Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts

ARTICLE *in* BIOCHEMISTRY · MARCH 2012

Impact Factor: 3.02 · DOI: 10.1021/bi3001825 · Source: PubMed

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts

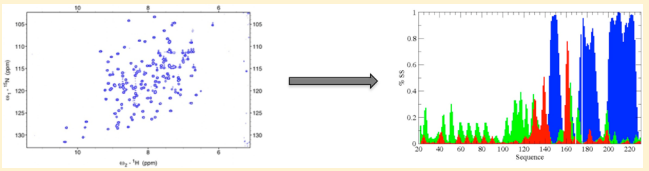
Carlo Camilloni,[†] Alfonso De Simone,[†,§] Wim F. Vranken,[‡,⊥] and Michele Vendruscolo*[,†]

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

[‡]European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, U.K.

Supporting Information

ABSTRACT: One of the major open challenges in structural biology is to achieve effective descriptions of disordered states of proteins. This problem is difficult because these states are conformationally highly heterogeneous and cannot be represented as single structures, and therefore it is necessary to characterize their conformational properties in terms of probability distributions. Here we show that it is possible to obtain highly quantitative information about particularly important types of probability distributions, the populations of secondary structure elements (α-helix, β-strand, random coil, and polyproline II), by using the information provided by backbone chemical shifts. The application of this approach to mammalian prions indicates that for these proteins a key role in molecular recognition is played by disordered regions characterized by highly conserved polyproline II populations. We also determine the secondary structure populations of a range of other disordered proteins that are medically relevant, including p53, α-synuclein, and the Aβ peptide, as well as an oligomeric form of αB-crystallin. Because chemical shifts are the nuclear magnetic resonance parameters that can be measured under the widest variety of conditions, our approach can be used to obtain detailed information about secondary structure populations for a vast range of different protein states.



Significant advances have been made recently in the use of NMR chemical shifts for protein structure determination, as it has been shown that they can be used to define the structures of proteins in solution[1−7] and in the solid state,[3,8] as well as the structures of protein complexes.[9,10] At least for proteins up to ∼130 residues in length, these methods appear to provide structures at a resolution comparable to that of more standard NMR methods.

Additional interest in the use of chemical shifts in structural biology is generated by the fact that even when the information provided by chemical shifts is not sufficient to define the tertiary structures of proteins, these observables can provide very accurate information about their secondary structures. For instance, in the case of native states, several highly accurate methods for mapping backbone chemical shifts to backbone dihedral angles[11,12] and for identifying secondary structure elements[12−17] exist. It has also been long recognized that the analysis of the secondary chemical shifts, which represent the differences between measured and random coil chemical shifts,[13,18−20] can be used to define the presence of α-helices and β-strands in partially structured states. In the chemical shift index (CSI) analysis, values of −1, 0, and 1 are introduced, which correspond to α-helix, random coil, and β-strand, respectively, and are assigned to each residue on the basis of the chemical shift deviations relative to the statistics of random coil values.[13] The quantitative characterization of secondary structure elements in disordered states has been recently further improved through the introduction of the SSP method, which provides secondary structure preferences in terms of a score between 0 and 1 for α-helices and between 0 and −1 for β-sheets.[16]

As essentially all proteins at various points during their lifetimes populate disordered states, which in many cases are of key importance in determining their functional or dysfunctional behavior,[21−23] it would be very desirable to further develop methods capable of exploiting chemical shifts to achieve a detailed characterization of the conformational properties of such states. In this work, we show that it is possible to use backbone chemical shifts to determine the populations of secondary structure elements in cases ranging from fully structured native states to partially folded intermediates and intrinsically disordered states. Many of the examples that we discuss illustrate situations in which the characterization of structural disorder in terms of secondary structure populations is conducted for proteins with important functional roles (e.g., p53 and αB-crystallin) or proteins involved in disease (e.g., Aβ, α-synuclein, and mammalian prions). The possibility of obtaining highly quantitative structural information about such important states of proteins through an analysis of chemical shifts is particularly attractive, as these parameters can often be measured quite readily even in cases when structural information is difficult to obtain otherwise.

Received: February 12, 2012
Published: February 23, 2012

   dx.doi.org/10.1021/bi3001825 | *Biochemistry* 2012, 51, 2224−2231

Once chemical shifts are available, the method that we introduce produces results very readily because it achieves a direct mapping between chemical shifts and secondary structure populations, without requiring an explicit characterization of the structure and dynamics of proteins in terms of three-dimensional conformations or conformational ensembles, as in other methods that have been proposed recently.[24] The quality of the determination of the secondary structure populations is assessed by comparison with other NMR and circular dichroism measurements. A web server (http://www-vendruscolo.ch.cam.ac.uk/d2D/) is available for performing the calculations.

METHODS

Parametrization of the δ2D Method. To parametrize the δ2D method, we used a database of high-resolution native state conformations for which chemical shift measurements are available.[19] In this database, we extracted three libraries of secondary structure elements, which were composed by all α-helical fragments longer than three residues (4480 α-helical fragments for a total of 55470 residues), all β-strand fragments longer than one residue (7038 β-strand fragments for a total of 36871 residues), and all PPII fragments longer than three residues (2419 residues, which were identified following the method of Berisio et al.[25]). In the parametrization procedure, 10% of the structures were extracted for testing and the remaining 90% were used to calculate the average chemical shifts and the correction factors for each residue in the three types of secondary structure elements.

