Conformation of the Backbone in Unfolded Proteins

Zhengshuang Shi,[†,‡] Kang Chen,[†] Zhigang Liu,[†] and Neville R. Kallenbach*[,†]

Department of Chemistry, New York University, 100 Washington Place, New York, New York 10003-5180, and Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

Received December 14, 2004

Contents

1. Introduction

Despite its theoretical and practical importance, protein folding remains among the most fundamental unsolved problems in the life sciences. The challenge of predicting folded structure from sequence alone remains unmet, despite promising recent advances and the availability of a wealth of new structures and sequence data. Mechanistically, much still remains to be learned about the process of acquiring a native fold from unfolded proteins. Thermodynamically, the folding reaction in many small single domain proteins, U (unfolded) ↔ N (native), is a reversible, all-or-none, two-state transition, in which only the breakage and formation of noncovalent bonds are involved. In well-studied cases, the process is cooperative, with a preponderance of individual molecules either fully folded or fully unfolded at any time. The unfolded state provides the starting point from which the polypeptide chain acquires its nascent structure and thus dictates the earliest events in the folding reaction. A full understanding of the mechanism of the protein folding process requires a detailed structural, dynamic, and thermodynamic characterization of both the starting and the final states. The final states have been characterized at high resolution by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, as demonstrated by the structures of a large set of proteins in the protein data bank (PDB). The starting states have evaded detailed structural and energetic characterization due to their heterogeneous nature, solubility problems, and complex dynamics.

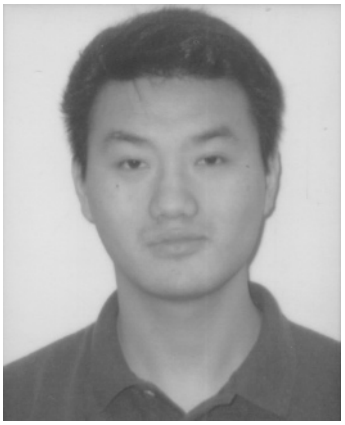
Tanford's hydrodynamic experiments on the chain dimensions of proteins denatured in guanidinium chloride (GuHCl) laid the foundation for thinking about unfolded proteins in terms of a polymeric random coil model.[1] What he showed from measurements of sedimentation and viscosity is that the overall dimensions (the radius of gyration, R_g) of a series of proteins denatured in urea or GuHCl vary with the number of residues (n) in the protein according to a simple power law, n^{γ}, with an exponent γ that is consistent with predictions from polymer chemistry.[2] According to calculations by Flory,[2] there should be no single dominant backbone conformation in an unfolded polypeptide chain. Recent results[3−9] suggest that polyproline II (P_{II}) is a dominant backbone conformation in unfolded peptides. As is discussed below in detail, these two seemingly opposed views of unfolded states of proteins are not mutually exclusive and can actually be reconciled by assuming that unfolded proteins can be represented by a locally scaled P_{II}/β equilibrium,[10] that on a long-range scale is consistent with coil-like behavior,[11] as originally proposed by Krimm and Tiffany.[12] It is necessary to emphasize that this model is substantially different from any classical random coil model.

The fact that unfolded proteins conform to polymeric random coils led to the acceptance of the interpretation that unfolded proteins *are* random coils for over three decades.

* To whom correspondence should be addressed. E-mail: nrk1@nyu.edu.
† New York University.
‡ University of Pennsylvania.

10.1021/cr040433a CCC: $59.00
Published on Web 04/14/2006

Zhengshuang Shi was born in Hubei, China, in 1971. He received his B.S. degree in Chemistry (1992) from Lanzhou University and his M.S. degree in Organic Chemistry (1995) from Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences. He worked at Fudan University Medical Center as a research associate and instructor from 1995 to 1997. Then he moved to the U.S. and obtained his Ph.D. in Biophysical Chemistry at NYU in 2002, working for Professor Neville Kallenbach at NYU and Tobin Sosnick at The University of Chicago. He did his first postdoc with Professor Josh Wand at University of Pennsylvania from 2002 to 2005. He is currently working for Professor Natalie Ahn as an HHMI postdoctoral associate at University of Colorado at Boulder. The problems in which he is currently most interested include protein folding and energetics, NMR structure biology, and the role of allostery in MAP kinases signaling transduction pathways.



