JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Sequence-Specific Mapping of the Interaction between Urea and Unfolded Ubiquitin from Ensemble Analysis of NMR and Small Angle Scattering Data

Jie-rong Huang,[†,‡,§] Frank Gabel,[†,‡,§] Malene Ringkjøbing Jensen,[†,‡,§] Stephan Grzesiek,[∥] and Martin Blackledge*[,†,‡,§]

[†]CEA, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, Grenoble 38027, France

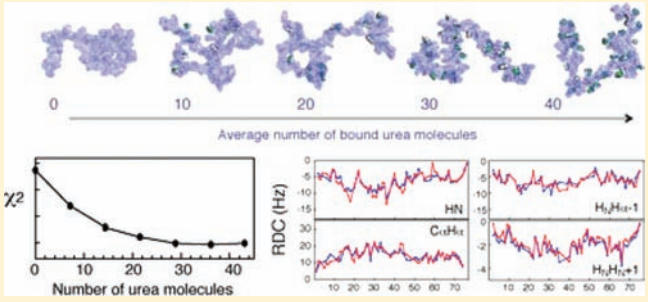
[‡]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, Grenoble 38027, France

[§]UJF-Grenoble 1, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, Grenoble 38027, France

[∥]Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

Supporting Information

ABSTRACT: The molecular details of how urea interacts with, and eventually denatures proteins, remain largely unknown. In this study we have used extensive experimental NMR data, in combination with statistical coil ensemble modeling and small-angle scattering, to analyze the conformational behavior of the protein ubiquitin in the presence of urea. In order to develop an atomic resolution understanding of the denatured state, conformational ensembles of full-atom descriptions of unfolded proteins, including side chain conformations derived from rotamer libraries, are combined with random sampling of explicit urea molecules in interaction with the protein. Using this description of the conformational equilibrium, we demonstrate that the direct-binding model of urea to the protein backbone is compatible with available experimental data. We find that, in the presence of 8 M urea, between 30 and 40% of the backbone peptide groups bind a urea molecule, independently reproducing results from a model-free analysis of small-angle neutron and X-ray scattering data. Crucially, this analysis also provides sequence specific details of the interaction between urea and the protein backbone. The pattern of urea-binding along the amino-acid sequence reveals a higher level of binding in the central part of the protein, a trend which resembles independent results derived from chemical shift mapping of the urea−protein interaction. Together these results substantiate the direct-binding model and provide a framework for studying the physical basis of interactions between proteins and solvent molecules.



INTRODUCTION

Characterizing the conformational energy landscape of unfolded proteins helps us to understand the protein folding problem,[1] disease-related protein misfolding events,[2] and the relation between protein sequence and function in intrinsically disordered proteins.[3] The small hydrophilic osmolyte urea is commonly used to study the behavior of unfolded proteins because of its so-called "good" solvent characteristics.[4] Although urea is widely used as a denaturant in biophysical and biochemical experiments, the molecular details of how urea actually unfolds proteins are largely unknown. Two different mechanisms have been proposed, relying either on a "direct" interaction between urea and the protein backbone via hydrogen-bonds[5] or an "indirect" interaction whereby urea disrupts the water structure surrounding the protein and thereby acts as a better solvent to hydrophobic groups.[6] The resolution of this question will provide important insight into the molecular basis of protein stability and remains a key challenge for the interpretation of protein folding studies.

Although molecular dynamics (MD) simulations can in principle provide an atomic resolution description of the basic mechanism of protein-denaturation,[7−17] there is still concern about the ability of currently available potential energy force fields to correctly capture the behavior of unfolded proteins in solution. These problems probably stem from the increased relevance of accurate solvent descriptions compared to folded proteins, against which recently improved force fields have been gauged,[18−21] and the importance of fluctuating weak interactions such as solvent−solute hydrogen bonding that contribute significantly to the behavior of disordered proteins in solution. Although there has been considerable progress in the development of specific force fields for the unfolded state,[22−24] unfolding events are expected to occur on time scales that are rarely accessed by state-of-the-art computational power, so that it still remains challenging to ensure efficient

Received: December 20, 2011
Published: February 6, 2012

   dx.doi.org/10.1021/ja2118688 | J. Am. Chem. Soc. 2012, 134, 4429−4436

sampling of the conformational energy phase space available to the unfolded state.

Nuclear magnetic resonance (NMR) spectroscopy provides a unique experimental tool for the study of unfolded proteins, providing time and ensemble averages of parameters reporting on local and long-range structure over a broad range of time scales. While under certain conditions it may be possible to measure cross-relaxation rates between protons belonging to the protein and the urea, such approaches can be complicated by a number of factors, including the possibly short residence times of the solvent, hydrogen-exchange, and long-range dipolar interactions between bulk and bound solvent.[25−28] The ability to map the conformational behavior of urea-unfolded proteins using residual dipolar couplings (RDCs) measured in partially aligned samples has recently been demonstrated for the proteins ubiquitin and GB1.[29,30] A popular method for the interpretation of experimental data applies restrained molecular dynamics simulations using a hybrid potential energy function to drive multiple copies of the protein into an ensemble of structures that are in agreement with experimental data.[31−34] Alternatively, over the past few years, we have developed a statistical coil model[35−37] to progressively map and comprehend the conformational energy landscape of the unfolded state. Using a multistep procedure, we first define expected values of experimentally accessible parameters, such as RDCs, for a disordered protein sequence that samples amino-acid specific backbone dihedral angle potential energy surfaces. These simple predictions demonstrated good agreement between the theoretical and experimental data, thereby validating this computational approach to the construction of disordered protein ensembles. The ability to predict experimental parameters from the unfolded state then provides the basis for identifying and characterizing the presence of transiently populated secondary structural elements or long-range order in the otherwise disordered state.[38,39] In combination with enhanced conformational sampling from accelerated molecular dynamics simulation[40] and statistical ensemble sample-and-select procedures, we have demonstrated the ability of these ensemble approaches to accurately map the backbone conformational behavior of denatured and intrinsically disordered states.[41−45]

Here we use a novel approach to the interpretation of experimental data from urea-denatured proteins, generating conformational ensembles of all-atom descriptions of the unfolded protein, including side chain rotamer conformations, and combining these to develop a molecular model of urea molecules in interaction with the protein. This model is used in combination with extensive experimental NMR and small-angle X-ray scattering (SAXS) data to test the relevance of the direct binding model of urea to the protein backbone and to identify the binding propensities of each amino acid along the protein sequence.