By generalizing the method that we used to define the random coil chemical shifts from the amino acid sequence of a protein,[19] we set up three different predictors for the chemical shifts in α-helical elements

$$\delta^{H}_{iA} = \delta^{0}_{iA} + \alpha_i^{-}\delta^{1}_{iBA} + \alpha_i^{+}\delta^{1}_{iAC} + \alpha_i^{2+}\delta^{2}_{iAxC} + \alpha_i^{3+}\delta^{3}_{iAxxC} + \alpha_i^{4+}\delta^{4}_{iAxxxC}$$

and in β-strand and PPII elements

$$\delta^{E}_{iA} = \delta^{0}_{iA} + \alpha_i^{2-}\delta^{1}_{iBxA} + \alpha_i^{-}\delta^{1}_{iBA} + \alpha_i^{+}\delta^{1}_{iAC} + \alpha_i^{2+}\delta^{2}_{iAxC}$$

where δ^0 represents the average chemical shift value for each residue and the coefficients (α) are a set of parameters optimized to maximize the correlation between the experimental and predicted chemical shift values. The parameters δ^1–δ^4 are the correction factors due to the nearest neighbors along the polypeptide chain; the use of more parameters would lead to overfitting, at least in the case of the database used here. In the case of α-helical fragments, we used a different set of neighbor residues with respect to β-strands and PPII. The tables with the average values, the correction tables, and the parameters are available for download from the δ2D website (http://www-vendruscolo.ch.cam.ac.uk/d2D/).

We then assessed the performance of the individual predictors for α-helices, β-strands, and PPII secondary structures on the fragments taken from the proteins removed for testing. From the test, we measured the average error in the prediction of each chemical shift for each conformation. These errors are then used to set the width of Lorentzian functions, as specified below. We found that these functions provided slightly better performance than Gaussian functions (see Figures S1–S3 of the Supporting Information). The errors in the predictions could arise from the fact that not all the amino acids in the proteins we used in the database populate exactly an ideal single secondary structure, and that the values of the chemical shifts of neighboring atoms are weakly correlated (Figure S4 of the Supporting Information). Furthermore, errors can also arise from the different sensibility of each atom to the different secondary conformations. By taking account of these effects, it should be possible in the future to improve the quality of the calculation of secondary structure populations from chemical shifts.

By combining the specific predictors of chemical shifts for the different types of secondary structure elements and by comparing the result to the measured chemical shifts, we extrapolate the relative populations of the different secondary structures. We thus write the secondary structure population for each residue as

$$P_{\{H,E,C,P\}}(\delta^{C\alpha}, \delta^{C\beta}, \delta^{C'}, \delta^{H\alpha}, \delta^{HN}, \delta^{N}) = \left[\prod_{i=(C\alpha,C\beta,C',H\alpha,H,N)} \mathbf{a}^{i}_{\{H,E,C,P\}} G (\delta^{i} - \delta^{i}_{\{H,E,C,P\}}) \right] \Big/ \left[\sum_{j=\{H,E,C,P\}} \prod_{i=(C\alpha,C\beta,C',H\alpha,H,N)} \mathbf{a}^{i}_{\{H,E,C,P\}} G (\delta^{i} - \delta^{i}_{j}) \right]$$

where G is a normalized Lorentzian distribution centered on the chemical shift value provided by the predictors, the standard deviation is given by the error in the prediction of that chemical shift in that secondary structure calculated from the 10% test database, and $\mathbf{a}$ is a 6×4 matrix of free parameters optimized through a Monte Carlo procedure on the whole database (apart for the 10% previously removed) to maximize the agreement between the assigned secondary structures (i.e., the most populated secondary structures) and the experimental ones.

Accuracy and Precision of the δ2D Method. The accuracy in the predictions obtained by the δ2D method for native proteins can be estimated from Figure 1, which indicates an error of ∼10% in assigning the most populated secondary structures if one takes into account an error of ∼5% by DSSP and STRIDE. The corresponding accuracy for unstructured proteins can be estimated from the comparison with CD measurements shown in Figure 6, which suggests an error of ∼4%. To estimate the precision of the δ2D method, we repeated twice the Monte Carlo optimization procedure to obtain the $\mathbf{a}$ matrix. We found an average variation in the secondary structure population per residue of ∼8%. Importantly, if the δ2D method is used to estimate the changes in secondary structure populations for a given protein under different conditions, the error is reduced to ∼2%, a result that provides confidence in the reliability of the analysis shown in Figure 4 for the Aβ peptide.

dx.doi.org/10.1021/bi3001825 | Biochemistry 2012, 51, 2224–2231

Random Coil Optimization. We further optimized the CamCoil method[19] by iteratively generating the random coil reference database by only keeping those residues that populate a coil structure for more than 96%. In this way, a new set of reference random coil values was obtained (see Table S1 of the Supporting Information).

RESULTS AND DISCUSSION

The δ2D Method. In this work, we describe the δ2D method, which translates a set of chemical shifts into probabilities of occupation of secondary structure elements. The approach that we describe is based on a generalization of a method that we used to define the random coil chemical shifts from the amino acid sequence of a protein.[19] To develop the δ2D method, the first step is to set up three further sequence-based predictors, which provide the ideal chemical shifts for α-helical, β-sheet, and polyproline II (PPII) secondary structures. The distinction between β-sheet and PPII secondary structures is relevant because, although they occupy nearby regions in the Ramachandran map, they are characterized by very different patterns of hydrogen bonds. The prediction of other types of secondary structure types will also become possible using the approach presented here (see Methods) as the number of chemical shifts deposited in publicly available databases will increase. The values provided by the four predictors currently implemented in the δ2D method (α-helical, β-sheet, PPII, and random coil secondary structures) represent the chemical shifts that would be measured for a given amino acid sequence in a state with a 100% population of a given secondary structure element. In the second step, by combining these predictions and by comparing the results with the experimental chemical shifts, we extract the most probable populations of secondary structure elements (see Methods).