Kang Chen was born in Wuhu (China) in 1979. He received his B.S. (2000) in Biochemistry from Fudan University and an M.S. (2002) in Biomedical Science from the University of Connecticut Health Center. He received his Ph.D. (2006) in Chemistry from New York University. As a graduate student with Gregory P. Mullen, he studied the structure and function of a DNA repair enzyme at UConn. Following Mullen's unexpected death in 2002, he resumed his Ph.D. study in biophysics, mentored by Neville R. Kallenbach at New York University. His thesis work is on the structure and thermodynamic characterization of unfolded peptides using NMR and CD spectroscopy. During his time in New York, he also worked on the NMR analysis of protein anisotropic motion with W. Clay Bracken at Weill Medical College of Cornell University.

Zhigang Liu was born in 1976 in Urumqi, China, and graduated with a B.S. degree in Chemistry (1999) and an M.S. degree in Analytical Chemistry (2002) from Sichuan University. He then moved to the U.S. and is currently carrying out graduate work in Neville Kallenbach's lab at New York University and hopes to obtain his Ph.D. in Chemistry in 2007. His thesis research deals with the role of PII conformation in the backbone of unfolded proteins and new designs for antimicrobial peptides.

Neville Kallenbach was born in Johannesburg, South Africa, and received his B.S. (1958) in Chemistry from Rutgers University and his Ph.D. in Chemistry from Yale University (1961) with Lars Onsager. After NSF and NIH Postdoctoral Fellowships at UCSD with Bruno Zimm, he moved to the University of Pennsylvania in 1964, where he became Professor of Biology in 1972. In 1987 he was appointed Professor and Chairman of Chemistry at NYU, a position he held until 1995. He has been a Guggenheim Fellow (1972–73) and NIH Senior Research Fellow (1981–82). He has won the Herman and Margaret Sokol Prize for Faculty Research (1994) as well as teaching awards (1992, 2003) at NYU.

The random coil model is still used as a framework for describing unfolded proteins[13,14] despite persistent doubts.[12,15−18] The model assumes that unfolded proteins represent an ensemble of featureless random chain molecules with very large associated chain entropy. The process of folding is then perceived to be one in which this enormous entropy is reduced to the point that a unique native state can be attained, achieved by favorable enthalpy changes associated with acquisition of hydrophobic interactions, internal H-bonds, and van der Waals and electrostatic interactions.[19] A common conceptualization of protein folding consists of a funnel, in which the entropy is proportional to the diameter and the favorable enthalpy terms accumulate as the diameter shrinks. Modern efforts to study folding are frequently cast in terms of a "landscape" embedded in the model that U states are essentially coils.[14] One recent publication, for example, states explicitly:

Protein folding is inherently a heterogeneous process because of the very large number of microscopic pathways that connect the myriad unfolded conformations to the unique conformation of the native structure.[14]

Recently, several lines of compelling evidence[3−9] have converged to reveal that the backbone conformation of unfolded proteins is predisposed to a particular type of structure, that is, polyproline II (P_{II}). In its extended form, P_{II} is a left-handed helix with backbone dihedral angles (ϕ, ψ) = (−75°, +145°) and three residues per turn, resulting

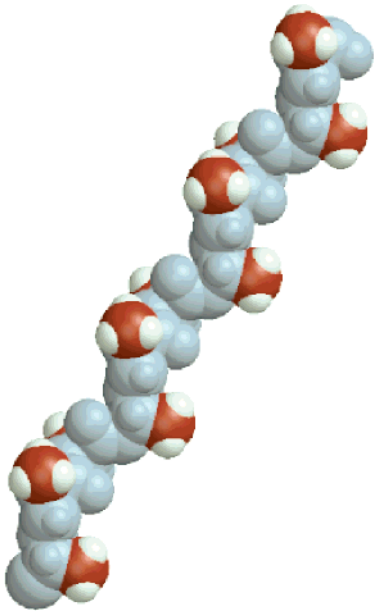


Figure 1. Idealized 12-residue segment of the polyalanine P_{II} helix. The backbone is shown in blue-gray; β carbons are in red, and their hydrogens are in white. Unlike the more familiar α helix, a P_{II} helix is left-handed ($\phi, \psi = -75°, +145°$). It has three residues per turn; that is, every third side chain is collinear, forming three parallel columns spaced uniformly around the long axis of the helix. In solution, significant fluctuations from the idealized structure shown here probably occur. (Reprinted with permission from ref 8. Copyright 2002 National Academy of Sciences, U.S.A.)