RESULTS

Incorporating Side Chains into the Statistical Coil Model. To incorporate van der Waals type interactions into the statistical coil model, it is essential to correctly describe amino acid side chains. This is achieved here by modifying the Flexible-MECCANO approach[36] to build explicit side chains in conformations randomly selected from known rotameric libraries (see Materials and Methods). Comparison of RDCs, Ramachandran plots, and the distribution of the radius of gyration (R_g) shows that the existing local and global sampling characteristics of Flexible-MECCANO are not significantly perturbed by the addition of side chains (Figure S1 of the Supporting Information).

Modeling the Protein−Urea Interaction. The physical basis of the interaction between urea and the protein is the subject of debate between the "indirect" mechanism, where urea disrupts the surrounding water structure, and the "direct" mechanism, where urea forms hydrogen bonds (H-bonds) with the protein backbone. There is increasing evidence from both simulation and experimental data in support of the direct-binding mechanism.[7,10,12,14,46,47] MD simulation shows that urea molecules have a higher propensity to bind to the protein backbone than to side chains,[7,14,48−50] although simulation results conflict as to whether urea acts as an H-bond donors or acceptor. Using a recent, long time scale (1 μs) MD simulation of 8 M urea-denatured lysozyme, Berne and co-workers found that urea preferentially bound to the amide proton of the protein backbone through its oxygen atom.[7] Parinello and co-workers recently combined parallel tempering and metadynamics simulation to assess the denaturing effect of urea on the β-hairpin of protein GB1, and they found evidence that urea directly interacts with the protein backbone through H-bonds.[15] We note that a recent experimental study using vibrational sum frequency spectroscopy, which directly probes the interfacial interaction of urea molecules, demonstrated that the carbonyl group of the urea molecule has a tendency to be directed toward the protein surface at lower pH and to adopt the inverse orientation at higher pH.[51] The relevance of this observation for our model of urea-denaturation of ubiquitin is underlined by the experimental observation that ubiquitin is not unfolded in 8 M urea unless it is studied at acid pH. Finally, our recent model-free analysis of SAXS/SANS data on urea-denatured ubiquitin also supports the hypothesis that urea directly binds to the protein backbone.[52]

To investigate whether the direct-binding mode of interaction is compatible with a large quantity of experimental data, we have built explicit molecular ensembles of the protein ubiquitin bound to urea, treating urea molecules as H-bond acceptors and constructing them onto the amide protons of the protein backbone. The urea molecule geometry is based on published neutron diffraction measurements.[53] The H-bond length is fixed at 2.5 Å, and the angle (N−H···O) is fixed at 180°. The urea molecules can randomly rotate along the H-bond axis (Figure 1A). We have tested different physically reasonable H-bond conformations[54] and find similar results in terms of predicted NMR data and R_g values (Figure S2 of the Supporting Information). Urea molecules are added randomly on specific binding sites according to a given saturation rate (the percentage of potential binding sites populated per solute molecule). The molecules are added after a residue has been built. An all-atom steric clash checking function is performed again after adding one urea molecule. If there is clash between the latest added urea molecule and the built atoms, the algorithm returns one residue and rebuilds with different ϕ/ψ angles in the statistical coil database.

Determination of the Urea−Protein Binding Rate from Experimental RDCs. Ensembles were calculated by applying different saturation rates, ranging from 0% to 60%. With increasing number of urea molecules, the $\langle R_g \rangle$ increases from 26.8 Å (without urea) to 34.8 Å (~43 urea molecules at a saturation rate of 60%, Table 1 and Figure S3A of the Supporting Information). Not surprisingly, the population of the extended regions (βP and βS; see legend, Table 1)

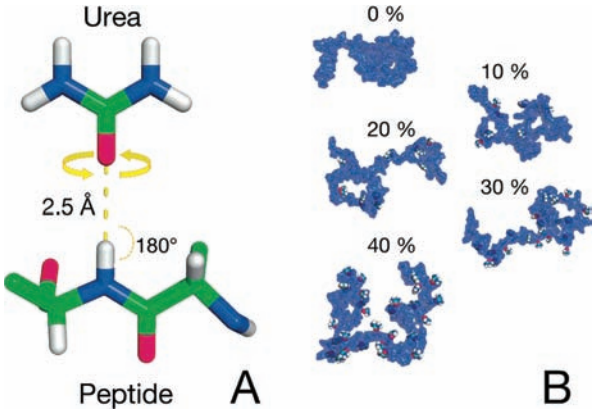


Figure 1. Illustrative representation of the urea-binding model. (A) Urea binding conformation used in the main text. The H-bond length is fixed at 2.5 Å, and the H-bond angle N−H···O is fixed at 180°. The urea molecule can randomly rotate along the axis of the H-bond. (B) Representative urea sampling with different saturation rates for each binding site from zero urea to 40%.