Native States. Before presenting the results of the application of the δ2D method to non-native states, we report a test of its performance in the case of native states, as it is well-known that for these states the information provided by chemical shifts can be used to identify the positions of secondary structure elements with high accuracy.[13−17] Several methods for performing this task have been compared recently.[12] Although there are several factors that influence the results of this type of comparison, such as the specific set of proteins used in the test and the random coil reference values, by testing the δ2D method by adopting the same procedure used in ref 12, we found that it performs at least as well as other existing methods (Figure 1).

Denatured States. In addition to the prediction of native secondary structure elements, we further tested the δ2D method on the prediction of random coil chemical shifts. To conduct a consistency test, we used a set of seven denatured proteins that were recently used to develop the CamCoil method,[19] finding that the δ2D method has an A_3 index of accuracy of 100% in six cases and 98.5% in the seventh one; the A_3 index indicates the percentage of predictions of the three-state secondary structure class (α-helix, β-strand, and coil) that are identical to the class in the reference structure.[12] We also considered a second set of nine unfolded proteins recently used to derive the ncIDP predictor of random coil chemical shifts.[20] To develop the ncIDP predictor, only the regions of these proteins that were considered to be fully random coil were used.[20] In this context, our results illustrate how our approach is capable of identifying regions of significant secondary structure populations in otherwise highly unstructured proteins (Figure S5 of the Supporting Information).

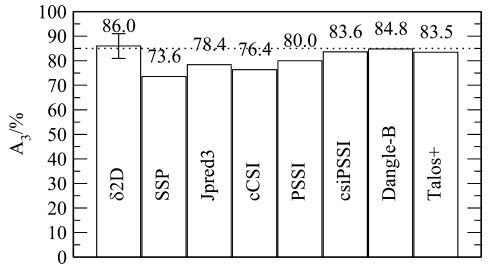


Figure 1. Comparison of the accuracy of different methods currently available for the prediction of secondary structure elements in native states: δ2D (this work), SSP,[16] JPRED3,[17] cCSI,[13] PSSI,[14] PsiCSI,[15] Dangle-B,[12] and Talos+.[11] The error bar for the δ2D method represents the standard deviation in the predictions for the proteins in the database used in a recent study.[12] The A_3 index represents the percentage of predictions of the three-state secondary structure class (α-helix, β-strand, and coil) that are identical to the class in the reference structure.[12]

After these initial tests, we applied the δ2D method to study the progressive loss of secondary structure upon urea and pH unfolding of ACBP, a protein for which a highly accurate series of chemical shift measurements is available.[26] We found that an increase in urea concentration and a decrease in pH disrupt progressively the secondary structure organization of this protein, but still at 5 M urea and pH 2.3, there are residual portions of the four α-helices that are still populated with a probability of ≤10% (Figure S6 of the Supporting Information).

We conducted a further validation on a "reference ensemble" test.[27,28] In this test, a set of reference chemical shifts are calculated from an ensemble of conformations and then used in turn to calculate using the δ2D method the secondary structure populations in the same ensemble. Such secondary structure populations are then compared with the corresponding reference populations, which are those directly defined for the structures making up the reference ensemble. We conducted this test by considering as a reference ensemble a recently determined ensemble of conformations representing the unfolded state of drk SH3.[29] Two sets of reference chemical shifts were calculated for this ensemble using Sparta+[30] and ShiftX2,[31] and the reference secondary structure populations were calculated using Stride[32] for α-helical, β-sheet, and coil secondary structure elements, and the Stapley and Creamer convention for PPII secondary structure elements.[25,33] The results that we obtained demonstrate that the δ2D method is capable of recovering the secondary structure populations within an error that is comparable to the differences between Sparta+ and ShiftX2 in the calculations of chemical shifts (Figure S7 of the Supporting Information). We note that secondary structures can also be identified using a Ramachandran map convention.[29] In this case, secondary structures are defined by the fraction of residues in the α-helical or β-sheet regions of Ramachandran space,[29] while in Stride,[32] an energy function is used that takes account dihedral angle, Lennard-Jones, and hydrogen bond terms.[32] Additional differences concern how coil and PPII secondary structures are treated. In the Ramachandran map convention, PPII regions are part of the β-sheet section of Ramachandran space, while in the Stapley and Creamer convention, they are considered as separate, as they do not form hydrogen bonds; coils are defined as regions

dx.doi.org/10.1021/bi3001825 | Biochemistry 2012, 51, 2224−2231

that sample both α-helical and β-sheet sections of Ramachandran space, while Stride defines them as regions not in one of the other secondary structure types.

Intermediate States. As an example of the application of the δ2D method to intermediate states, we compared the secondary structure populations in the native and intermediate states of Im7, a four-helix bundle protein.[34] The chemical shifts of the L53A/I54A double mutant, which completely populates the intermediate state under native conditions,[34] were used to perform the calculations relative to the intermediate state. The results that we found indicate that α-helix III is rather unstable in the native state, with α-helical turns formed with about a 60% probability; this region is nearly completely disrupted in the intermediate state (Figure 2). The overall predicted α-helical populations of 51% and 44%, for the native and intermediate states, respectively, are in agreement with the populations derived from CD spectra (50% and 45%, respectively).[34] These results are also in overall agreement with those recently determined from hydrogen exchange measurements for the intermediate state of wild-type Im7[34] (cf. Figure 2b in ref 34) and indicate that the δ2D method provides a quantitative assessment of the degree of formation of secondary structure elements even in the absence of information about their tertiary organization.

