate that conformational propensities in the unfolded state largely depend on the local chain chemistry.¹¹ Although the compatibility to the experimental data presents no proof for such a coil model, because the number of degrees of freedom by far exceeds the number of experimentally determined parameters, the coil model constitutes a very plausible quantitative concept for the unfolded state that can be checked against other observables. In addition, averaging over the coil distributions of analytical expressions for parameters can be used to produce theoretical baseline values. Deviations of experimental data from these predictions may then indicate additional order in the unfolded state.²³ A recently proposed method goes beyond this basic application of the coil model and derives residue-specific ϕ, ψ distributions from sparse J -coupling data (${}^3J_{\text{HNH}\alpha}$ and ${}^3J_{\text{H}\alpha\text{N}}$) by a maximum entropy formalism that minimizes the amount of additional information in the target distributions relative to a coil distribution prior.²⁶

Since ${}^1\text{H}^\alpha$, ${}^{13}\text{C}^\alpha$, ${}^{13}\text{C}^\beta$, and ${}^{13}\text{C}'$ chemical shifts are very sensitive to polypeptide backbone conformations, they are also well suited to report on the unfolded state. A similar averaging over coil conformations as applied to the prediction of ${}^3J_{\text{HNH}\alpha}$ couplings has been used successfully to explain experimental variations of ${}^1\text{H}^\alpha$ chemical shifts in unfolded proteins.¹⁵ Bias towards α -helical or β -sheet conformations in the unfolded state can be detected easily by comparing ${}^1\text{H}^\alpha$, ${}^{13}\text{C}^\alpha$, ${}^{13}\text{C}^\beta$, and ${}^{13}\text{C}'$ chemical shifts with empirical random-coil values.²⁷⁻³⁰ To achieve reliable quantification of secondary structure propensities from such conformation-dependent chemical shifts, various programs have been developed in recent years.^{8,31}

Examples that have used chemical shifts to derive secondary structure propensities in unfolded states include proteins related to misfolding diseases; e.g., high β -sheet propensity is found for the natively unfolded Tau protein involved in Alzheimer's disease in regions that contribute to misfolding and aggregation.³² In contrast, residual helical propensity detected in γ -synuclein seems to protect the protein from aggregation, whereas the involvement of the highly homologous α -synuclein in Parkinson's disease has been rationalized by a lower helical propensity.⁸

RESIDUAL DIPOLAR COUPLINGS

As compared to three-bond J -couplings, chemical shifts, and NOEs, RDCs may provide a far more extensive and quantitative description of local and, possibly, global order in the unfolded state. RDCs in solutions result from the average over the dipolar interaction between two nuclei i and j ,

$$D = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2 r^3} \left\langle \frac{(3 \cos^2 \theta(t) - 1)}{2} \right\rangle = D_{\max} \langle P_2(\cos \theta(t)) \rangle, \quad (1a)$$

$$D_{\max} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2 r^3}, \quad (1b)$$

where θ is the instantaneous angle of the internuclear vector with respect to the magnetic field, and the internuclear distance r is assumed fixed. The average within the angular parentheses corresponds to an ensemble average over the sample and a time average up to the millisecond range. For a solute that tumbles in an isotropic solution, the average in Eqs. (1a) and (1b) vanishes. However, the addition of suitable anisotropic alignment media to a solution³³ or even the intrinsic anisotropic susceptibility of certain solutes³⁴ can lead to weak anisotropic orientation of solutes and, hence, to observable RDCs. Geometric information can then be extracted from the measured RDCs as an angular orientation of the internuclear vector relative to the external magnetic field.

A number of different alignment media have been developed in the past ten years, reviewed recently by Prestegard *et al.*³⁵ The most common ones are lipid bicelles,³³ filamentous bacteriophages,^{36,37} liquid crystalline ethylene glycol/alcohol phases,³⁸ and strained polyacrylamide gels.^{39,40} The alignment mechanisms depend on the interaction with these media: whereas uncharged substances such as uncharged lipids or polyacrylamide orient the solutes mostly by steric exclusion, charged substances exert electric forces onto the charge distribution of the solute in addition to steric exclusion. For work with unfolded proteins under harsh denaturing conditions such as low pH, high temperature, or high urea concentrations, inert polyacrylamide gels have proven particularly useful.

In their pioneering work, Shortle and Ackerman applied strained polyacrylamide gels to the $\Delta 131\Delta$ mutant of staphylococcal nuclease under strongly denaturing conditions in 8M urea.¹⁸ Surprisingly, the denatured protein showed well observable RDCs, which correlated to values measured under nondenaturing conditions, where the protein is in an only partially unfolded form. The latter had been shown to have some resemblance to the native state topology by paramagnetic relaxation using 14 extrinsic spin labels.¹⁷ This led the authors to conclude that the overall spatial positioning and orientation of different chain segments in 8M urea should also be similar to that in the folded state. As described below, this interpretation may have to be revised to some extent following more recent work on the interpretation of RDCs from unfolded proteins, which are based either on random flight polymer chain models or on amino-acid-specific conformational propensities.

The nature of the RDC average

For the discussion of RDCs from unfolded polypeptide chains, it is instructive to consider the nature of the averaging of Eq. (1a) in more detail. Writing the average of Eq. (1a) over N molecules in explicit form yields

$$D = D_{\max} \frac{1}{N} \sum_{k=1}^N \frac{1}{t_{\max}} \int_{t=0}^{t_{\max}} P_2(\cos \theta_k(t)) dt, \quad (2)$$

where t_{\max} is the maximal time, during which experimental averaging occurs, i.e., on the order of tens to hundreds of milliseconds.

Certain limits may be derived for the time scale of this averaging process:

- (1) Usually, the alignment experiments are carried out such that no appreciable line broadening is observed for the multiplets, e.g., the $^1\text{H}-^{15}\text{N}$ doublet, through which the dipolar couplings are observed. If different molecules experience different averages over $P_2(\cos \theta(t))$ during t_{\max} , then the multiplet lines should be smeared out according to the different resulting RDCs. Since this is not the case, all molecules apparently have a similar average of $P_2(\cos \theta(t))$. This seems only possible if the single molecules sample very similar parts of the conformational space during t_{\max} . This, in fact, means that every single molecule should sample most of the accessible conformational space during t_{\max} , because, otherwise, the sampled parts of the conformational space and the respective averages would be different. Thus, the molecular average in Eq. (2) can be omitted:

$$D = D_{\max} \frac{1}{t_{\max}} \int_{t=0}^{t_{\max}} P_2(\cos \theta(t)) dt. \quad (3)$$

- (2) Since no appreciable line broadening and no additional lines are observed from the alignment, the solute molecules are in fast exchange between a state that interacts with the orienting medium and a free form. If one assumes a model of two-site exchange between bound and free forms, a limit on the exchange times may be derived based on the time scale of translational diffusion. Typical free path lengths of various alignment media are on the order of 100 nm. For example, steric alignment by uncharged bicelles such as dimyristoylphosphatidylcholine (DMPC)/dihexanoylphosphatidylcholine (DHPC) is usually carried out at concentrations of about 5% (w/v) \approx 5% (v/v). The bicelle thickness is about 5 nm. Assuming that the bicelles are perfectly planar parallel sheets, the mean distance between bicelles should be 95 nm. A similar order of magnitude is also found for the distance between the lamellar layers of L_{α} phases in ethylene glycol/alcohol/water mixtures⁴¹ or for the mean distance between bacteriophages at typical concentrations of 30 mg/ml.⁴² At room temperature, a protein of 10 kDa diffuses in aqueous solution across 100 nm within about 5 μs (see, e.g., Ref. 43). Thus, the lifetime of the free molecule until it hits the alignment medium should be on the microsecond time scale for two-site exchange. The situation is probably more complicated for polyacrylamide gels where the solute protein may rather be in almost continuous contact with the polyacrylamide mesh, and

diffusion may be more hindered. In fact, a certain increase in rotational correlation times is already noticeable for folded ubiquitin at polyacrylamide concentrations of 4%–7% (w/v).⁴⁰

- (3) The residual alignment itself is weak with order parameters $S = \langle P_2(\cos \theta) \rangle$ in the range of 10^{-4} – 10^{-3} . Assuming two-site exchange, the lifetime of the aligned state (contact time) should be 10^3 – 10^4 times smaller than the lifetime of the free state. This would correspond to contact times on the order of 0.5–5 ns. This time scale is similar to typical protein rotational correlation times and larger than fast backbone motions in flexible parts of a protein, indicating that averaging over a certain fraction of the orientational and conformational distribution may already occur during a single contact event.