in an axial translation of 3.20 Å (Figure 1). This new evidence has stimulated a reappraisal of the structure of unfolded proteins, causing some but not all in the field to adopt a new view,[20−23] reviving a proposal first advanced by Krimm and his group.[12] In their model, unfolded proteins contain substantial locally ordered P_{II} structure but conform nevertheless to long-range disordered chains. Their argument was based on the fact that the UV circular dichroism (CD) spectra of disordered chains of charged poly(Glu) or poly(Lys) resemble those of oligomers of proline and do not conform at all to predictions for a random coil.[24,25] Studying models such as poly(Lys) and poly(Glu), a physical argument based on electrostatics suggested to Krimm that an extended three-fold helical structure might best accommodate charged Lys^+ or Glu^- side chains in a simple dielectric medium.[26] Since the CD spectra of proteins denatured in urea or GuHCl also resembled those of proline peptides, they concluded that P_{II} must be the dominant structure in unfolded proteins as well.[12] It is worth emphasizing that the strong CD band near 198 nm that was attributed to random coils is in fact inconsistent with any freely rotating chain, as was noted earlier by Kauzmann and Schellman, among others.

We argue here that Krimm's model is consistent with the weight of available evidence, including Tanford's hydrodynamic data,[1] newer, more precise SAXS results,[27] and spectroscopic analyses,[3−9,12,15−18] as well as kinetic and loop closure experiments.[28−31] The P_{II} model is still not widely accepted, and attention has been paid mostly by spectroscopists[15−18,32,33] until recently. This part of the history and the recent converging lines of evidence from combined theoretical and spectroscopic studies including UV CD,[17,33] NMR, two-dimensional vibrational spectroscopy (2D-IR),[34] vibrational circular dichroism (VCD),[35] and vibrational Raman optical activity (ROA)[36,37] as of early 2002 were reviewed[23] in a special volume of *Advances in Protein Chemistry* devoted to unfolded proteins, in which it was pointed out that a new view of unfolded states is now timely.[21] In this review, we survey recent evidence that P_{II} is a major backbone conformation in unfolded proteins and discuss some reasons that this might be so. This contribution is intended to complement an earlier one[23] in *Advances in Protein Chemistry*.

The paucity of structural information on unfolded proteins reflects three major problems. First, unfolded proteins tend to be less soluble and more prone to associate in water than native proteins. This limits the range of concentrations over which unfolded proteins can be studied without adding cosolvents. Second, the mainstream tools for high-resolution analysis of native protein structure—NMR and single crystal X-ray diffraction analysis—are not directly applicable to unfolded protein molecules. Third, of course, is the fact that the structure in unfolded proteins is innately fluctuating and dynamic. Five model systems have played major roles in developing our understanding of unfolded proteins:

(i) Denatured proteins in solvents such as urea and GuHCl, systems similar to those studied by Tanford.[1]

(ii) Coil libraries,[38−46] essentially compendia of structural data from the PDB consisting of regions of high-resolution structures of native proteins in which residues from standard secondary structures such as α helix and β strands have been excluded.

(iii) Short peptides[8] that cannot fold to form α helix or β sheets under native folding conditions. Studies of these include binding of peptides to SH3 domains that recognize Pro-rich peptides as well as sequences that adopt the P_{II} conformation.[47,48]

(iv) Charged polypeptides[9] such as poly(Glu) and poly(Lys) which cannot fold due to electrostatic repulsion among side chains at neutral pH.

(v) "Natively unfolded" proteins[49,50] including synuclein, calpastatin, and elastin, that are unfolded *in vitro* or as part of their functional active cycles. This group includes molecules such as glutens that are rich in Pro and Gln.[51]