increases with the number of bound urea molecules while the population in the αR and αL regions decreases (Figure S3B of the Supporting Information and Table 1). Theoretical RDCs calculated from ensembles with different saturation rates of urea are compared to experimental data (Figure 2A). Addition of urea molecules to the statistical coil model induces the sampling of more extended ϕ/ψ angles as observed in previous studies,[55,56] so that D_{HNHN+1} and D_{HNHN+2} are significantly reduced with increasing urea saturation rate. The χ^2 analysis shows that, with increasing urea saturation rate, the agreement between experimental and predicted values is improved and the χ^2 reaches a minimum when an average of approximately 30 urea molecules are bound to each protein (corresponding to a saturation rate of 40%, Figure 2B). The χ^2 value obtained at this saturation rate is similar to that calculated from previous hypothesis-driven approaches, where the extended region of Ramachandran space was progressively sampled more than the standard coil library (around 80% in the extended region).[56] We also note that, in our previous "model-free" approach for analyzing SAXS/SANS data from urea-unfolded ubiquitin, we found that around 10 urea molecules were bound to ubiquitin at pH 6.0, and a further 20 urea molecules were recruited when the pH was reduced to 2.5.[52]

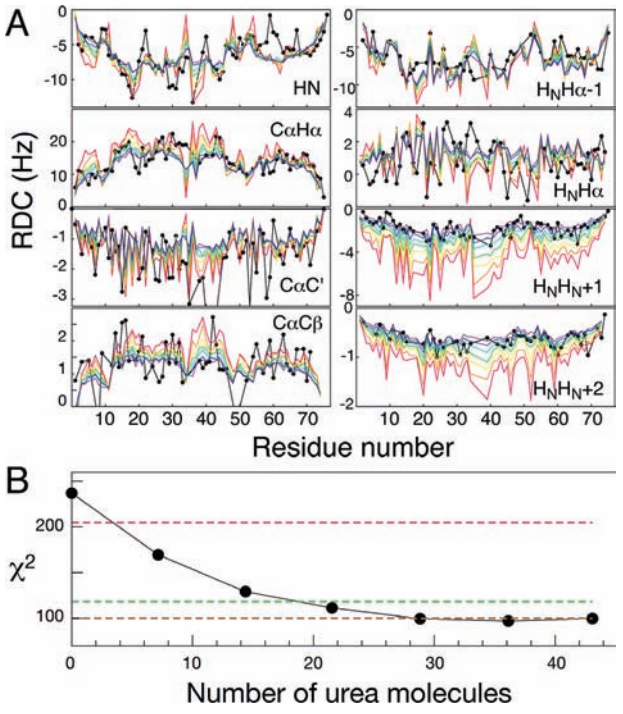


Figure 2. Analysis of predicted RDCs and experimental data for different models. (A) Comparison of experimental RDCs for urea-denatured ubiquitin and predicted values from different percentages of urea saturation rates. Experimental data are shown as black points, and predicted values are color coded as follows: 0% (red), 10% (orange), 20% (yellow), 30% (green), 40% (blue), 50% (indigo), 60% (purple). (B) χ^2 calculated from the different urea binding models and those generated from different levels of extendedness. χ^2 from the urea binding model along with the number of urea molecules is shown in solid black circles and lines. χ^2 values from different levels of extendedness are shown in dashed lines: 59.0% (red), 74.2% (green), 81.2% (blue), and 85.2% (yellow). χ^2 values are not normalized, and the total number of RDCs considered in this figure is 458.

Experimental RDCs Define a Representative Ensemble with a Site-Specific Urea Binding Pattern. This analysis shows that improved reproduction between experimental and theoretical RDCs is obtained by randomly recruiting approximately 30 urea molecules onto the statistical coil model of the peptide chain. To develop a site-specific model of urea binding, we then use the genetic algorithm ASTEROIDS,[42,44] in combination with the experimental RDCs

Table 1. Comparison of Physical Properties of Different Urea-Binding Models[a]

rate[b]	# of urea[c]	$\langle R_g \rangle$ (Å)	$p(\beta P)$ (%)[d]	$p(\beta S)$ (%)	$p(\alpha L)$ (%)	$p(\alpha R)$ (%)
0%	0	26.82	29.6	37.2	4.3	28.9
10%	7.2 ± 2.5	27.74	30.8	38.8	4.2	26.2
20%	14.4 ± 3.4	28.91	31.9	40.6	4.0	23.5
30%	21.5 ± 3.9	30.03	33.3	42.4	3.7	20.6
40%	28.2 ± 4.1	31.59	34.6	44.3	3.3	17.8
50%	36.4 ± 4.0	33.02	36.0	46.2	2.9	14.9
60%	43.1 ± 4.1	34.83	37.4	48.2	2.5	12.0
ASTEROIDS[e]	23.3 ± 18.0	31.90	33.4	42.2	3.9	20.5

[a]To quantify the ϕ/ψ angle distribution in different urea binding ensemble models, the Ramachandran space is divided into four parts: (αL) $\phi > 0°$; (αR) $\phi < 0°$, $-120° < \psi < 50°$; (βP) $-90° < \phi < 0°$, $\psi > 50°$, or $\psi < -120°$; (βS) $-180° < \phi < -90°$, $\psi > 50°$, or $\psi < -120°$. [b]Saturation rate for sampling a urea molecule on the binding site of a protein. [c]Number of urea molecules averaged over entire ensemble at such a saturation rate. [d]Population of ϕ/ψ angles of an ensemble model separated in the Ramachandran space. [e]Selected ensemble against experimental RDCs from a pool generated with a randomly assigned rate.

dx.doi.org/10.1021/ja2118688 | J. Am. Chem. Soc. 2012, 134, 4429−4436

(in total 458) to select a representative ensemble of urea-bound conformers. A pool of 50,000 structures with a randomly assigned saturation rate of urea molecules, from an occupancy of 0 to 100% of the available binding sites, was constructed. The similar average number of urea molecules on each binding site (Figure S4A of the Supporting Information) in the pool indicates that no bias for sampling urea molecule is induced from the prebuilt polypeptide. The number of urea molecules that one model structure possesses is also well distributed (Figure S4B of the Supporting Information) except those structures having more than 65 urea molecules, which are slightly less populated because of the difficulty in constructing such extended chains. Two hundred structures are selected from a pool of 50,000 structures with a randomly assigned number of urea molecules bound to randomly selected sites along the backbone. The genetic algorithm ASTEROIDS is used as described previously to make this selection.[42,44] The comparison of RDCs calculated from the selected ensemble and experimental data is shown in Figure 3. In order to estimate