Global and local molecular order from RDCs

For rigid molecules, the average in Eq. (3) is usually expressed as an average over the orientation of the entire molecule, which is described by the alignment tensor or Saupe order matrix:

$$D = D_{\max} A_{zz} [P_2(\cos \vartheta) + (\eta/2) \sin^2 \vartheta \cos 2\varphi], \quad (4)$$

where A_{zz} is the zz component of the alignment tensor in the tensor frame, η is the rhombicity defined as $\eta = (A_{xx} - A_{yy})/A_{zz}$, and $\{\vartheta, \varphi\}$ are the respective polar coordinates of the internuclear vector. In case the molecule is only approximately rigid, such as a folded protein, the same formalism is usually still applied with the modification that local order of the dipolar vector is described by a local order parameter.^{44–46} The validity of this approach hinges on the assumption that the global alignment is not strongly coupled to local fluctuations, such that a common global alignment tensor can be used for all the different local conformations. This assumption seems reasonable for folded proteins, but variations in order parameters from different alignment media and different internuclear vectors have been interpreted recently as evidence for a more complicated behavior.⁴⁷

In contrast to the folded state, it is obvious for unfolded proteins that the global shape and the corresponding alignment will depend strongly on local conformations, e.g., when a turn residue changes its backbone angles. Thus, it clearly has no sense to assume a common alignment tensor for all the different conformations of an unfolded peptide chain. Therefore, the full average of Eq. (3) has to be calculated. The alignment experiments on unfolded polypeptides are usually also carried out such that no significant line broadening occurs. From the arguments given above, it follows that also a single unfolded molecule should sample close to all of its accessible conformations together with the respective possible orientations within the time t_{\max} .

A viable approximation for an unfolded protein (see below) may be to assume that each individual conformation orients according to a specific alignment tensor, such that the total average of Eq. (3) can be written as

$$D = D_{\max} \frac{1}{M} \sum_{k=1}^M A_{k,zz} [P_2(\cos \vartheta_k) + (\eta_k/2 \sin^2 \vartheta_k \cos 2\varphi_k)], \quad (5)$$

where the sum over k extends now over all M molecular conformations and the corresponding alignment tensors that the molecule samples during the time t_{\max} .

Perturbation of unfolded protein conformations by interactions with the alignment medium

Before addressing the different approaches to the interpretation of RDCs in unfolded proteins, it should be emphasized that these rely on the assumption that local conformational properties of the unfolded polypeptide are not significantly perturbed by the interaction with the alignment medium, and that, therefore, the RDCs report on the behavior of the chain in more physiological environments. This assumption is clearly very difficult to prove. Often, it is justified by the absence of significant chemical shift differences between isotropic and anisotropic spectra, and by the lack of nonuniform broadening of peaks that may arise from specific interactions between the unfolded chain and the aligning medium. However, this reasoning is not convincing, since the very low relative population of the aligned state (10^{-4} – 10^{-3}) together with fast chemical exchange between bound and free states will lead to negligible chemical shift and line broadening effects.

A far better way would be to show that RDCs obtained in different alignment media imply identical conformational preferences. This entails considerable difficulties, since, besides polyacrylamide gels, many other media cannot be used under harsh unfolding conditions and are also prone to stronger interactions with unfolded polypeptides. Thus, steric alignment media such as bicelles or liquid crystalline ethylene glycol/alcohol phases³⁸ show pronounced hydrophobic interactions that often render the peptide resonances broad or unobservable.⁴⁸ Conversely, it seems very plausible that electrostatically aligning media such as filamentous phages cause perturbations of the polypeptide conformations due to their strong electric forces. For these reasons, no systematic comparisons have been performed and only sparse data are available, which rather argue for minor perturbations by the alignment medium. For example, very similar RDC patterns were observed for intrinsically unfolded α -synuclein aligned either by phages or C8E5 ethylene glycol/octanol phases.⁴⁹ Similarly, a study on small peptides⁵⁰ showed a break in the uniform RDC profile around a tryptophan residue for alignment both by stretched polyacrylamide gels and by phages. Furthermore, the RDC profile of phage-aligned α -synuclein could be predicted with reasonable accuracy from a statistical coil model (see below) that ignores perturbation of the polypeptide conformation by electric field effects.⁵¹

As an alternative to comparisons of RDC data from different alignment media, the RDC-derived conformational preferences have been validated in many studies by other NMR parameters obtained under isotropic conditions, e.g., chemical shifts, scalar couplings, and relaxation data. Indeed, sequence profiles of such parameters similar to the RDC pat-

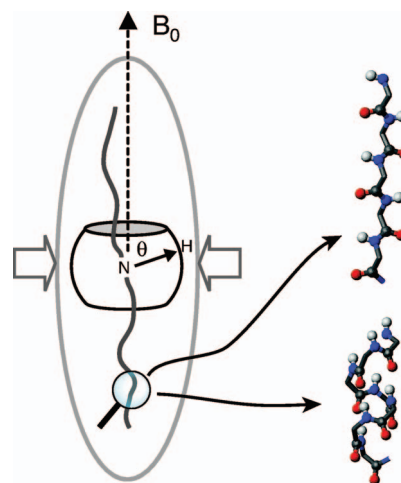


FIG. 2. (Color) Residual order of a bond vector (NH) in an elongated unfolded protein oriented by steric alignment in an elongated cavity that is parallel to the magnetic field. The depicted situation corresponds to radially squeezed gels. Extended conformations (right top) will contribute negative ${}^1D_{\text{NH}}$ RDCs to the ensemble average, as the NH vectors are predominantly orthogonal to the magnetic field and the integral over $P_2(\cos \theta)$ is negative [Eq. (3)]. In contrast, NH vectors are parallel to the helix axis in helical conformations. For longitudinal alignment of the helices, the NH vector mostly samples directions parallel to the magnetic field, causing an inversion of the sign of RDCs from the dependence on $P_2(\cos \theta)$. The predominance of extended conformations in highly unfolded proteins leads to mostly negative ${}^1D_{\text{NH}}$ RDCs in radially squeezed gels.

terns have been found for apomyoglobin,⁵² acyl-CoA binding protein (ACBP),⁵³ T4 fibrin foldon,⁵⁴ model peptides,⁵⁰ urea-denatured ubiquitin,⁵⁵ and Tau.⁵⁶

Interpretation of RDCs in the unfolded state by random flight polymer models

As already noticeable in the data of Shortle and Ackerman,¹⁸ ${}^1D_{\text{NH}}$ RDCs of unfolded proteins aligned in longitudinally stretched gels tend to have a unique negative sign throughout the chain. Moreover, despite significant differences between single residues, the RDC profiles seem to follow a broad bell-shaped distribution, with larger couplings at the midpoint of the chain, falling off to negligible values at the chain termini. This behavior has been explained by a simple random flight model of a linear homopolymer put forward by Annala and co-workers.^{57,58} In this model, the restrictions of bond movements due to the covalent polymer geometry lead to nonvanishing RDCs even when the torsion angles along the polymer chain adopt random conformations. Figure 2 depicts this situation for an extended random peptide chain that is aligned parallel to the magnetic field in an elongated cavity. For the N-H vector shown in the figure, the average $(1/t_{\max}) \int_{t=0}^{t_{\max}} P_2(\cos \theta(t)) dt$ will be negative, and hence, ${}^1D_{\text{NH}}$ should be negative due to the differing signs of γ_{H} and γ_{N} in Eqs. (1a) and (1b).

In the initial study,⁵⁷ the unfolded protein was described as a series of connected segments of given length, which undergo a restricted random walk that is numerically integrated over all available orientations of each segment. In the presence of an obstacle, orientational sampling is more restricted in the center of the chain than at the termini. This leads to stronger RDCs in the center and a bell-shaped dis-

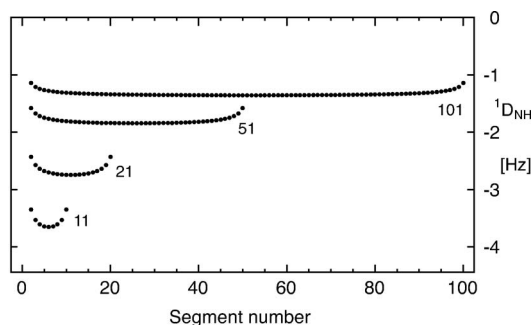


FIG. 3. $^1D_{NH}$ RDCs for a random flight chain model of the unfolded polypeptide chain. Profiles are obtained from an analytical description for chains of length 11, 21, 51, and 101. [Data redrawn with permission from O. I. Obolensky, K. Schlepckow, H. Schwalbe, and A. V. Solov'yov, *J. Biomol. NMR* 39, 1 (2007). Copyright 2007, Journal of Biomolecular NMR.]

tribution of the size of RDCs along the chain. The model also predicts that longer random chains should give smaller RDCs since the overall average shape will be more spherical than for shorter chains. Recently, Obolensky *et al.* have presented an improvement of this approach,⁵⁹ which remedies some theoretical shortcomings of the initial model and gives an extension to a true three-dimensional random walk model that can be solved in closed analytical form at any position in the chain. While the predicted absolute size of RDCs differs significantly from the initial model, general features are confirmed, i.e., RDCs along the chain should follow a flat-bell-shaped distribution and RDCs should be larger for shorter chains (Fig. 3).