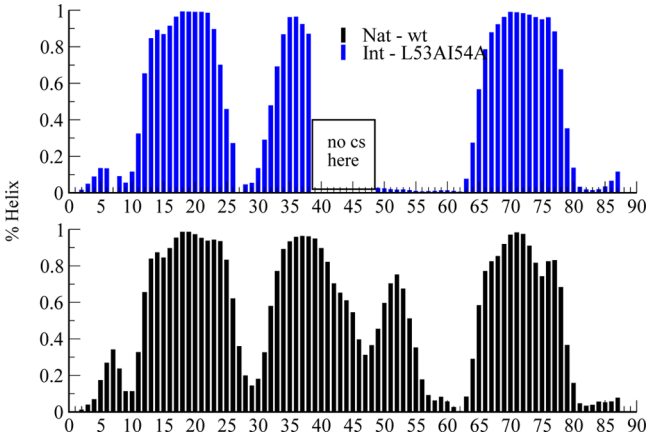


Figure 2. Comparison of the secondary structure populations in the native (black, bmr7316) and intermediate (blue, bmr7317) states of Im7,[34] obtained with the δ2D method. The degrees of formation of α-helices I and IV and the N-terminal region of α-helix II are essentially the same in the native and intermediate states. By contrast, α-helix III is marginally stable in the native state and nearly completely disrupted in the intermediate state.

Intrinsically Disordered States. The δ2D method is particularly useful for the characterization of the conformational properties of intrinsically disordered proteins. We present an analysis of the results of the application of this approach to a wide range of proteins of this type. We first discuss three cases, N_{TAIL}, ctFCP1, and p53, for which an accurate independent characterization of the secondary structure populations has been conducted, so that they serve as validation tests for the δ2D method. We then present results for the Aβ peptide, α-synuclein, and the human prion protein, which are important because of their medical relevance.

N_{TAIL}. We applied the δ2D method to identify the presence secondary structure elements in N_{TAIL}, the C-terminal intrinsically disordered domain of the Sendai virus nucleoprotein.[35] The molecular recognition element of N_{TAIL} has been shown, using detailed analysis of multiple residual dipolar couplings, to contain a conformationally fluctuating α-helical element, whose core is formed by residues 479−484 and populated with a probability of ∼75−82%.[35] By using the δ2D method, we find this region to have a similar α-helical content (85%) (Figure 3a). These results indicate that the analysis of chemical shifts and of residual dipolar couplings leads to quantitatively similar results in the case of this protein. These results are also consistent with those of a recent study[24] in which the α-helical region was identified by an analysis of the chemical shifts that involved the generation of a pool of three-dimensional conformations of the protein and their subsequent selection through a score taking account of the agreement between experimental chemical shifts and those back-calculated from the structures in the pool. In that study, however, the population of the α-helical element was not estimated directly from the chemical shifts.

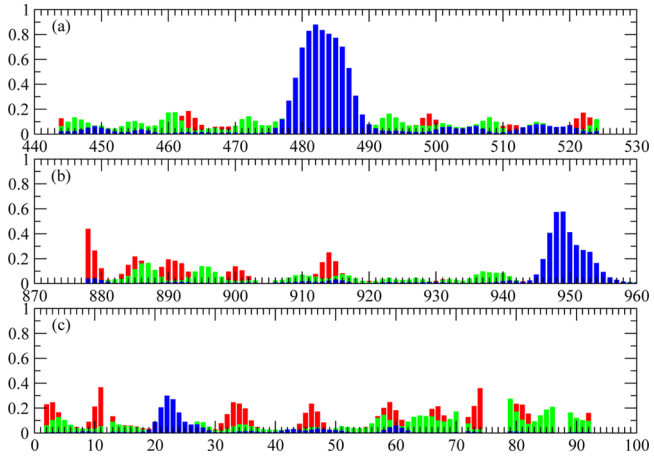


Figure 3. Validation of the calculations of secondary structure populations using the δ2D method (secondary structure populations are colored red for β-sheets, blue for α-helices, and green for PPIIs): (a) N_{TAIL}, the C-terminal domain of the Sendai virus nucleoprotein (bmr15123),[24] (b) ctFCP1, a construct corresponding to residues 879−961 of the TFIIF-associating component of the RNA polymerase II C-terminal domain phosphatase (bmr16296),[36] and (c) the N-terminal domain (residues 1−92) of p53 (bmr17760).[38,39]

ctFCP1. This is a construct corresponding to residues 879−961 of the TFIIF-associating component of the RNA polymerase II C-terminal domain phosphatase (bmr16296).[36] In a recent study (Figure 3b), the formation of a transient α-helical region was characterized as corresponding to approximately residues 945−955. The calculations that we conducted using the δ2D method are fully consistent with this result and quantify the α-helical content as ∼35% within this region.

N-Terminal Domain of p53. The tumor suppressor p53 protein plays key roles in a variety of cellular pathways, including those involved in the maintenance of the integrity of the human genome, in the control of apoptosis, in cell-cycle arrest, and in DNA repair.[37] p53, which functions by forming a homotetramer, is a 393-residue protein formed by seven domains. The highly sophisticated regulation of the activity of p53 takes advantage of the organization of the protein into two folded domains (tetramerization and DNA-binding core domains), which are linked together and flanked by five natively unfolded domains, including the N-terminal region (residues 1−92), which includes the transcription−activation

dx.doi.org/10.1021/bi3001825 | *Biochemistry* 2012, 51, 2224−2231

domain (TAD, residues 1−42), which activates transcription factors, the activation domain 2 (AD2, residues 43−63), which is important for apoptotic activity, and the proline-rich domain (residues 64−92), which is important for the apoptotic activity of p53. Here we have applied the δ2D method to analyze the conformational properties of the N-terminal region, for which a set of backbone chemical shifts is available.[38,39] Our results suggests that there is an α-helix loop in the region of residues 21−25, which are within the nuclear export signal region (Figure 3c), and two long PPII regions that correspond to residues 62−86, which is within the proline-rich region (residues 61−93). We also found that there are no signs of extended β-sheet structures. These results are consistent with recent experimental studies,[38,39] as in both cases the extent of α-helix formation is estimated to be ∼10−20% in this region.