In the following section, we review new results from theoretical and experimental studies on all these systems since 2002, many of which reinforce the conclusion that P_{II} is a major backbone structure in unfolded short Ala peptides. In section 3, we discuss the generality of the occurrence of P_{II} structure in peptides and proteins with sequence composition other than Ala or Pro residues. We then review recent efforts directed toward establishing a P_{II} propensity scale that includes each amino acid. In section 4, we review coil libraries[38−46] which analyze the nonhelix nonsheet regions of high-resolution structures of native proteins to gain insight into unfolded states of proteins. In section 5, we review some theoretical and experimental studies directed toward explaining why there is a biased structure in the first place and why this biased backbone conformation is P_{II}. Reasons that P_{II} dominates as a major backbone conformation in unfolded proteins have emphasized the role of solvent hydration,[4,52−54] steric effects,[55−57] and side chain−backbone interaction[53,54] as well as a combination of these.[23] Here we discuss the potential role of hyperconjugation,[58,59] which was first applied in an analysis of the structure of collagen by Raines and co-workers,[60−62] and suggest that hyperconjugation may provide additional help in answering the puzzling question of why P_{II} should be preferred over β, for example.

In section 6, we discuss some implications of the proposition that P_{II} is a major backbone structure in unfolded proteins. These include its impact on the protein folding problem, development and refinement of force fields, the structure of natively unfolded proteins, and mechanisms of protein fibril formation. In addition, we review the CD assignments for P_{II} relative to other types of structure, including various turns or loop structures. In section 7, we compare the model originally derived by Krimm and Tiffany,[12] which is endorsed by our group[8,23,63] and a growing number of experimentalists and theoreticians[9,21,22,35,36,54,56,64−69] (referred to here as the P_{II} model), with three other major alternative models of unfolded proteins that have been advanced. In section 8, we conclude by attempting to outline some possible critical future work.

2. Recent Evidence that P_{II} Is a Major Conformation in Short Ala-Based Peptides

2.1. Spectroscopy of Short Peptides

A compelling reason to study very short peptides is that these systems do not fold because they are unable to develop the longer range cooperative interactions required to form α helix or β sheets and, hence, can be investigated under native folding conditions. Short fragments offer the simplest models with which to compare experiment and theory, a major objective for modern protein chemistry. Recently, a variety of spectroscopic studies[4−7] on the conformation of the smallest protein subunits, including blocked alanine (Ac-Ala-NHMe, AAMA) and trialanine, have revealed that these small peptides adopt a well-defined structure in water, significantly P_{II}. The backbone conformation of a seven-residue alanine peptide (XAO)[8] shows that it also predominantly adopts P_{II} based on CD and NMR,[8] as well as resonance Raman[70] and Raman optical activity (ROA) measurements.[71] Analysis of this peptide thus provided a benchmark for assigning spectra of other peptides and proteins to P_{II}.

Since the small methyl side chain of alanine corresponds closely with the methylene group present in all amino acids except Pro and Gly, short Ala-based peptides are arguably representative models for the backbone in natively unfolded proteins. If short alanine peptides favor the P_{II} structure, then P_{II} might in fact be the major backbone conformation in unfolded proteins, in contrast to the more common belief that they are unstructured or statistical random coils. The existence of a well-defined local structure in unfolded proteins has far-reaching conceptual implications for the protein folding problem that we discuss in section 6.

Woutersen *et al.* have investigated AAMA and trialanine in a detailed study using nonlinear time-resolved vibrational spectroscopy.[72] The new work confirms their previous conclusion that trialanine adopts a predominantly P_{II} structure. They find in this study that AAMA has a single well-defined P_{II} structure. Moreover, they have analyzed the spectral inhomogeneity of the corresponding amide I bands in both peptides and conclude that the band in trialanine is significantly heterogeneously broadened relative to that in AAMA. From this they suggest that trialanine exists as an equilibrium of two conformations: one (∼80%) P_{II} and the other (∼20%) the α_R structure. However, the two peptides differ in charge and capping of the ends—AAMA is neutral with caps at both ends, whereas trialanine is +1 charged at its N-terminus under the experimental conditions. How much these differences contribute to the conformational results is hard to judge. This point is of concern also in evaluating the results obtained by Schweitzer-Stenner and his group,[52,64,73−75] who have used a combined spectroscopic approach including polarized Raman, Fourier transform infrared spectroscopy (FTIR), VCD, and CD to examine the structure in a variety of Ala-based short peptides including Ac-A_2, A_3, KA_2, SA_2, and A_4. They conclude that Ac-A_2, A_3, KA_2, and SA_2 exist as a mixture of P_{II} and extended β-type conformations in D_2O[73,74] whereas A_4 is predominantly P_{II}.[75]