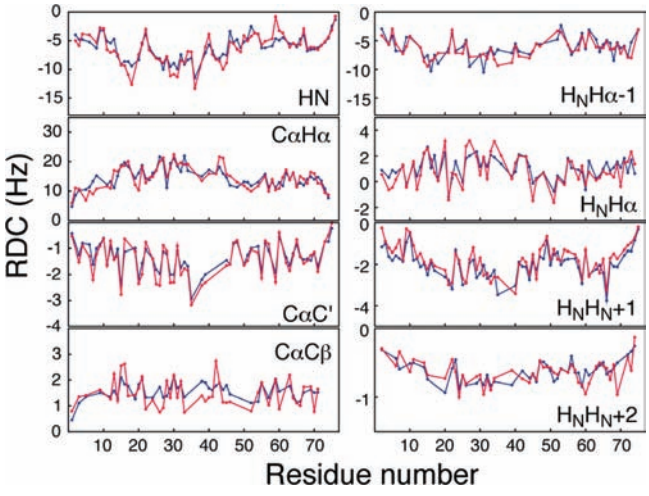


Figure 3. Comparison of ASTEROIDS selected results and experimental residual dipolar couplings. The experimental data are shown in red lines, and back-calculated RDCs from selected structures are shown in blue.

the uncertainty in the procedure, the selection was repeated 50 times. A consistency check was performed, by removing 5% of the experimental couplings to validate the number of structures present in the subensemble (Figure S5 of the Supporting Information).

The backbone dihedral angle distribution and $\langle R_g \rangle$ calculated from the selected ensemble are consistent with the calculations from the ensemble with a 40% urea saturation rate (Table 1). Notably, the average number of urea molecules drops from around 30 as determined using the nonspecific approach described above, to around 23 for the selected ensembles that are in best agreement with the experimental RDCs (Table 1). This small discrepancy results from site-specific analysis and reflects the nonuniform binding along the chain. The average number of urea molecules on each binding site was determined from the selected ensemble, revealing an intriguing pattern of urea binding along the amino-acid sequence that is apparently nonrandom (Figure 4A). The most noticeable trend is that residues 20−25 and the C-terminal part of urea-denatured ubiquitin recruit the fewest urea molecules while the central part of the protein recruits more urea compared to the average

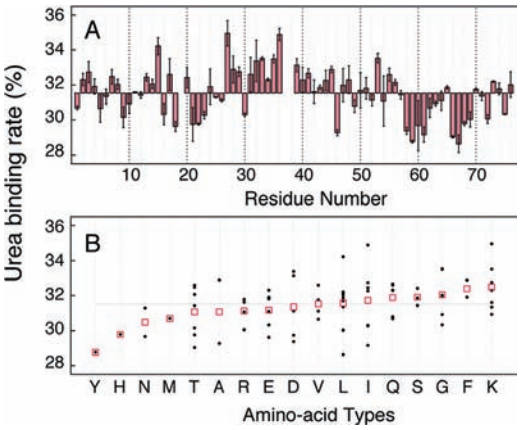


Figure 4. Distribution of urea molecules. (A) Normalized number of urea molecules at each binding side from the ASTEROIDS selection plotted with respect to the primary sequence (averaged over 50 selections). The normalized number of urea molecules at each binding side in the pool distribution is 50%. The error is propagated from the deviation of two independent results of 25 selections. (B) The same number as in Figure 4A, plotted against amino-acid types: Black dots represent the value for each residue, and red squares represent the averaged values for every amino-acid type.

and compared to the pool. Amino-acid-specific binding behavior may also exist (Figure 4B): the polar amino-acids tyrosine, histidine, and asparagine recruit fewer urea molecules, while lysine, glycine, and phenylalanine on average recruit more. However, only one system is available in this study, so that statistics are necessarily poor. In addition, most types of amino acids show less obvious dependence on binding behavior. This may suggest that urea-binding properties are more significantly encoded in the primary sequence context rather than the individual amino-acid type.

RDC-Selected Urea-Binding Model Is Consistent with NMR Chemical Shift Mapping and SAXS Experiments. A series of titration experiments from 0 to 8 M urea under identical buffer conditions and pH were recorded using $^{1}H-^{15}N$ HSQC spectra (Figure 5B). The spectrum of 8 M urea-denatured ubiquitin at pH 2.5 was assigned previously.[55] Most of the resonances at other urea concentrations can be identified from this assignment. Below 3 M urea concentration, the presence of both folded and unfolded peaks in the spectrum indicates slow exchange kinetics for these two states. Peaks from the population of unfolded ubiquitin thus can be traced down to 2 M urea. The differences between the weighted average chemical shifts of ^{1}H and ^{15}N ($\Delta\delta_{av}$) at 2 M urea and at 8 M urea concentration were calculated and plotted along the amino-acid sequence (Figure 5A). Showing an overall similarity to the distribution of urea binding presented above, the average chemical shift differences are again larger in the central part of ubiquitin while the peaks belonging to the C-terminal part shift less. This qualitative similarity indicates that the chemical shift differences arise from the direct binding of urea to the protein backbone, in agreement with the ASTEROIDS-selected ensemble on the basis of RDCs. Peaks belonging to the N-terminal region, on the other hand, bear no similarity to the urea-binding model. This is possibly not surprising because an unfolding event is also known to occur with increasing amounts of urea: as observed from Φ-value analysis, the N-terminal β-hairpin is populated in the transition state ensemble during unfolding,[57] and even at a 8 M urea concentration at pH 2.5, it