Aβ. According to the amyloid hypothesis, a key initiating pathogenic event in Alzheimer's disease is the assembly of neurotoxic aggregates formed by the Aβ peptide.[40] As a consequence, the development of therapeutic interventions for this medical condition may be facilitated by the characterization of the conformational properties of the monomeric soluble form of this peptide. To take a step in this direction, we applied the δ2D method to the chemical shifts measured for two major forms of the peptide.[41] By applying the δ2D method to the chemical shifts measured for the Aβ(1−40) and Aβ(1−42) variants (Figure 4), we found that the probability of formation

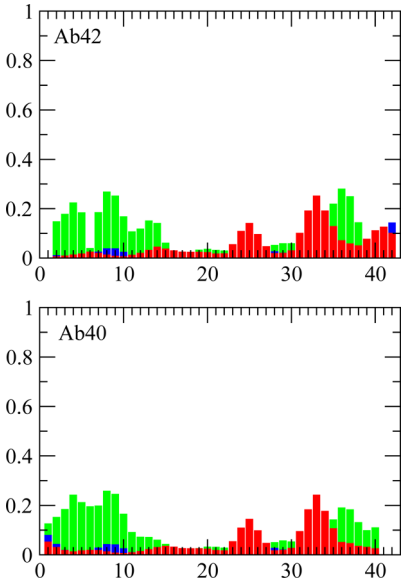


Figure 4. Comparison of the secondary structure populations for the two main variants of the Aβ peptide obtained by the δ2D method. Chemical shifts have been measured[41] for the Aβ(1−40) (bmr17796) and Aβ(1−42) (bmr17794) peptides. Secondary structure populations are colored red for β-sheets, blue for α-helices, and green for PPIIs. The random coil population is not shown explicitly but can be inferred from the condition that the sum of the four types of secondary structure populations should be equal to one.

of β-sheet structure was slightly higher for Aβ(1−42) than for Aβ(1−40), which is in agreement with the experimental observation that the former peptide aggregates more readily.[41]

α-Synuclein. This 140-residue natively unfolded protein self-assembles into intracellular inclusions, known as Lewy bodies, in dopaminergic neurons of patients suffering from Parkinson's disease.[42] To understand the different aggregation behaviors of its variants, β-synuclein and γ-synuclein,[43] we applied the δ2D method to available chemical shift data[16,44,45] to investigate whether there are differences in the secondary structure populations that these three proteins exhibit in their soluble monomeric states. Our results indicate that α-synuclein has a slightly higher β-sheet population than γ-synuclein, which in turn has a much higher β-sheet population than β-synuclein (Figure S8a−c of the Supporting Information). These results are consistent with the known aggregation behavior of these three proteins,[43] and with the observation that α-synuclein and γ-synuclein have toxic in vivo effects, while β-synuclein tends instead to have a protective effect.[46] In addition, we considered two membrane-bound forms of α-synuclein,[45,47] which identify the known regions of high α-helical content (Figure S8d,e of the Supporting Information).

Human Prion Protein (PrP). PrP, which is a 253-residue protein characterized by the presence of a disordered N-terminal domain and an ordered C-terminal domain, is involved in sporadic, inherited, or infectious forms of Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Sheinker syndrome, and fatal familial insomnia.[48] The key event in the pathogenesis of these human diseases is the conversion of the normal α-helix-rich and protease-sensitive cellular isoform of the protein (PrP^C) into a β-sheet-rich aggregated form (PrP^{Sc}). Although a number of high-resolution X-ray and NMR structures of the ordered region of PrP are available,[49] much less is known about the conformational propensities of the N-terminal region. To improve our understanding of the structural basis of the conversion between the PrP^C and PrP^{Sc} forms, we applied the δ2D method to characterize the secondary structure populations in PrP^C. In the ordered C-terminal domain of PrP^C, the δ2D method identifies the α-helical regions (residues 144−153, 172−189, and 200−223) as nearly 100% populated, and the β-strand region (residues 161−164) as being nearly 80% populated (red histogram, Figure 5a). By contrast, the β-strand region (residues 128−131) at the interface between the disordered N-terminal domain and the ordered C-terminal domain is characterized as having a low population, suggesting that this region exhibits significant conformational fluctuations (red histogram, Figure 5a). These findings are consistent with the observation that the hydrogen−deuterion exchange rates are much higher in the first β-strand region (residues 128−131) than in the second (residues 161−164),[49] and with the suggestion that the region of residues 120−130, which is predicted to have a very high intrinsic aggregation propensity,[50] plays an important role in the initiation of the aggregation process of PrP by exploring non-native conformations particularly prone to forming aberrant intermolecular interactions.[50]