Although the description of a natural amino acid sequence as a homopolymer is clearly unrealistic, these models capture many aspects of the physical situation of an aligned unfolded polypeptide and provide a simple conceptual framework for the qualitative understanding of the experimental observations. Again, deviations of RDC profiles from the model predictions can then be interpreted as variations of order.

EXPERIMENTAL RDC PROFILES OF UNFOLDED POLYPEPTIDES

A recent, comprehensive list of unfolded polypeptides for which RDC sequence profiles have been measured has been provided in Ref. 59. In the following, we discuss general features of several examples.

As already mentioned, the all-negative H^N RDC profile of $\Delta 131\Delta$ staphylococcal nuclease in 8M urea¹⁸ resembled the profile for this destabilized protein in standard buffer conditions, indicating also some resemblance of the tertiary topology to the native protein. However, with the exception of eglin,⁶⁰ no such correlations between native and denatured states have been found in other proteins. Examples include protein GB1,⁶¹ apomyoglobin,⁵² full-length staphylococcal nuclease, and a further staphylococcal nuclease truncation mutant.⁶² In the latter study, RDCs of the full-length protein and the truncation mutant were uncorrelated in the folded state, which can be explained by the different alignment properties of the two proteins. However, the all-negative RDCs in the denatured states of both proteins showed corre-

lation ($R=0.86$), apparently due to similar local chain properties and the lack of a larger stabilized structure.

Residual secondary structure propensities have been derived from variations within the RDC sequence profile of various denatured proteins. Based on a change in magnitude or even sign of $^1D_{NH}$, residual helical propensities were identified in acid-denatured states of apomyoglobin⁵² and the acyl-CoA binding protein (ACBP)⁵³ for regions which form helices in the native state. Such a change in sign is expected for α -helical segments which orient parallel to the helix axis, since the NH vector is parallel to this orientation axis and not perpendicular as in extended conformations (see also Fig. 2).⁵² In contrast to the acid-denatured state, no such helical propensities were detected in urea- or guanidinium-HCl-denatured forms of both proteins. Similarly, the particular stability of a short turn sequence (KDGE) in a β -hairpin structure of the fibritin foldon domain was detected from variations in the RDC profile above the melting transition.⁵⁴ The latter supports a role of turn sequences in the nucleation of β -sheet folding. The RDC-derived secondary structure propensities in the aforementioned studies also correlated to chemical shifts⁵²⁻⁵⁴ and to backbone dynamics derived from relaxation experiments.⁵²

Conformational studies on natively unfolded proteins have attracted particularly strong interest due to the surprising prevalence of natively unfolded proteins in eukaryotic genomes and due to the high biomedical relevance of these proteins. RDCs have been used to detect local structure in the Tau protein involved in Alzheimer's disease, where four short repeats show a sign inversion of $^1D_{NH}$ values that are rationalized as local turn conformations.^{63,64} These turns prove surprisingly resistant to the addition of high concentrations of urea. Functional importance of the turns is suggested as they interrupt regions with high β -sheet propensity and as they contribute to the binding of polyanions that enhance the aggregation of Tau.^{63,64} Variations in the RDC profile (Fig. 4) have also been detected for the protein α -synuclein⁴⁹ involved in Parkinson's disease, which are apparently caused by hydrophobic or electrostatic long-range interactions (see below).

RDCs can serve not only to study residual structure in the unfolded state, but also to follow conformational changes during, or prior to, unfolding. For the α -helical ribonuclease S-peptide, a gradual decrease of the size of RDCs, indicative of helix destabilization, was observed with increasing temperatures or decreasing salt concentration.⁶⁵ Thermal unfolding was also studied for the fibritin foldon domain by measuring $^1D_{NH}$ and $^1D_{CaH\alpha}$ RDCs.⁵⁴ Above the melting transition, both types of RDCs converged to small, but non-zero values that showed residual structure in the β -turn and around prolines (Fig. 5). The sequence-specific RDC melting behavior also correlated with chemical shift changes. Similarly, Ding *et al.* have studied the thermal unfolding of protein GB1 and identified two melting hotspots on the basis of differential RDC temperature behavior.⁶¹

A further free variable in the characterization of unfolded polypeptides by RDCs is the amino acid sequence itself. If the sequence dictates local conformational preferences, then substitutions of specific amino acids should have

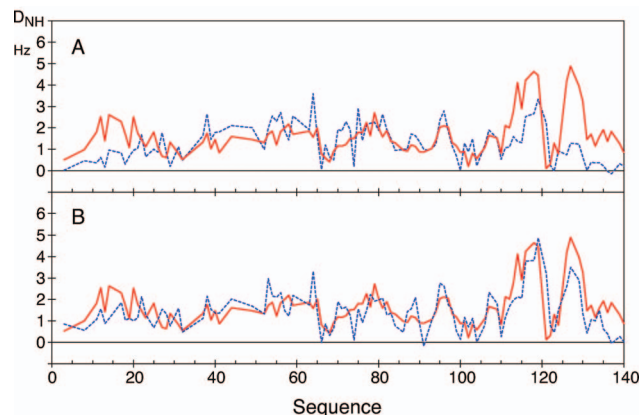


FIG. 4. (Color) Detection of long-range order in α -synuclein by $^1D_{NH}$ RDCs. (A) Experimental data (solid, red) obtained by alignment in Pf1 bacteriophages under high salt concentrations. Under these conditions, alignment is mainly steric. Simulated data (dashed, blue) were obtained by the FM approach without long-range contacts. (B) Experimental data (solid, red) as in (A). Simulated data (dashed, blue) were obtained by FM with the additional constraint of long-range contacts between residue regions 1-20 and 121-140. [Data are taken with permission from P. Bernado, C. W. Bertocini, C. Griesinger, M. Zweckstetter, and M. Blackledge, *J. Am. Chem. Soc.* **127**, 17968 (2005). Copyright 2005, Journal of the American Chemical Society.]

significant effects on RDC patterns. This approach has not been explored very widely. RDC data by Ohnishi and Shortle on various small peptides of different lengths⁴⁸ were mainly interpreted in terms of relative chain stiffness, but not detailed enough to derive amino-acid-specific effects. More recently, Dames *et al.*⁵⁰ systematically studied the influence of amino acid substitutions X on model peptides EGAAXAASS as monitored by $^1D_{NH}$ and $^1D_{CaH\alpha}$ RDCs in strained polyacrylamide gels. The RDCs showed a specific dependence on the substitution X that correlates to steric or hydrophobic interactions with adjacent amino acids. In particular, the RDCs for glycine and proline substitutions indicated less or more order, respectively, than other amino acids. Furthermore, RDCs for aromatic substitutions by tryptophan and tyrosine gave evidence of a kink in the peptide backbone, whereas the substitution with the β -turn sequence KNGE differed significantly from single amino acid substitutions. Terminal and next neighbor effects on the RDC profile could also be demonstrated.

STATISTICAL COIL MODELS OF THE UNFOLDED STATE

In the light of the variations of observed RDCs along an unfolded peptide chain from the bell-shape prediction of the simple homopolymer random flight models, it was clearly important to enhance these models by amino-acid-specific conformational behavior. In a first attempt, higher or lower flexibility of single glycine and proline residues, respectively, were incorporated into simulations of random homopolymers,⁵⁸ showing that RDC profiles changed to lower or higher order parameters in the corresponding regions. Similar profiles could also be obtained from all-atom molecular dynamics simulations of these polymers, where the RDCs of single snapshots were calculated by prediction of the steric alignment tensors using PALES.⁶⁶

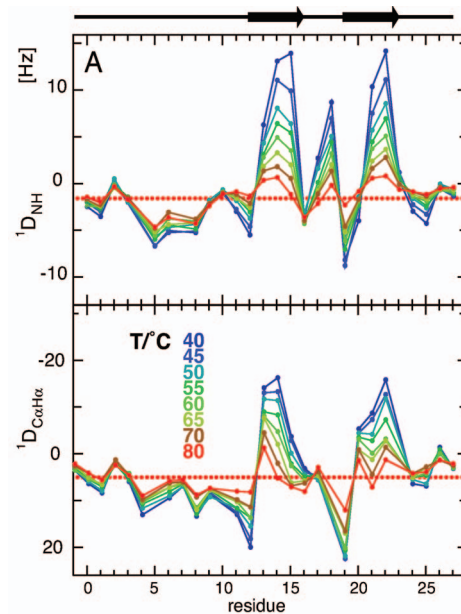


FIG. 5. (Color) $^1D_{NH}$ and $^1D_{CaH\alpha}$ RDCs in the T4 fibrin foldon hairpin between 40 and 80 °C at pH 2. Averages of the $^1D_{NH}$ and $^1D_{CaH\alpha}$ RDCs at 80 °C are shown as dotted lines. RDCs serve to directly monitor the order-disorder transition in the structured hairpin region (residues 12-16 and 19-23, indicated on top) during thermal unfolding. [Figure adapted with permission from S. Meier, S. Guthe, T. Kiefhaber, and S. Grzesiek, *J. Mol. Biol.* **344**, 1051 (2004). Copyright 2004, Journal of Molecular Biology.]