dx.doi.org/10.1021/ja2118688 | J. Am. Chem. Soc. 2012, 134, 4429−4436

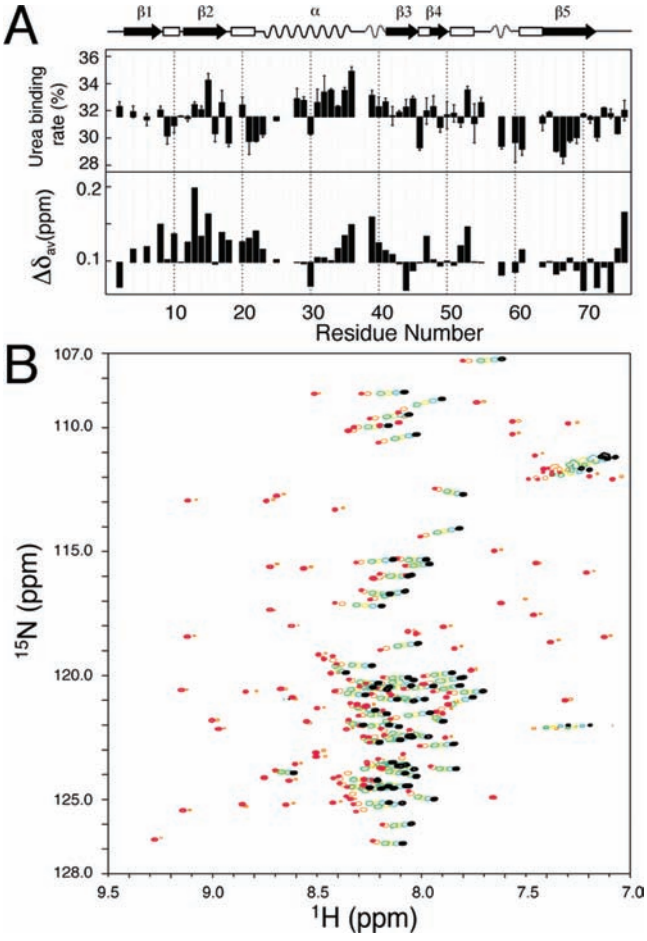


Figure 5. Correlation between urea-binding pattern and chemical shift mapping. (A) Upper panel: the same as Figure 4A but residues not in the bottom panel are removed for ease of comparison. Bottom panel: the average chemical shift difference between 2 and 8 M urea-denatured ubiquitin spectra. The missing residues are proline or overlapped peaks. (B) HSQC spectra for different concentrations of urea: 2 M (red), 3 M (orange), 5 M (green), 6 M (yellow), 7 M (blue), and 8 M (black). The "baseline" represents the mean value and takes into account the general upfield shift of ^{1}H resonances with increasing concentrations of urea.

still retains a certain amount of residual structure, as revealed from different NMR experiments.[33,58] Chemical shift titrations will therefore be expected to reflect these complex unfolding events, so that additional shifts may be observed, in addition to any possible urea binding to the protein backbone.

We also use SAXS data recorded in the presence of 8 M urea to test this model. Predicted SAXS curves for the selected ASTEROIDS ensemble and the pool ensemble were calculated using the program CRYSOL.[59] SAXS curves of ubiquitin urea showed significant interparticle interactions at low scattering angles, especially for low urea concentration (Figure S6 of the Supporting Information).[60] Because of the difficulty in excluding that this phenomenon may impact on the effective radius of gyration derived from the low angle region of the plot, only the parts of the curves lying between the scattering angles of 0.06 and 0.4 Å^{-1} were used for further analysis (Figure 6). This region is nevertheless informative, as it should report on the distribution function of short to medium range distances that will be sensitive to binding of urea to the protein backbone (Figure S7). Over this range the theoretical curve calculated

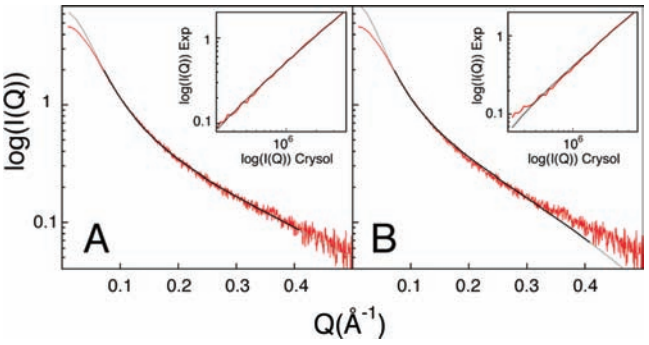


Figure 6. Comparison of experimental and calculated SAXS curves. (A) Curve calculated from the ASTEROIDS selected ensemble (black, the part used in linear fit; gray, the full predicted curve) compared to experimental data (red line). (B) Curve calculated from the pool of randomly sampled urea molecules. The correlations between calculated and experimental curves are shown in inset figures (red lines), and linear regression lines are shown in black. χ^2 values given in the text report on the difference between experimental curves and linear regression lines.

from the ensemble selected against RDCs is found to agree significantly better with the experimental SAXS data compared to the curve calculated from the pool (χ^2 falls from 0.110 to 0.014). The RDCs therefore refine the local conformation and the distribution of urea molecules, and thus, they modify the shape of the theoretical SAXS curve in the 0.06 and 0.4 Å^{-1} range.