The δ2D method also suggests that several regions (around residues 20−30, 40−110, and 130−140) populate a PPII structure with nearly 40% probability (green histogram, Figure 5a). It is remarkable that the PPII populations in the region of residues 90−200 are highly conserved among all known different mammalian prion species (Figure 5b), thus suggesting that a tight control of the probability of formation of PPII structure has a functional role, most likely for molecular recognition. These regions include the PHGGGWGQ tandem repeats (residues 51−91), which have also been reported to form PPII elements by Raman spectroscopy,[51] and the region (residues 100−110) before the interface between the structured and unstructured domains, which has been identified as SH3 binding.[52] We also compared the populations of secondary

dx.doi.org/10.1021/bi3001825 | *Biochemistry* 2012, 51, 2224−2231

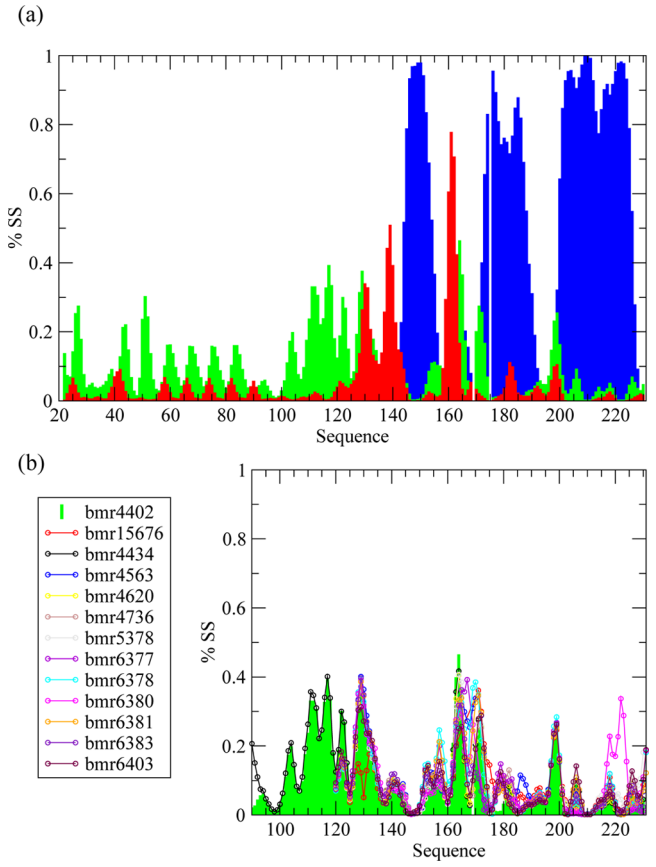


Figure 5. Secondary structure populations of the human prion protein. (a) In the ordered C-terminal region, the δ2D method identifies the regions that form ordered α-helical (blue) and β-sheet (red) structures,[49] while in the disordered N-terminal region, a prevalence of PPII transient secondary structure elements is detected. (b) Our analysis also revealed that regions of high PPII secondary structure populations are highly conserved among all known mammalian prion proteins, suggesting a crucial role of these regions in molecular recognition events.

structure elements in the denatured forms of human and mouse PrP in the reduced and oxidized states (Figure S9 of the Supporting Information).[53]

Analysis of the Disordered Proteins in the BMRB. In addition to the cases discussed so far, to perform a systematic analysis of the performance of the δ2D method, we also conducted calculations of the secondary structure populations for all the denatured or intrinsically disordered proteins for which chemical shifts that were available in the BMRB at the time of writing (Table S1 and Figure S10 of the Supporting Information). This analysis is consistent with the suggestion that PPII is a fairly common motif in unfolded proteins,[54] as well as with the results about PPII secondary structure populations for the individual proteins discussed above.

Oligomeric States. To illustrate the use of the δ2D method in cases in which proteins form conformationally heterogeneous assemblies, we analyzed the oligomeric state of αB-crystallin, which is a small heat shock protein.[55] Our results (Figure S11 of the Supporting Information) indicate the presence of β-strand regions whose locations are in close agreement with those in a structure recently determined by solid state NMR methods.[56] The β-strand regions that we identified have populations ranging from ~70 to ~90%, consistent with the polydisperse nature of αB-crystallin oligomers, which is a prerequisite for their function as molecular chaperones.[56]

Validation by Circular Dichroism. The results that we have presented can be compared with those obtained from the measurement of independent observables reporting on secondary structure populations. In this context, circular dichroism (CD) is particularly suitable because it can provide detailed information about these populations.[57] Our results indicate that the two techniques provide similar results (Figure 6).

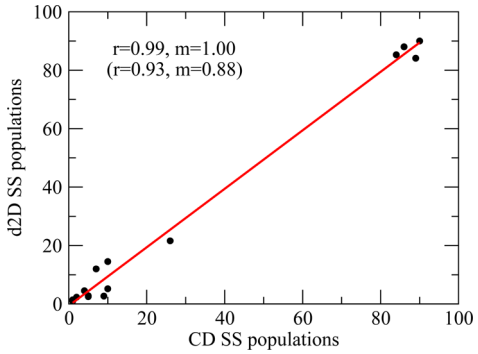


Figure 6. Comparison of the secondary structure populations as derived by the δ2D method from chemical shifts and from circular dichroism (CD) measurements. The numbers in parentheses indicate the correlation obtained by considering only the cases in which secondary structure elements have a population of <30%. The proteins used are α-synuclein, Aβ(1−40), DARRP-32, colicin, merozoite, IA3, p27, and p53-NT, for which CD and NMR data are available from the literature and were collected under approximately similar conditions.

CONCLUSIONS

The realization that many native proteins are intrinsically disordered or contain large unfolded regions has prompted the development of experimental and computational methods capable of characterizing the conformational properties of unstructured states. Further interest in such states stems from the fact that they play crucial roles in the protein folding, misfolding, and aggregation processes. The use of NMR spectroscopy in combination with computational methods has the potential to provide accurate information about the structure and dynamics of proteins even in states that are highly disordered. In this context, we have presented an approach that brings together NMR and computational methods to provide information about the populations of secondary structure elements in disordered states of proteins.