However, these simple enhancements did not provide the power to predict RDCs for arbitrary “random-coil” amino acid sequences. Two very similar approaches were, therefore, developed almost simultaneously that explicitly account for the heteropolymeric nature of the peptide chain, using random sampling of distinct amino-acid-specific ϕ/ψ propensities to construct a large number of conformers of the protein of interest.^{21,22} In both approaches, amino-acid-specific conformational energy basins were derived from a structural database of nonsecondary structural elements in folded proteins,^{15,25} an approach discussed above for the prediction of NMR scalar couplings and chemical shifts in unfolded chains.^{11,15,16,23} In the case of Jha *et al.*,²¹ the authors tested a number of conformational database sampling approaches by comparing the relevance of eliminating α -helical, β -sheet, and turn conformations from the complete database and found that removal of all three structural elements turned out to be the most appropriate. The authors explicitly incorporated nearest neighbor effects into the peptide chain construction algorithm by introducing an additional statistical potential that accounted for nonbonding interactions with the adjacent amino acids. This resulted in a better reproduction of experimental data as compared to neglecting such next neighbor effects. Clearly, this finding that neighboring amino acids influence the conformational sampling properties of a given amino acid is highly relevant to the understanding of nucleation in protein folding.

In the case of Bernado *et al.*,²² the selected database was derived from a recent compilation of 500 high resolution crystal structures.⁶⁷ α -helical and β -sheet conformations were again removed from the sampled structures, although turn conformations were retained. Additional pseudoamino

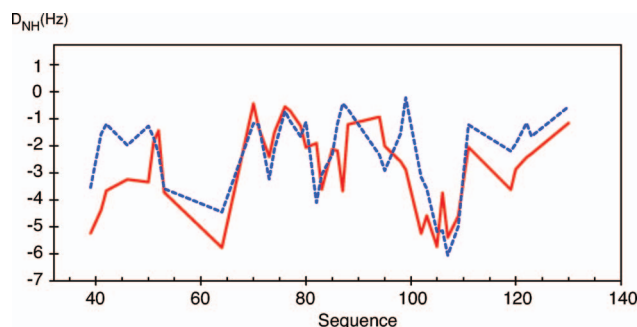


FIG. 6. (Color) Comparison of experimental (solid line, red) and predicted (dashed line, blue) $^1D_{NH}$ RDCs in urea-denatured staphylococcal nuclease mutant $\Delta 131\Delta$. RDCs are predicted by the coil model using sidechain steric exclusion. [Figure adapted with permission from P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok, and M. Blackledge, Proc. Natl. Acad. Sci. U.S.A. **102**, 17002 (2005). Copyright 2005, Proceedings of the National Academy Sciences of the United States of America.]

acids were defined to account for specific backbone dihedral angle sampling occurring, for example, for residues preceding prolines. Steric interactions between residue-specific spheres centered on the β -carbon atoms of each amino acid (α -proton in the case of glycines) were used to roughly account for van der Waals type interactions and to avoid overlap in the chain. The problem of steric clashes was dealt with by stopping the chain propagation whenever contacts occur, removing the previous two amino acids, and rebuilding the chain with other randomly selected backbone dihedral angles until no contacts are encountered.

In both approaches, conformers are constructed by randomly sampling these amino-acid-specific conformational basins. Dipolar couplings are then predicted from the ensemble of structures either by averaging the orientation of the vector with respect to the inertia tensor axis²¹ or by averaging RDCs [Eq. (5)] predicted for each copy of the ensemble²² using a shape-based alignment algorithm.^{66,68}

The latter approach, termed flexible-Meccano or FM, was applied in the first instance to a two-domain viral protein, protein X, comprising an apparently entirely unfolded and a folded domain.²² The advantages of this system are quite clear, as the relative level of alignment of the two interdependent domains provides a quantitative test of the validity of the approach. $^1D_{NH}$ and $^2D_{C'NH}$ RDCs were accurately reproduced throughout the protein. It is worth noting that in this and all other systems so far analyzed using the FM approach, the convergence of average RDCs requires the production of on the order of 50 000 structures for a molecule of around 100 amino acids.

Good agreement between experimental and FM-predicted RDCs was also found for $^1D_{NH}$ measured in a number of urea-unfolded proteins;²² in particular, the distribution of RDCs from $\Delta 131\Delta$ mutant of staphylococcal nuclease was accurately reproduced simply on the basis of local conformational propensities, notably in the absence of the previously invoked residual tertiary fold (Fig. 6). Intriguingly, despite similar reproduction of experimental RDCs from urea-unfolded apomyoglobin by the approaches from Bernado *et al.* and Jha *et al.*, the latter did not reproduce the $\Delta 131\Delta$ mutant of staphylococcal nuclease as convincingly.

The consistent reproduction of experimental profiles establishes the statistical coil sampling approach as a method for predicting baseline values that result directly from the conformational properties of the primary sequence and can be considered as random coil. As with the other models described, departure from such baseline values can then be interpreted in terms of specific local or long-range conformational behavior.

In one such study, the local inversion of $^1D_{NH}$ RDCs in four repeated domains (R1–R4) of a 130 amino acid construct of the natively unfolded protein Tau could not be reproduced by the simple coil FM approach.⁶⁴ Unrestrained extended molecular dynamics simulations were, therefore, performed, which revealed strong tendencies to form turns, identified as β I-turns for repeats R1–R3. Incorporation of the backbone dihedral sampling resulting from the molecular dynamics simulations into the statistical coil model closely reproduced experimental RDC values, which suggested that the local β I-turn conformations and their associated populations were probably accurate.

It is apparent that the effective order of the polypeptide chain that is induced by sampling the amino-acid-specific backbone conformational wells is dictated by the torsion angle distributions present in the database. For example, glycine, sampling widely distributed parts of the ϕ/ψ space, will induce less order and, therefore, lower RDCs, than preproline/proline pairs, whose conformational sampling are much more restricted. Cho *et al.*⁶⁹ have noted recently that averaging the amino acid sidechain bulkiness over a five amino acid window reproduces $^1D_{NH}$ RDCs along the chain with similar accuracy as the coil model. Apparently, this procedure translates the bulkiness, which is a measure for near neighbor interactions, to the order along the backbone, which is probed by the RDCs. Similar ideas have previously been exploited for the interpretation of transverse relaxation rates measured in denatured proteins.¹³ The fact that an average bulkiness extending further than over just the adjacent neighbors better reproduces the RDC data supports the idea that steric effects beyond nearest neighbors restrict the conformational search in the unfolded state. Indeed, a recent study on three-bond scalar couplings measured in short peptides of differing lengths⁷⁰ indicated that backbone dihedral angle distributions depend strongly on the overall length of the peptide. This underlines not only the caution that must be exercised when translating information derived from peptides into more general rules applicable to unfolded proteins, but also highlights the influence that relatively distant neighbors in the primary chain may have on local conformational properties.