DISCUSSION

We have used extensive experimental NMR and SAXS data in combination with statistical coil ensembles to study the conformational behavior of urea-unfolded proteins. In order to develop an atomic resolution understanding of the denatured state, conformational ensembles of full-atom descriptions of unfolded proteins, including side chain conformations derived from rotamer libraries, are combined with random sampling of explicit urea molecules in interaction with the protein. We adopt a model whereby urea molecules directly bind to the protein backbone through H-bonds as proposed from MD simulations and independent sources of experimental data.[7,10,12,14,46,47]

The addition of urea molecules to the statistical coil model modifies the conformational energy landscape of the protein backbone via van der Waals repulsive forces, and this modification leads to significant improvement in the reproduction of extensive experimental RDCs measured from 8 M urea denatured ubiquitin. This improvement allows us to identify the average number of urea molecules that bind each ubiquitin molecule, closely reproducing results from a recent model-free analysis from SAXS/SANS data.[52]

To analyze the distribution of urea binding over the primary sequence, we apply the previously developed genetic algorithm ASTEROIDS,[42,44] to select subensembles in agreement with the experimental RDCs. This analysis demonstrates that urea binds preferentially in the central region of the protein, with binding propensity tailing off toward the termini. This profile shows qualitative similarity to the amplitude of site-specific chemical shift differences observed between low and high concentrations of urea. The validity of the RDC-selected subensemble of urea-bound conformers is further supported by the observation that the reproduction of the predicted SAXS

dx.doi.org/10.1021/ja2118688 | J. Am. Chem. Soc. 2012, 134, 4429–4436

curve is significantly improved in the scattering angle range reporting on medium range distances.

These results are consistent with the results from Pettitt and co-workers, who studied a ten-residue polyalanine peptide in different osmolyte systems using MD simulation, finding that urea molecules make better van der Waals contacts with the protein than water and TMAO.[12] It was also observed from Berne's group that urea unfolds a protein through its stronger dispersion interaction with the protein compared to water molecules.[7] Stumpe and Grubmüller hypothesized from simulations of model compounds that the denaturing mechanism of urea results from its significantly larger volume than that of a water molecule which entropically disrupts the protein structure.[14] The overall picture developed here also compares closely with that resulting from a recent combination of parallel tempering and metadynamics simulation of the effect of urea on the β-hairpin of protein GB1, that found clear evidence that urea preferentially interacts with the protein backbone through H-bonds and that this interaction results in elongation of the peptide chain at higher urea concentrations.[15] The results shown here are in agreement with single molecule FRET studies of proteins showing that end to end distances in denatured proteins increase with the concentration of urea.[61] The expansion of unfolded protein structures was also observed from SAXS[62] and pulse-field gradient NMR[63] with the addition of denaturant. As observed in this study, the increase of binding rate expands the structure, and it is interesting to note that the determined binding rate of around 40% coincides with the approximate volume occupied by urea in an 8 M solution (36.4%). As urea is proposed to weakly bind to the protein,[64] the binding percentage may simply represent the competitive binding of urea to the protein with respect to water. It was also observed in MD simulation that 34−44% of total H-bonds between peptide and solution are contributed by urea molecules and the rest are from water molecules.[12]

To summarize, using a combination of an explicit molecular ensemble based on a statistical coil model, we demonstrate that the direct binding model of urea to the protein backbone is compatible with extensive experimental NMR and small angle scattering data when approximately 40% of backbone peptide groups bind to urea. This level of urea binding closely reproduces numbers derived from a model-free analysis of small angle neutron and X-ray scattering data. We note that the results presented in this study do not rule out the indirect binding model, which would be intrinsically more difficult to model at atomic resolution. The pattern of urea-binding along the amino-acid sequence revealed from this analysis indicates a higher level of binding in the central part of the protein, a trend which resembles independent results derived from chemical shift mapping. These results substantiate the direct-binding model and more generally provide a framework for studying the molecular basis of interactions between proteins and solvent.

MATERIAL AND METHODS

Protein Purification. The purification and preparation of denatured ubiquitin (8 M urea, 10 mM glycine-HCl buffer at pH 2.5) in isotropic solution or in stretched polyacrylamide gel are described elsewhere.[55]

NMR Measurements and Analysis. Experimental RDCs for ubiquitin except $^1D_{C\alpha C\beta}$ were taken from a previous publication.[55] The method for measuring $^1D_{C\alpha C\beta}$ is based on HN(CO)CA-type experiments; the evolution time for Cα was set to be longer than $1/2J_{C\alpha C\beta}$ for sufficient coupling development. Here, a 2048*(^{1}H) × 240*(^{13}C) × 48*(^{15}N) data set (where n* denotes n complex data points) with acquisition times of 85.2, 72.0, and 26.2 ms, respectively, was used.

The titration of urea was started from two samples of 300 μM ubiquitin with 0 or 8 M urea, both at 10 mM glycine-HCl pH 2.5, and a temperature of 25 °C. Specific amounts of solution were mixed to achieve urea concentrations 1 and 7 M, 2 and 6 M, 3 and 5 M at each step. By this procedure, the concentration of ubiquitin and buffer and pH were kept identical along the titration with a minimum amount of protein sample usage. Standard ^{1}H−^{15}N HSQC spectra were recorded on a Bruker DRX 800 MHz spectrometer equipped with a cryoprobe.

Ensemble Modeling. In the original Flexible-MECCANO (FM) algorithm, amino-acid-specific volumes were represented by spheres placed at the Cβ positions (or Cα for Gly), and a steric clash was detected when the spheres overlapped.[36,65] If necessary, all-atom descriptions of side chains were added by other programs, e.g. SCCOMP.[66] In this study, energy-minimized internal coordinates of each amino acid are used as a polypeptide building block (J.-r.H. and M.B., unpublished), so that side chains can be added explicitly immediately after each backbone amino acid has been constructed. The side chain is rotated around all possible χ angles according to the stagger rotamer population derived from the PISCES[67] purged Coil Library[68] in the built backbone ϕ/ψ space to find a steric-clash-free accommodation.