We have shown that it is possible to determine secondary structure populations in structured or unstructured states of proteins if chemical shift measurements of backbone atoms are available. Since it is not yet clear whether it will become possible in the future to use chemical shifts in the absence of any other experimental information to determine ensembles of tertiary structures representing highly dynamical states of proteins, the results that we have presented indicate that chemical shifts can already provide valuable information at the secondary structure level.

dx.doi.org/10.1021/bi3001825 | *Biochemistry* 2012, 51, 2224−2231

■ ASSOCIATED CONTENT

Supporting Information

Further analyses of unstructured proteins with the δ2D method (Table S1 and Figures S5−S11) and additional information about the validation of the δ2D method (Table S2 and Figures S1−S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +44 1223 763873. E-mail: mv245@cam.ac.uk.

Present Addresses

[§]Division of Molecular Biosciences, Imperial College, South Kensington Campus, London SW7 2AZ, U.K.

[⊥]Department of Structural Biology, Vrije Universiteit Brussels, Pleinlaan 2, 1050 Brussels, Belgium.

Funding

This work was supported by FEBS (C.C.), EPSRC (A.D.S.), Innoviris (W.V.), and BBSRC (M.V.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Joseph March and Julie Forman-Kay for having sent us the structures of the unfolded state of drk SH3 and for insightful discussions.

■ ABBREVIATIONS

NMR, nuclear magnetic resonance; PPII, polyproline II; BMRB, Biological Magnetic Resonance Bank.

■ REFERENCES

(1) Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007) Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9615−9620.

(2) Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G. H., Eletsky, A., Wu, Y. B., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D., and Bax, A. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A. 105*, 4685−4690.

(3) Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR 43*, 63−78.

(4) Berjanskii, M., Tang, P., Liang, J., Cruz, J. A., Zhou, J. J., Zhou, Y., Bassett, E., MacDonell, C., Lu, P., Lin, G. H., and Wishart, D. S. (2009) GENMR: A web server for rapid NMR-based protein structure determination. *Nucleic Acids Res. 37*, W670−W677.

(5) Shen, Y., Bryan, P. N., He, Y. N., Orban, J., Baker, D., and Bax, A. (2010) De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci. 19*, 349−356.

(6) Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J., Liu, G. H., Ramelot, T. A., Eletsky, A., Szyperski, T., Kennedy, M. A., Prestegard, J., Montelione, G. T., and Baker, D. (2010) NMR structure determination for larger proteins using backbone-only data. *Science 327*, 1014−1018.

(7) Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R., and Kay, L. E. (2010) A transient and low-populated protein-folding intermediate at atomic resolution. *Science 329*, 1312−1316.

(8) Robustelli, P., Cavalli, A., and Vendruscolo, M. (2008) Determination of protein structures in the solid state from NMR chemical shifts. *Structure 16*, 1764−1769.

(9) Montalvao, R. W., Cavalli, A., Salvatella, X., Blundell, T. L., and Vendruscolo, M. (2008) Structure determination of protein-protein complexes using NMR chemical shifts: Case of an endonuclease colicin-immunity protein complex. *J. Am. Chem. Soc. 130*, 15990−15996.

(10) Das, R., Andre, I., Shen, Y., Wu, Y. B., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperski, T., and Baker, D. (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl. Acad. Sci. U.S.A. 106*, 18978−18983.

(11) Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR 44*, 213−223.

(12) Cheung, M. S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010) Dangle: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson. 202*, 223−233.

(13) Wishart, D. S., and Sykes, B. D. (1994) Chemical-shifts as a tool for structure determination. *Methods Enzymol. 239*, 363−392.

(14) Wang, Y. J., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci. 11*, 852−861.

(15) Hung, L. H., and Samudrala, R. (2003) Accurate and automated classification of protein secondary structure with psicsi. *Protein Sci. 12*, 288−295.

(16) Marsh, J. A., Singh, V. K., Jia, Z. C., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795−2804.

(17) Cole, C., Barber, J. D., and Barton, G. J. (2008) The JPRED3 secondary structure prediction server. *Nucleic Acids Res. 36*, W197−W201.

(18) Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970−2978.

(19) De Simone, A., Cavalli, A., Hsu, S. T. D., Vranken, W., and Vendruscolo, M. (2009) Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J. Am. Chem. Soc. 131*, 16332−16333.

(20) Tamiola, K., Acar, B., and Mulder, F. A. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc. 132*, 18000−18003.

(21) Dobson, C. M. (2003) Protein folding and misfolding. *Nature 426*, 884−890.

(22) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol. 6*, 197−208.

(23) Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett. 579*, 3346−3354.

(24) Jensen, M. R., Salmon, L., Nodet, G., and Blackledge, M. (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc. 132*, 1270−1271.

(25) Berisio, R., Loguercio, S., De Simone, A., Zagari, A., and Vitagliano, L. (2006) Polyproline helices in protein structures: A statistical survey. *Protein Pept. Lett. 13*, 847−854.

(26) Modig, K., Jurgensen, V. W., Lindorff-Larsen, K., Fieber, W., Bohr, H. G., and Poulsen, F. M. (2007) Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett. 581*, 4965−4971.

(27) Kuriyan, J., Petsko, G. A., Levy, R. M., and Karplus, M. (1986) Effect of anisotropy and anharmonicity on protein crystallographic refinement: An evaluation by molecular-dynamics. *J. Mol. Biol. 190*, 227−254.

(28) Richter, B., Gsponer, J., Varnai, P., Salvatella, X., and Vendruscolo, M. (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR 37*, 117−135.

(29) Marsh, J. A., and Forman-Kay, J. D. (2012) Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins 80*, 556−572.