Refinement of the coil model using a larger number of experimental parameters

We have seen that the steric alignment of unfolded proteins and the detection of mostly $^1D_{NH}$ RDCs provide important information on their conformational behavior that can be reproduced to reasonable accuracy by using the simple coil models of the unfolded state derived from the conformational preferences in folded proteins. However, it is clear that the agreement is not perfect and that also other models may

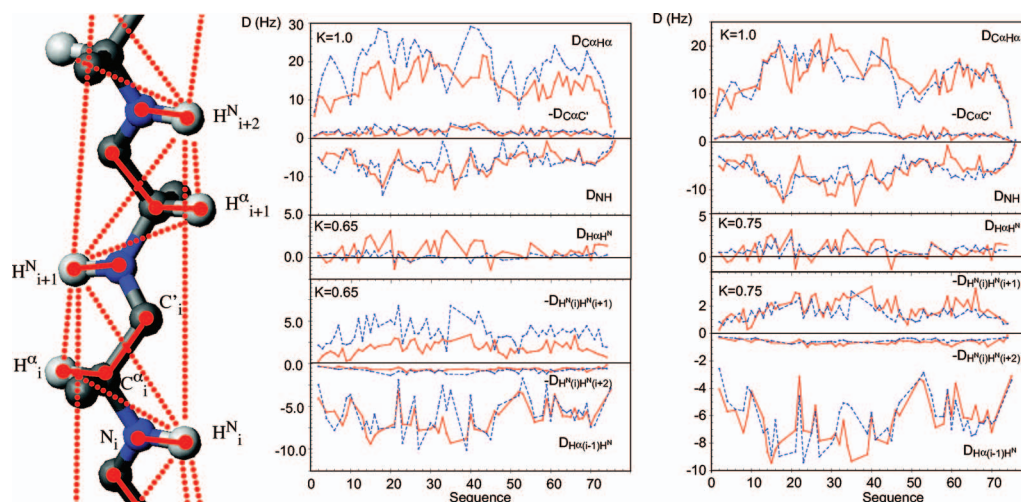


FIG. 7. (Color) Left: Strand of β -sheet conformation showing an extended set of RDCs measured to refine the conformational model of urea-denatured ubiquitin. Middle, right: Agreement of the observed RDCs (solid, red) with predictions (dashed, blue) from the coil model (middle) and from a modified coil model assuming increased sampling of extended backbone conformational basins under denaturing conditions (right). Values of K indicate additionally necessary scaling factors. [Figure adapted with permission from S. Meier, S. Grzesiek, and M. Blackledge, *J. Am. Chem. Soc.* **129**, 9799 (2007). Copyright 2007, Journal of the American Chemical Society.]

be possible. To achieve a more precise description, refinement of these models will be necessary, which requires the test against a larger number of experimental parameters. Alignment by electrostatic media can yield additional linearly independent RDCs that can help in the triangulation of the internuclear vector.⁷¹ It has recently been demonstrated that alignment of natively unfolded proteins in electrostatically aligning media can also be successfully predicted.⁵¹ This study took conformers calculated using FM, and predicted RDCs aligned in bacteriophage using a simplified alignment model that approximates the electrostatic interaction between an unfolded protein and an ordered liquid-crystal particle as that between the protein sidechain charges and the electric field of the phage, and again averaged the resulting RDCs over all protein conformers. The potential for extracting meaningful information from additional, differently aligning media has also been recently demonstrated by Gebel *et al.*⁷²

An additional approach to a better definition of local conformational sampling in unfolded proteins is to measure more RDCs from the peptide chain. Although the study of protein X described above reproduced both $^1D_{NH}$ and $^2D_{C'NH}$ RDCs, relatively few studies have extended the investigation of RDCs beyond the analysis of $^1D_{NH}$ couplings. In the conformational study of the model peptides EGAAAX-AASS described above,⁵⁰ $^1D_{NH}$ couplings were reasonably reproduced using the statistical coil model, but the distribution of $^1D_{CaHa}$ couplings deviated more strongly.

In a recent study, we sought to extend our understanding of the structural behavior of unfolded proteins further by measuring a more extensive set of RDCs in urea-denatured ubiquitin. Up to seven couplings per peptide unit were measured, including $^1H^N-^1H^N$ and $^1H^N-^1H^\alpha$ RDCs obtained in deuterated and protonated ubiquitin, respectively (Fig. 7, left). The FM approach was then used to predict expected one-bond RDCs ($^1D_{NH}$, $^1D_{CaHa}$, and $^1D_{CaC'}$) from the unfolded chain.¹⁹ While the resulting profiles along the chain

were well reproduced, the different types of RDCs ($^1D_{NH}$, $^1D_{CaHa}$, and $^1D_{CaC'}$) required significantly different scaling factors to reproduce their absolute values. Further analysis revealed that this was possibly due to a larger population of extended conformations than had been predicted by the standard coil library. A similar comparison of calculated and experimental sequential, interproton ($^1H^N_i-^1H^N_{i+1}$, $^1H^N_i-^1H^N_{i+2}$, and $^1H^N_i-^1H^\alpha_{i-1}$) RDCs again indicated an increased sampling of extended conformations in the presence of urea. This hypothesis was tested by performing statistical coil calculations using modified amino-acid-specific potentials to sample more extended conformations. This bias then simultaneously reproduced all measured one-bond RDCs with a single scaling factor (Figs. 7 and 8). A further scaling factor of 0.75 had to be applied to predicted sequential, interproton RDCs to also reproduce these experimental data. The latter may be caused by interpeptide dynamic averaging effects that are not

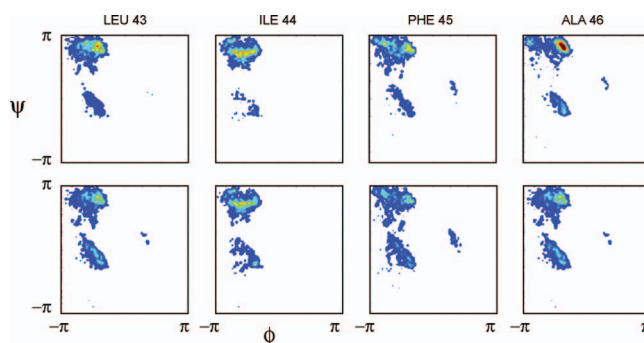


FIG. 8. (Color) Top: Conformational distributions of four sequential amino acids in denatured ubiquitin derived by increased sampling of extended backbone conformational basins in order to optimally reproduce measured multiple RDCs (Fig. 7). Bottom: Comparison to the database sampling of the unbiased coil model. Red indicates higher populations, yellow intermediate populations, and blue lower populations. [Figure reprinted with permission from S. Meier, S. Grzesiek, and M. Blackledge, *J. Am. Chem. Soc.* **129**, 9799 (2007). Copyright 2007, Journal of the American Chemical Society.]

(yet) fully described by the biased conformational ensemble. A simultaneous analysis of $^3J_{\text{HNH}\alpha}$ scalar couplings measured under the same conditions suggested that neither polyproline II nor extended β -regions dominate the additional sampling of extended conformations.

The increased sampling of extended conformations was also applied to the previously mentioned two-domain protein X containing a long, natively unfolded chain in the absence of urea. The couplings simulated under these conditions extensively violate experimental $^1D_{\text{HN}}$ and $^2D_{\text{C'NH}}$ RDCs, whereas standard database propensities reproduce the experimental data very closely. Thus, the necessity to use more extended conformations only for a urea-denatured protein seems to indicate that the binding of urea to the polypeptide extends an unfolded amino acid chain. In summary, this study showed that the information from different RDCs in the peptide unit is highly complementary and that it may be used as a stringent test for the accuracy of current models of the unfolded state.

DETECTION OF LONG RANGE ORDER IN UNFOLDED STATES

In the previous sections, we have discussed experimental data and models that describe local conformational preferences in the unfolded state. However, it is clear that the unfolded state ensemble also comprises less or more stable long-range contacts that may or may not be similar to the native state. These long-range contacts may be induced by long-range hydrophobic or electrostatic interactions and, hence, not be predictable from strictly local propensities of amino acid conformations. How short-range and long-range order play together to form stable native states is a question under intense debate. For example, the hierarchical complementarity principle⁷³ proposes that short-range interactions favorable for the native state also promote such favorable longer-range interactions at successive levels of the structural hierarchy of a protein.

A number of NMR techniques can be used to detect transient long-range contacts in unfolded proteins. For a comprehensive recent review of the respective advantages and disadvantages of these methods, see Ref. 74. Transient hydrophobic sidechain contacts in similar arrangement as in the native state were detected by NOEs already in the study of the urea-denatured 434 repressor.⁴ Similarly, relaxation data both on backbone and sidechain nuclei in numerous unfolded or truncated proteins^{12,13,75,76} have indicated restricted mobility due to long-range hydrophobic contacts. These relaxation studies often have been combined with mutational studies to validate the importance of particular sidechain interactions.