Steric clash checking was applied according to hard-sphere van der Waals radii for every atom between the side chain and the built backbone. The hard-sphere radii for different atom types are based on the values used in Xplor-NIH[69] with the scaling factor (0.76) used conventionally in the final stage of structural refinement. These scaled radii are very similar to a recent MD study from Rose and co-workers on their modified allowed Ramachandran space survey.[70] If the side chain cannot find a nonclash conformation within a certain number of trials (defaulted as 100 times), the algorithm returns to the previous residue and builds another conformation using a different ϕ/ψ combination.

The algorithm can be modified to add side chains after the whole backbone is constructed. In this case, there would be no difference in backbone sampling compared to the spherical side chain model. However, for the purpose of sampling urea molecules, adding side chains while building the protein backbone is the most efficient approach. This all-atom description for the statistical coil model facilitates further analysis of SAXS data, chemical shift prediction, and the urea-binding model.

RDC Prediction. Theoretical RDCs were calculated on the basis of the assumption of steric exclusion[71,72] using an efficient in-house written algorithm.[33] In short, the maximal extension of a molecule for each direction of a unit sphere is calculated. The probability for finding the molecule in a certain orientation is then derived as the volume that can be occupied by the molecule between two infinitely extended, parallel planes relative to the total distance between the planes. The alignment tensor then corresponds to the average over all orientations of second rank spherical harmonics weighted by this probability. The theoretical RDCs are then calculated from the alignment tensor:

$$D_{k,ij}^{\mathrm{calc}} = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S_{k,m}^{*}\frac{Y_{2m}(\Theta_{k,ij},\,\Phi_{k,ij})}{r_{k,ij}^{3}} \qquad (1)$$

where $D_{k,ij}$ represents the RDC between nuclei i and j for ensemble member k with individual alignment tensor $S_{k,m}$ (written in irreducible form[73]), Y_{2m} are spherical harmonics, $r_{k,ij}$, $\Theta_{k,ij}$, and $\Phi_{k,ij}$ are the polar coordinates of the internuclear vector, and γ are the nuclear gyromagnetic ratios. These RDC values for the individual structures were then averaged over all members of the ensemble to obtain the calculated value:

$$D_{ij}^{\mathrm{calc}} = \frac{1}{N}\sum_{k=1}^{N} D_{k,ij}^{\mathrm{calc}} \qquad (2)$$

The size of the ensemble (N) throughout this article is 50,000.

In the section on ASTEROIDS selection, a local alignment window size of 15 amino-acids in length is applied in combination with a baseline correction.[42,44]

χ^2 Analysis. χ^2 analysis is used to indicate the agreement between experimental and theoretical values. It is defined as follows:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{D_i^{\mathrm{obs}} - D_i^{\mathrm{calc}}}{\sigma_i} \right)^2 \quad (3)$$

where σ_i is the experimental error and the summation runs over all observed data n.

Average Chemical Shift Calculation. The weighted average chemical shift is calculated according to the following equation:[74]

$$\Delta\delta_{\mathrm{av}} = \sqrt{\frac{(\Delta\delta_{\mathrm{H}})^2 + \left(\frac{1}{5}\Delta\delta_{\mathrm{N}}\right)^2}{2}} \quad (4)$$

where $\Delta\delta_{\mathrm{H}}$ and $\Delta\delta_{\mathrm{N}}$ are the chemical shift differences between two ^{1}H−^{15}N HSQC spectra for amide proton and nitrogen chemical shifts, respectively.

SAXS Measurement and Analysis. The SAXS measurements were recorded on the ID14-3 BioSAXS beamline at the European Synchrotron Radiation Facility (ESRF Grenoble, France). The sample−detector distance was 2.6 m, and the X-ray wavelength used was 0.931 Å (13.32 keV). 50 μL of each protein solution was loaded into a flow-through quartz capillary cell at 25 °C. The total exposure time was 100 s per sample. The 2D diffraction patterns were normalized to an absolute scale and azimuthally averaged to obtain the intensity profiles $I(Q)$, within BSxCuBE (ESRF beamline data collection software). Solvent contributions (buffer backgrounds collected before and after every protein sample) were averaged and subtracted from the associated protein sample.

The theoretical curves were calculated with CRYSOL[59] with default settings except that the scattering angle was up to 0.6 Å^{-1}, and the order of harmonics and the order of the Fibonacci grid were maximized to 50 and 18, respectively, to optimize the resolution. Experimental data was interpolated with a cubic spline function to facilitate linear regression at the same values of scattering angles of calculated curves.

Calculation of the Radius of Gyration. The individual R_g value for each structure of an ensemble is the root-mean-squared distance toward its center of mass.[75] The averaged R_g of an ensemble is calculated by the root-mean-squared values of all individual R_g values (denoted as $\langle R_g \rangle$ throughout) as the Guinier analysis of SAXS data corresponding to an average over R_g of the individual molecules in the experimental ensemble.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Comparison of using side chain and explicit side chain descriptions. Comparison of the effect of different H-bond geometries between urea and protein backbone. R_g and ϕ/ψ angle distribution for different saturation rates of urea binding models. Distribution of urea molecules in ensemble calculations. Cross-validation of RDC data. SAXS curves of denatured ubiquitin for different urea concentrations. SAXS curves calculated from different urea binding models. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

martin.blackledge@ibs.fr

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to acknowledge the Agence National de Recherche for financial support from TAUSTRUCT - ANR MALZ 2010 (to M.B.), Proteindisorder - ANR JCJC 2010 (M.R.J.), National Science Council of Taiwan NSC98-2917-I-564-155 (to J.-r.H.), and SNF grant 31-132857 to S.G. We would like to acknowledge the ESRF for SAXS beam-time on ID14-3 in the framework of an IBS Block Allocation Group and Dr. Adam Round (local contact).