(30) Shen, Y., and Bax, A. (2010) Sparta plus: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR 48*, 13−22.
(31) Han, B., Liu, Y. F., Ginzinger, S. W., and Wishart, D. S. (2011) SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR 50*, 43−57.
(32) Heinig, M., and Frishman, D. (2004) Stride: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res. 32*, W500−W502.
(33) Stapley, B. J., and Creamer, T. P. (1999) A survey of left-handed polyproline II helices. *Protein Sci. 8*, 587−595.
(34) Gsponer, J., Hopearuoho, H., Whittaker, S. B. M., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E., and Vendruscolo, M. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein im7. *Proc. Natl. Acad. Sci. U.S.A. 103*, 99−104.
(35) Jensen, M. R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R. W. H., and Blackledge, M. (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Application to the molecular recognition element of sendai virus nucleoprotein. *J. Am. Chem. Soc. 130*, 8055−8061.
(36) Lawrence, C. W., Bonny, A., and Showalter, S. A. (2011) The disordered C-terminus of the RNA polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem. Biophys. Res. Commun. 410*, 461−465.
(37) Vousden, K. H., and Lane, D. P. (2007) P53 in health and disease. *Nat. Rev. Mol. Cell Biol. 8*, 275−283.
(38) Teufel, D. P., Freund, S. M., Bycroft, M., and Fersht, A. R. (2007) Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. *Proc. Natl. Acad. Sci. U.S.A. 104*, 7009−7014.
(39) Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M., and Fersht, A. R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5762−5767.
(40) Hardy, J., and Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science 297*, 353−356.
(41) Hou, L. M., Shao, H. Y., Zhang, Y. B., Li, H., Menon, N. K., Neuhaus, E. B., Brewer, J. M., Byeon, I. J. L., Ray, D. G., Vitek, M. P., Iwashita, T., Makula, R. A., Przybyla, A. B., and Zagorski, M. G. (2004) Solution NMR studies of the Aβ(1−40) and Aβ(1−42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc. 126*, 1992−2005.
(42) Spillantini, M. G., Schmidt, M. L., Lee, V. M. Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997) α-Synuclein in Lewy bodies. *Nature 388*, 839−840.
(43) Uversky, V. N., Li, J., Souillac, P., Millett, I. S., Doniach, S., Jakes, R., Goedert, M., and Fink, A. L. (2002) Biophysical properties of the synucleins and their propensities to fibrillate: Inhibition of α-synuclein assembly by β- and γ-synucleins. *J. Biol. Chem. 277*, 11970−11978.
(44) Bertoncini, C. W., Rasia, R. M., Lamberto, G. R., Binolfi, A., Zweckstetter, M., Griesinger, C., and Fernandez, C. O. (2007) Structural characterization of the intrinsically unfolded protein β-synuclein, a natural negative regulator of α-synuclein aggregation. *J. Mol. Biol. 372*, 708−722.
(45) Rao, J. N., Kim, Y. E., Park, L. S., and Ulmer, T. S. (2009) Effect of pseudorepeat rearrangement on α-synuclein misfolding, vesicle binding, and micelle binding. *J. Mol. Biol. 390*, 516−529.
(46) Ninkina, N., Peters, O., Millership, S., Salem, H., van der Putten, H., and Buchman, V. L. (2009) γ-Synucleinopathy: Neurodegeneration associated with overexpression of the mouse protein. *Hum. Mol. Genet. 18*, 1779−1794.
(47) Chandra, S., Chen, X. C., Rizo, J., Jahn, R., and Sudhof, T. C. (2003) A broken α-helix in folded α-synuclein. *J. Biol. Chem. 278*, 15313−15318.
(48) Prusiner, S. B. (1998) Prions. *Proc. Natl. Acad. Sci. U.S.A. 95*, 13363−13383.
(49) Zahn, R., Liu, A. Z., Luhrs, T., Riek, R., von Schroetter, C., Garcia, F. L., Billeter, M., Calzolai, L., Wider, G., and Wuthrich, K. (2000) NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U.S.A. 97*, 145−150.
(50) Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F., and Vendruscolo, M. (2008) Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol. 380*, 425−436.
(51) Blanch, E. W., Gill, A. C., Rhie, A. G. O., Hope, J., Hecht, L., Nielsen, K., and Barron, L. D. (2004) Raman optical activity demonstrates poly(L-proline) II helix in the N-terminal region of the ovine prion protein: Implications for function and misfunction. *J. Mol. Biol. 343*, 467−476.
(52) Lysek, D. A., and Wuthrich, K. (2004) Prion protein interaction with the C-terminal SH3 domain of GRB2 studied using NMR and optical spectroscopy. *Biochemistry 43*, 10393−10399.
(53) Gerum, C., Schlepckow, K., and Schwalbe, H. (2010) The unfolded state of the murine prion protein and properties of single-point mutants related to human prion diseases. *J. Mol. Biol. 401*, 7−12.
(54) Shi, Z. S., Chen, K., Liu, Z. G., and Kallenbach, N. R. (2006) Conformation of the backbone in unfolded proteins. *Chem. Rev. 106*, 1877−1897.
(55) Horwitz, J. (1992) α-Crystallin can function as a molecular chaperone. *Proc. Natl. Acad. Sci. U.S.A. 89*, 10449−10453.
(56) Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kuhne, R., Stout, J. R., Higman, V. A., Klevit, R. E., van Rossum, B. J., and Oschkinat, H. (2010) Solid-state NMR and saxs studies provide a structural basis for the activation of $\alpha\beta$-crystallin oligomers. *Nat. Struct. Mol. Biol. 17*, 1037−1043.
(57) Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers 89*, 392−400.