Paramagnetic relaxation enhancements

A particularly promising method for the sensitive detection of both native and non-native long-range interactions in unfolded proteins is PRE, the accelerated relaxation of NMR coherences by a paramagnetic spin label, which is usually attached covalently to the protein. Like NOEs, PREs have an r^{-6} distance dependence. However, due to the large gyro-

magnetic ratio of the electron, the effect is detectable over lengths larger than 20 Å, thus providing atomic distance information on a length scale that is highly relevant to distance distributions in unfolded protein domains and comparable to the radius of gyration in small protein domains. Long-range contacts have been detected by PREs in the denatured states of, e.g., $\Delta 131\Delta$,¹⁷ ACBP,^{77,78} α -synuclein,^{49,79} and the drk SH3 domain.⁸⁰ These PREs in unfolded proteins require an ensemble interpretation, as smaller distances in the ensemble will dominate the signal as a result of the r^{-6} dependence. Thus, imposing all distance restraints on a single conformer artificially favors compact structures.⁸¹ Ensemble averaging of several structures refined simultaneously by PRE-restrained Monte Carlo or molecular dynamics sampling has been applied by Vendruscolo and co-workers to detect both native and non-native contacts in unfolded ensembles of ACBP,⁷⁷ $\Delta 131\Delta$,⁸¹ and α -synuclein.⁷⁹ Likewise, native and non-native interactions in the drk SH3 domain have been detected by Marsh *et al.*⁸⁰ by selecting an ensemble of conformations fitting the PRE data from a set of pregenerated conformers. A similar approach has been applied to the analysis of small-angle x-ray scattering (SAXS) data by selecting fitting subsets from a large number of conformations via a genetic algorithm.⁸²

RDCs and long-range order

RDCs in unfolded proteins have also proven to be very effective reporters of long-range contacts. Even short-range dipolar couplings are affected by such interactions, because chain topology and alignment properties of the unfolded state ensemble change due to the presence of long-range interactions. Thus, studies on acid-denatured ACBP have shown that a mutation I27A alters the RDC profile even in distant parts, presumably due to a reduction of long-range interactions.⁵³ Strong variations in the RDC profile of α -synuclein⁴⁹ as a result of hydrophobic long-range interactions have been mentioned above. These interactions could also be corroborated by PREs and by the FM approach,⁸³ which showed that the experimental RDC profile could only be well reproduced when the structural ensemble was restricted to such conformers that had contacts between the N- and C-terminal regions (Fig. 4). The strong variations of the RDC profile of α -synuclein disappear upon addition of denaturant, apparently due to a screening of the long-range interactions. Loss of the interactions also occurs at elevated temperatures or upon polyamine binding, which both favor aggregation of α -synuclein *in vitro*. In addition, the mutations A30P and A53T, which are associated with increased neurotoxicity in familial forms of Parkinson's disease, have a similar effect.⁸⁴ This has suggested a protective role of the long-range interactions in α -synuclein against misfolding and aggregation.

A more direct approach to obtain long-range order from RDCs is the detection of long-range $^1\text{H}^{\text{N}}-^1\text{H}^{\text{N}}$ RDCs on the background of a deuterated protein. The deuteration reduces losses by dipolar truncation and makes it possible to detect distances of up to 8 Å in folded proteins.^{85,86} The approach has been tested on denatured ubiquitin (8M urea at pH

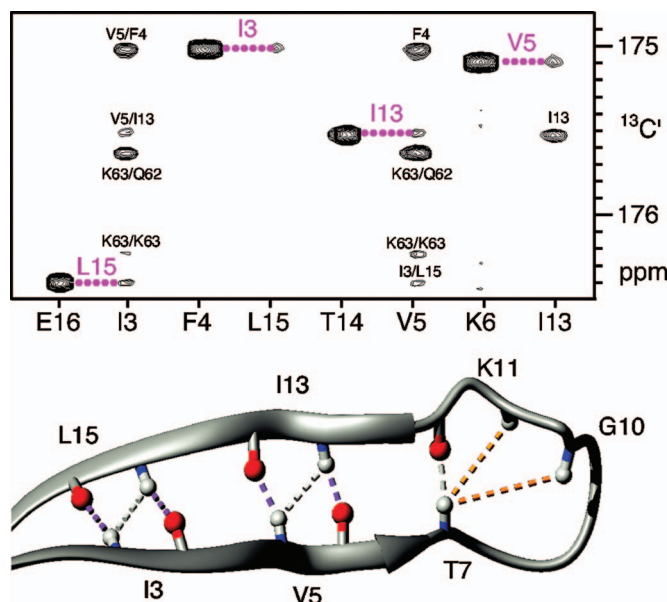


FIG. 9. (Color) Detection of hydrogen bonds in ubiquitin under highly denaturing conditions (8M urea). Top: $^1\text{H}/^{15}\text{N}$ strips from a three-dimensional long-range HNCQ-TROSY experiment showing $^3J_{\text{NC}'}$ correlations across H bonds. Peaks are labeled with assignment information. Hydrogen bond correlations are marked in magenta. Bottom: Summary of long-range interactions in the first hairpin of urea-denatured ubiquitin. Detected interactions are shown as dashed cylinders for $^3J_{\text{NC}'}$ (magenta) and D_{HNHN} (gold). Ambiguous correlations due to spectral overlap are shown in white. [Figure reprinted with permission from S. Meier, M. Strohmaier, M. Blackledge, and S. Grzesiek, *J. Am. Chem. Soc.* **129**, 754 (2007). Copyright 2007, Proceedings of the National Academy of Sciences of the United States of America].

2.5).⁵⁵ Besides a very large number of sequential $i, i+1$ and $i, i+2$ RDCs, also long-range RDCs could be detected that show a non-negligible population of native local structure in the N-terminal β -hairpin. This observation was substantiated by secondary chemical shifts, three-bond scalar couplings, and *trans*-hydrogen bond scalar couplings that establish the presence of closed hydrogen bonds across the β -hairpin (Fig. 9).

Combination of RDCs with complementary biophysical approaches

Mapping the vast conformational space available to unfolded peptide chains clearly presents significant experimental and analytical challenges. For this reason, it is essential to combine data from as many complementary biophysical approaches as possible in the quest to understand the nature of the unfolded state. One particular powerful experimental technique for characterizing the average overall dimensions of unfolded chains uses small-angle x-ray or neutron scattering.⁸⁷ Unfolded proteins have, thus, been shown to follow predictions from polymer theory for a random-coil chain, such that the radius of gyration $R_g = R_o N^{0.588}$, where N is the number of residues and $R_o = 2.0 \text{ \AA}$ is a constant that depends on the chain persistence length.⁸⁸ Over the past two decades, apparent contradictions between the random-coil nature of unfolded chains as detected from scattering experiments and evidence of local structure from spectroscopic techniques (in particular, from NMR) have resulted in the

so-called reconciliation problem.⁸⁹ It can simply be expressed as follows: If the polypeptide chain was, indeed, obeying conformational rules relevant to a random-coil chain, as small-angle scattering data appeared to demonstrate, how could this be reconciled with the detection of significant residual structure? While this is still an open question, the resolution of the apparent contradiction may simply lie in the specific sensitivities of the different techniques to long-range dimensions and local structure, respectively, and in the way the respective data are interpreted. A number of recent studies (e.g., Ref. 90 or 91) have, thus, presented atomic resolution models of denatured chains containing different degrees of local structure that do agree with the expected R_g power law described above. In this context, it is interesting to note that both the studies from Jha *et al.*²¹ and Bernado *et al.*²² demonstrated that reasonable reproduction of SAXS data can be derived from the same models that also reproduce experimental RDCs. Apparently, the local structure present in these ensembles does not contradict the level of extendedness detected using small-angle scattering.

PERSPECTIVES

Since the early 1990s, concepts inspired by physics have shifted views of the protein folding problem from the biochemical characterization of individual proteins towards the statistical mechanics aspects and the universal properties of polypeptide chains.⁹² The application of these theories has, however, been impeded by the lack of detailed data on the unfolded state. This situation is about to change. NMR-derived parameters on structure and dynamics at atomic resolution can now be tested against plausible models of unfolded proteins. Especially, RDCs and PREs obtained in solution have proven extremely useful for the description of local and long-range structures at high precision, and hold further promise for the detection of universal and protein-specific properties of the unfolded state.

Currently, the most successful model of the unfolded state is the coil model, which describes unfolded proteins by the conformations of nonsecondary structure regions in folded proteins. Using explicit all-atom model building, the coil library has been combined with restrictions from steric exclusion to reproduce RDCs and three-bond J -couplings reporting on local conformations and on overall chain properties. It is clear that unfolded backbone conformations are not uniform. They show sequence-specific biases, which are caused by the properties of individual amino acids via sidechain-sidechain and sidechain-backbone interactions. These conformational preferences can rationalize the finite experimental time scales of protein folding, as biased searches are much more efficient than full random searches. Despite the apparent success of statistical coil models for the prediction of local structure in unfolded chains, a number of points remain open.

At the moment, it is unclear how unique the description via the coil model is. Certainly, the problem of finding the best conformational distribution is ill-posed due to the limited number of experimental parameters; other models besides the coil model are possible. To rank such models and to

judge their compatibility with the experimental data, it will be necessary to develop robust mathematical criteria. It also remains to be explored whether a purely analytical description of conformational preferences can be derived from the experimental data that would offer further conceptual insights.

The current coil models also cannot predict long-range interactions in unfolded states. Such interactions comprising natively like structures such as β -hairpins or α -helices have been detected at significant population in many experimental studies even under highly denaturing conditions. The amount of such long-range, natively like interactions successively increases under less severe denaturing conditions.² An obvious next goal would be the successful prediction of such contacts. A possible first step in this direction could be the prediction of local secondary or supersecondary structure elements in conjunction with the already successful description of local conformations by the coil models.