■ REFERENCES

(1) Dyson, H. J.; Wright, P. E. *Chem. Rev* **2004**, *104*, 3607−3622.
(2) Dobson, C. *Nature* **2003**, *426*, 884−890.
(3) Fink, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35−41.
(4) Gong, H.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3321−3326.
(5) Schellman, J. A. *C. R. Trav. Lab. Carlsberg, Ser. Chim.* **1955**, *29*, 223−229.
(6) Tanford, C. *J. Am. Chem. Soc.* **1964**, *86*, 2050.
(7) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16928−16933.
(8) Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2009**, *113*, 12816−12824.
(9) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142−5147.
(10) Canchi, D. R.; Paschek, D.; García, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338−2344.
(11) Caflisch, A.; Karplus, M. *Structure* **1999**, *7*, 477−488.
(12) Kokubo, H.; Hu, C. Y.; Pettitt, B. M. *J. Am. Chem. Soc.* **2011**, *133*, 1849−1858.
(13) Zangi, R.; Zhou, R.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131*, 1535−1541.
(14) Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126−16131.
(15) Berteotti, A.; Barducci, A.; Parrinello, M. *J. Am. Chem. Soc.* **2011**, *133*, 17200−17206.
(16) Eberini, I.; Emerson, A.; Sensi, C.; Ragona, L.; Ricchiuto, P.; Pedretti, A.; Gianazza, E.; Tramontano, A. *J. Mol. Graph. Model.* **2011**, *30*, 24−30.
(17) Xue, Y.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2011**, *133*, 14614−14628.
(18) Mackerell, A. D. Jr.; Feig, M.; Brooks, C. L. 3rd. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(19) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(20) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36−L38.
(21) Li, D.-W.; Brüschweiler, R. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 6778−6780.
(22) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07−09.
(23) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673−699.
(24) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(25) Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974−980.
(26) Otting, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 259−285.
(27) Halle, B. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* **2004**, *359*, 1207−1223.
(28) Nucci, N. V.; Pometun, M. S.; Wand, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 12326−12329.
(29) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.
(30) Vajpai, N.; Gentner, M.; Huang, J.-R.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 3196−3203.

(31) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C.; Poulsen, F.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291−3299.
(32) Dedmon, M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. *J. Am. Chem. Soc.* **2005**, *127*, 476−477.
(33) Huang, J.-R.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694−705.
(34) Esteban-Martín, S.; Fenwick, R. B.; Salvatella, X. *J. Am. Chem. Soc.* **2010**, *132*, 4626−4632.
(35) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494−506.
(36) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 17002−17007.
(37) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13104.
(38) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M. *Structure* **2009**, *17*, 1169−1185.
(39) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W. H.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055−8061.
(40) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968−16975.
(41) Mukrasch, M. D.; Markwick, P.; Biernat, J.; Bergen, M.; von Bernadó, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235−5243.
(42) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908−17918.
(43) Wells, M.; Tidow, H.; Rutherford, T.; Markwick, P.; Jensen, M.; Mylonas, E.; Svergun, D.; Blackledge, M.; Fersht, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762−5767.
(44) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407−8418.
(45) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270−1272.
(46) Lim, W. K.; Rösgen, J.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2595−2600.
(47) Almarza, J.; Rincon, L.; Bahsas, A.; Brito, F. *Biochemistry* **2009**, *48*, 7608−7613.
(48) Tobi, D.; Elber, R.; Thirumalai, D. *Biopolymers* **2003**, *68*, 359−369.
(49) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346−7353.
(50) Klimov, D. K.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14760−14765.
(51) Chen, X.; Sagle, L. B.; Cremer, P. S. *J. Am. Chem. Soc.* **2007**, *129*, 15104−15105.
(52) Gabel, F.; Jensen, M. R.; Zaccaï, G.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 8769−8771.
(53) Pryor, A.; Sanger, P. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1970**, *A26*, 543.
(54) Lipsitz, R. S.; Sharma, Y.; Brooks, B. R.; Tjandra, N. *J. Am. Chem. Soc.* **2002**, *124*, 10621−10626.
(55) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799−9807.
(56) Bernadó, P.; Blackledge, M. *Biophys. J.* **2009**, *97*, 2839−2845.
(57) Went, H. M.; Jackson, S. E. *Protein Eng. Des. Sel.* **2005**, *18*, 229−237.
(58) Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 754−755.
(59) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.
(60) Shukla, A.; Mylonas, E.; Di Cola, E.; Finet, S.; Timmins, P.; Narayanan, T.; Svergun, D. I. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5075−5080.
(61) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14609−14614.
(62) Latypov, R. F.; Liu, D.; Jacob, J.; Harvey, T. S.; Bondarenko, P. V.; Kleemann, G. R.; Brems, D. N.; Raibekas, A. A. *Biochemistry* **2009**, *48*, 10934−10947.
(63) Sevcsik, E.; Trexler, A. J.; Dunn, J. M.; Rhoades, E. *J. Am. Chem. Soc.* **2011**, *133*, 7152−7158.
(64) Schellman, J. A. *Biopolymers* **1987**, *26*, 549−559.
(65) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(66) Eyal, E.; Najmanovich, R.; McConkey, B. J.; Edelman, M.; Sobolev, V. *J. Comput. Chem.* **2004**, *25*, 712−724.
(67) Wang, G.; Dunbrack, R. L. Jr. *Bioinformatics* **2003**, *19*, 1589−1591.
(68) Fitzkee, N. C.; Fleming, P. J.; Rose, G. D. *Proteins* **2005**, *58*, 852−854.
(69) Schwieters, C.; Kuszewski, J.; Clore, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47−62.
(70) Porter, L. L.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 109−113.
(71) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.
(72) van Lune, F.; Manning, L.; Dijkstra, K.; Berendsen, H.; Scheek, R. *J. Biomol. NMR* **2002**, *23*, 169−179.
(73) Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77−82.
(74) Grzesiek, S.; Stahl, S. J.; Wingfield, P. T.; Bax, A. *Biochemistry* **1996**, *35*, 10256−10261.
(75) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W H Freeman & Co.: San Francisco, CA, 2012.