Besides the definition of the conformational ensemble of a protein in the unfolded state, it is equally important to describe the transitions between the different conformations within this ensemble. Currently, predictions of these transitions and their time scales are only possible from extended molecular dynamics simulations.⁹³ The ability to define a statistical ensemble from RDCs, chemical shifts or scalar couplings, that average up to tens or hundreds of milliseconds, may provide further guidance to these molecular dynamics simulations and for the development of models that describe folding kinetics.

NMR also provides direct experimental data on the kinetics of folding via relaxation data that monitor local correlation times on the nanosecond time scale of Larmor frequencies and conformational exchange on the millisecond time scale of chemical shifts. Broadening of resonances by the latter phenomenon makes it possible to detect the presence of a few percent of non-native state under native, physiological conditions. Thus, relaxation dispersion experiments have proven sensitive enough to obtain populations, transition rates, and structural information on a lowly populated folding intermediate of the FYN SH3 domain in the absence of a denaturant.⁹⁴ Alternatively, optical spectroscopy can be used to obtain distance information between specific sites of lowly populated states under native conditions or in rapid mixing experiments. The high temporal resolution and sensitivity of optical spectroscopy provide an ideal complement to NMR for the detection of conformational fluctuations and intrachain contact formation rates on the femtosecond to millisecond time scale.⁹⁵⁻⁹⁸

Clearly, NMR experimental method development for the characterization of the unfolded state has not reached an end point. In particular, it seems possible to derive many more experimental parameters per amino acid by the characterization of sidechains and the use of additional alignment media for RDC experiments. This may provide crucial input for the refinement of current models. For RDCs, it will also be necessary to better characterize and understand the interactions with the alignment media and the exact nature of the averaging process. A further experimental route that is only starting to emerge is the characterization of unfolded proteins in

the solid state. NMR experiments on glassy frozen solutions have recently been used by Havlin and Tycko⁹⁹ to characterize the conformations of the chemically denatured villin headpiece subdomain. Site-specific conformational distributions could be obtained by line shape analysis of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ resonances at different denaturant concentrations to monitor the unfolding pathway of the protein.

In all described experimental methods, a variation of amino acid sequences may provide a handle to a systematic characterization of chain properties. In particular, this approach could be used to experimentally characterize smaller parts of an entire protein sequence in order to detect conformational preferences for secondary or supersecondary structure. It could then be explored whether the entire structure can be constructed from the combination of the conformational distributions of these local elements. This would effectively break the problem of building a folded structure of an entire protein apart into the more tractable problem of characterizing the conformational preferences of its constituents. Similar approaches of fragment assembly are used by the currently most successful methods for predicting protein structure from sequence alone.¹⁰⁰⁻¹⁰²

ACKNOWLEDGMENTS

We would like to thank Dr. R. Tycko, Dr. D. Torchia, Dr. A. Bax, Dr. S. Dames, and Dr. J. Huang for valuable discussions. This work was supported by SNF Grant No. 31-109712 (S.G.), the Carlsbergfond (S.M.), the CEA, the CNRS, the UJF Grenoble, and Grant No. NT05-4_42781 of the Agence Nationale de Recherche (M.B.).

¹G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16623 (2006).

²D. Shortle, *FASEB J.* **10**, 27 (1996).

³C. Tanford, K. Kawahara, and S. Lapanje, *J. Biol. Chem.* **241**, 1921 (1966).

⁴D. Neri, M. Billeter, G. Wider, and K. Wuthrich, *Science* **257**, 1559 (1992).

⁵A. T. Alexandrescu, C. Abeygunawardana, and D. Shortle, *Biochemistry* **33**, 1063 (1994).

⁶R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 20 (1992).

⁷C. Levinthal, *J. Chim. Phys. Phys.-Chim. Biol.* **65**, 44 (1968).

⁸J. A. Marsh, V. K. Singh, Z. Jia, and J. D. Forman-Kay, *Protein Sci.* **15**, 2795 (2006).

⁹H. Roder, G. Wagner, and K. Wuthrich, *Biochemistry* **24**, 7407 (1985).

¹⁰A. Miranker, S. E. Radford, M. Karplus, and C. M. Dobson, *Nature (London)* **349**, 633 (1991).

¹¹H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson, *Biochemistry* **36**, 8977 (1997).

¹²S. Schwarzinger, P. E. Wright, and H. J. Dyson, *Biochemistry* **41**, 12681 (2002).

¹³J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Ducharndt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe, *Science* **295**, 1719 (2002).

¹⁴W. Y. Choy, D. Shortle, and L. E. Kay, *J. Am. Chem. Soc.* **125**, 1748 (2003).

¹⁵L. Serrano, *J. Mol. Biol.* **254**, 322 (1995).

¹⁶L. J. Smith, K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton, and C. M. Dobson, *J. Mol. Biol.* **255**, 494 (1996).

¹⁷J. R. Gillespie and D. Shortle, *J. Mol. Biol.* **268**, 170 (1997).

¹⁸D. Shortle and M. S. Ackerman, *Science* **293**, 487 (2001).

- ¹⁹ S. Meier, S. Grzesiek, and M. Blackledge, *J. Am. Chem. Soc.* **129**, 9799 (2007).
- ²⁰ N. C. Fitzkee, P. J. Fleming, and G. D. Rose, *Proteins* **58**, 852 (2005).
- ²¹ A. K. Jha, A. Colubri, K. F. Freed, and T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13099 (2005).
- ²² P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok, and M. Blackledge, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17002 (2005).
- ²³ K. M. Fiebig, H. Schwalbe, M. Buck, L. J. Smith, and C. M. Dobson, *J. Phys. Chem.* **100**, 2661 (1996).
- ²⁴ A. Bundi and K. Wüthrich, *Biopolymers* **18**, 285 (1979).
- ²⁵ M. B. Swindells, M. W. MacArthur, and J. M. Thornton, *Nat. Struct. Biol.* **2**, 596 (1995).
- ²⁶ T. Massad, J. Jarvet, R. Tanner, K. Tomson, J. Smirnova, P. Palumaa, M. Sugai, T. Kohno, K. Vanatala, and P. Damberg, *J. Biomol. NMR* **38**, 107 (2007).
- ²⁷ O. W. Howarth and D. M. J. Lilley, *Prog. Nucl. Magn. Reson. Spectrosc.* **12**, 1 (1978).
- ²⁸ S. Spera and A. Bax, *J. Am. Chem. Soc.* **113**, 5490 (1991).
- ²⁹ D. Wishart, B. Sykes, and F. Richards, *Biochemistry* **31**, 1647 (1992).
- ³⁰ D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes, *J. Biomol. NMR* **5**, 67 (1995).
- ³¹ Y. Wang and O. Jardetzky, *Protein Sci.* **11**, 852 (2002).
- ³² M. D. Mukrasch, J. Biernat, M. von Bergen, C. Griesinger, E. Mandelkow, and M. Zweckstetter, *J. Biol. Chem.* **280**, 24978 (2005).
- ³³ N. Tjandra and A. Bax, *Science* **278**, 1111 (1997).
- ³⁴ J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9279 (1995).
- ³⁵ J. H. Prestegard, C. M. Bougault, and A. I. Kishore, *Chem. Rev. (Washington, D.C.)* **104**, 3519 (2004).
- ³⁶ M. R. Hansen, L. Mueller, and A. Pardi, *Nat. Struct. Biol.* **5**, 1065 (1998).
- ³⁷ G. M. Clore, M. R. Starich, and A. M. Gronenborn, *J. Am. Chem. Soc.* **120**, 10571 (1998).
- ³⁸ M. Ruckert and G. Otting, *J. Am. Chem. Soc.* **122**, 7793 (2000).
- ³⁹ R. Tycko, F. J. Blanco, and Y. Ishii, *J. Am. Chem. Soc.* **122**, 9340 (2000).
- ⁴⁰ H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, and S. Grzesiek, *J. Biomol. NMR* **18**, 303 (2000).
- ⁴¹ G. Salamat and E. W. Kaler, *Langmuir* **15**, 5414 (1999).
- ⁴² M. Zweckstetter and A. Bax, *J. Biomol. NMR* **20**, 365 (2001).
- ⁴³ C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980), Vol. 2.
- ⁴⁴ J. Tolman, J. Flanagan, M. Kennedy, and J. Prestegard, *Nat. Struct. Biol.* **4**, 292 (1997).
- ⁴⁵ J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, and R. Bruschweiler, *J. Am. Chem. Soc.* **123**, 6098 (2001).
- ⁴⁶ P. Bernado and M. Blackledge, *J. Am. Chem. Soc.* **126**, 4907 (2004).
- ⁴⁷ M. Louhivuori, R. Otten, T. Salminen, and A. Annala, *J. Biomol. NMR* **39**, 141 (2007).
- ⁴⁸ S. Ohnishi and D. Shortle, *Proteins* **50**, 546 (2003).
- ⁴⁹ C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, and M. Zweckstetter, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1430 (2005).
- ⁵⁰ S. A. Dames, R. Aregger, N. Vajpai, P. Bernado, M. Blackledge, and S. Grzesiek, *J. Am. Chem. Soc.* **128**, 13508 (2006).
- ⁵¹ L. Skora, M. K. Cho, H. Y. Kim, S. Becker, C. O. Fernandez, M. Blackledge, and M. Zweckstetter, *Angew. Chem., Int. Ed.* **45**, 7012 (2006).
- ⁵² R. Mohana-Borges, N. K. Goto, G. J. Kroon, H. J. Dyson, and P. E. Wright, *J. Mol. Biol.* **340**, 1131 (2004).
- ⁵³ W. Fieber, S. Kristjansdottir, and F. M. Poulsen, *J. Mol. Biol.* **339**, 1191 (2004).
- ⁵⁴ S. Meier, S. Guthe, T. Kiefhaber, and S. Grzesiek, *J. Mol. Biol.* **344**, 1051 (2004).
- ⁵⁵ S. Meier, M. Strohmeier, M. Blackledge, and S. Grzesiek, *J. Am. Chem. Soc.* **129**, 754 (2007).
- ⁵⁶ M. D. Mukrasch, M. von Bergen, J. Biernat, D. Fischer, C. Griesinger, E. Mandelkow, and M. Zweckstetter, *J. Biol. Chem.* **282**, 12230 (2007).
- ⁵⁷ M. Louhivuori, K. Paakkonen, K. Fredriksson, P. Permi, J. Lounila, and A. Annala, *J. Am. Chem. Soc.* **125**, 15647 (2003).
- ⁵⁸ M. Louhivuori, K. Fredriksson, K. Paakkonen, P. Permi, and A. Annala, *J. Biomol. NMR* **29**, 517 (2004).
- ⁵⁹ O. I. Obolensky, K. Schlepckow, H. Schwalbe, and A. V. Solov'yov, *J. Biomol. NMR* **39**, 1 (2007).
- ⁶⁰ S. Ohnishi, A. L. Lee, M. H. Edgell, and D. Shortle, *Biochemistry* **43**, 4064 (2004).
- ⁶¹ K. Ding, J. M. Louis, and A. M. Gronenborn, *J. Mol. Biol.* **335**, 1299 (2004).
- ⁶² C. O. Sallum, D. M. Martel, R. S. Fournier, W. M. Matousek, and A. T. Alexandrescu, *Biochemistry* **44**, 6392 (2005).
- ⁶³ N. Sibille, A. Sillen, A. Leroy, J. M. Wieruszkeski, B. Mulloy, I. Landrieu, and G. Lippens, *Biochemistry* **45**, 12560 (2006).
- ⁶⁴ M. D. Mukrasch, P. Markwick, J. Biernat, M. V. Bergen, P. Bernado, C. Griesinger, E. Mandelkow, M. Zweckstetter, and M. Blackledge, *J. Am. Chem. Soc.* **129**, 5235 (2007).
- ⁶⁵ A. T. Alexandrescu and R. A. Kammerer, *Protein Sci.* **12**, 2132 (2003).
- ⁶⁶ M. Zweckstetter and A. Bax, *J. Am. Chem. Soc.* **122**, 3791 (2000).
- ⁶⁷ S. Lovell, I. Davis, W. Arendall, P. de Bakker, J. Word, M. Prisant, J. Richardson, and D. Richardson, *Proteins* **50**, 437–450 (2003).
- ⁶⁸ A. Almond and J. B. Axelsen, *J. Am. Chem. Soc.* **124**, 9986 (2002).
- ⁶⁹ M. K. Cho, H. Y. Kim, P. Bernado, C. O. Fernandez, M. Blackledge, and M. Zweckstetter, *J. Am. Chem. Soc.* **129**, 3032 (2007).
- ⁷⁰ J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, *J. Am. Chem. Soc.* **129**, 1179 (2007).
- ⁷¹ J. Sass, F. Cordier, A. Hoffmann, A. Cousin, J. G. Omichinski, H. Lowen, and S. Grzesiek, *J. Am. Chem. Soc.* **121**, 2047 (1999).
- ⁷² E. B. Gebel, K. Ruan, J. R. Tolman, and D. Shortle, *J. Am. Chem. Soc.* **128**, 9310 (2006).
- ⁷³ N. Fitzkee, P. Fleming, H. Gong, N. Panasik, T. Street, and G. Rose, *Trends Biochem. Sci.* **30**, 73 (2005).
- ⁷⁴ T. Mittag and J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **17**, 3 (2007).
- ⁷⁵ W. Y. Choy, and L. E. Kay, *J. Am. Chem. Soc.* **125**, 11988 (2003).
- ⁷⁶ W. Y. Choy, F. A. Mulder, K. A. Crowhurst, D. R. Muhandiram, I. S. Millett, S. Doniach, J. D. Forman-Kay, and L. E. Kay, *J. Mol. Biol.* **316**, 101 (2002).
- ⁷⁷ K. Lindorff-Larsen, S. Kristjansdottir, K. Teilum, W. Fieber, C. M. Dobson, F. M. Poulsen, and M. Vendruscolo, *J. Am. Chem. Soc.* **126**, 3291 (2004).
- ⁷⁸ S. Kristjansdottir, K. Lindorff-Larsen, W. Fieber, C. M. Dobson, M. Vendruscolo, and F. M. Poulsen, *J. Mol. Biol.* **347**, 1053 (2005).
- ⁷⁹ M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, *J. Am. Chem. Soc.* **127**, 476 (2005).
- ⁸⁰ J. A. Marsh, C. Neale, F. E. Jack, W. Y. Choy, A. Y. Lee, K. A. Crowhurst, and J. D. Forman-Kay, *J. Mol. Biol.* **367**, 1494 (2007).
- ⁸¹ C. J. Francis, K. Lindorff-Larsen, R. B. Best, and M. Vendruscolo, *Proteins* **65**, 145 (2006).
- ⁸² P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, *J. Am. Chem. Soc.* **129**, 5656 (2007).
- ⁸³ P. Bernado, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge, *J. Am. Chem. Soc.* **127**, 17968 (2005).
- ⁸⁴ C. Bertoncini, C. Fernandez, C. Griesinger, T. Jovin, and M. Zweckstetter, *J. Biol. Chem.* **280**, 30649 (2005).
- ⁸⁵ Z. Wu and A. Bax, *J. Am. Chem. Soc.* **124**, 9672 (2002).
- ⁸⁶ S. Meier, D. Haussinger, P. Jensen, M. Rogowski, and S. Grzesiek, *J. Am. Chem. Soc.* **125**, 44 (2003).
- ⁸⁷ J. Lipfert and S. Doniach, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307 (2007).
- ⁸⁸ J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dohager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, and K. W. Plaxco, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491 (2004).
- ⁸⁹ I. S. Millett, S. Doniach, and K. W. Plaxco, *Adv. Protein Chem.* **62**, 241 (2002).
- ⁹⁰ N. C. Fitzkee and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12497 (2004).
- ⁹¹ H. T. Tran and R. V. Pappu, *Biophys. J.* **91**, 1868 (2006).
- ⁹² E. Shakhnovich, *Chem. Rev. (Washington, D.C.)* **106**, 1559 (2006).
- ⁹³ W. F. van Gunsteren, R. Burgi, C. Peter, and X. Daura, *Angew. Chem., Int. Ed.* **40**, 351 (2001).
- ⁹⁴ D. M. Korzhnev, X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson, and L. E. Kay, *Nature (London)* **430**, 586 (2004).
- ⁹⁵ O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9597 (1999).
- ⁹⁶ A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgartel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 105 (2007).
- ⁹⁷ D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2655 (2007).
- ⁹⁸ T. A. Laurence, X. Kong, M. Jager, and S. Weiss, *Proc. Natl. Acad. Sci.*

- U.S.A. **102**, 17348 (2005).
- ⁹⁹R. H. Havlin and R. Tycko, Proc. Natl. Acad. Sci. U.S.A. **102**, 3284 (2005).
- ¹⁰⁰K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, J. Mol. Biol. **268**, 209 (1997).
- ¹⁰¹G. Chikenji, Y. Fujitsuka, and S. Takada, Proc. Natl. Acad. Sci. U.S.A. **103**, 3141 (2006).
- ¹⁰²D. Shortle, Protein Sci. **11**, 18 (2002).