Following the structural analysis of a soluble seven-residue Ala peptide (XAO),[8] Kallenbach's group selected a smaller model peptide,[63] $AcGGAGGNH_2$, to investigate the backbone conformation of the central Ala flanked by two Gly residues on each end. In this peptide both ends are blocked to avoid charge effects while the flanking glycines have a relatively high solubility in water. Their high degree of conformational freedom imposes minimal constraints on the conformation of the central alanine. They report that $AcGGAGGNH_2$ has the P_{II} conformation, with an apparent thermal stability lower than that of XAO. Introducing additional Ala residues in the series $AcGGA_nGGNH_2$ reveals no significant cooperative effect as n increases from 2 to 4, indicating that the P_{II} structure is only weakly cooperative (Figure 2; see the figure

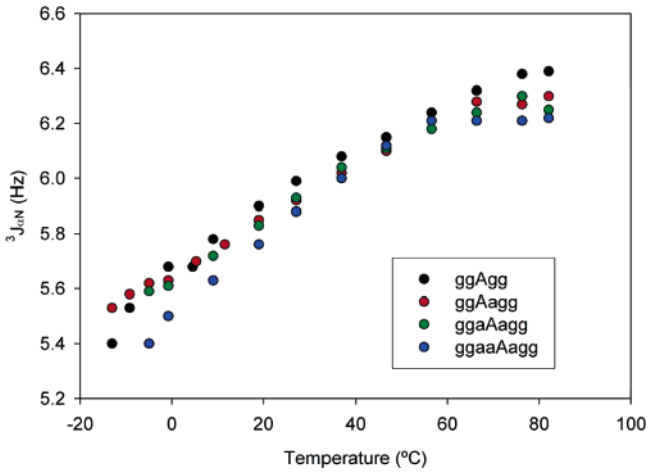


Figure 2. Temperature profiles of $^3J_{\alpha N}$ coupling constants of individual alanines in a series of peptides $AcGGA_nGGNH_2$ (n = 1−4). The measurements were carried out on the ^{15}N labeled alanines shown in capitalized letters in the legend. If there is a strong cooperative effect, we would expect to see distinctive $^3J_{\alpha N}-T$ profiles for the series of peptides $AcGGA_nGGNH_2$ (n = 1−4). We conclude that P_{II} structure is only weakly cooperative, as we observe similar $^3J_{\alpha N}-T$ profiles for the series of peptides. (The related data analysis can be found in ref 76.)

legend).[76] If true for other side chains, this result should greatly simplify statistical thermodynamic treatment of the structure in unfolded proteins. For example, the helix−coil transition theories of Zimm and Bragg[77] or Lifson[78] can be modified quite directly to incorporate a temperature-dependent P_{II}/β equilibrium in the coil state rather than the constant weight assigned to the coil.

Barron's group has investigated a series of short Ala-based peptides with increasing chain length including A_2, A_3, A_4, A_5, and A_7 using ROA.[71] They identify a band near 1320 cm^{-1} in ROA spectra as P_{II}.[37] Their results suggest that the P_{II} content appears to increase with the length of the Ala chain, with A_7 being the most P_{II}-rich peptide, in agreement with the conclusion from a previous study that XAO has a

predominantly P_{II} structure.[8] The lack of cooperative behavior seen in $AcGGA_nGGNH_2$ relative to these results may reflect differences due to the ends of the chains. An interesting point in this study is that pH may perturb the ϕ and ψ angles of P_{II} structures adopted by short Ala peptides with charged ends.

Resonance Raman spectroscopy with excitation at 204 nm is a powerful technique for investigating the structural properties of the backbone in proteins.[79] The normally low intensity of Raman bands is enhanced in the presence of a strong absorption band such as the $\pi-\pi^*$ amide band in peptides. Resonance Raman spectroscopy of the amide bands offers several sensitive windows for investigation of peptide and protein conformation, including Am III bands at 1302 and 1244 cm^{-1}. Importantly, these bands are sensitive to the ψ dihedral angle in peptides, which is difficult to determine. They propose a general relation of the form

$$\nu = \nu_0 + A \sin(\psi + \delta)$$

and have fit the constants A and δ for AmIII3. Use of the XAO peptide model together with trialanine and A5 shows that a longer α helical peptide AP, with sequence AAAAA-$(AAARA)_3$A, unfolds to a predominantly P_{II} conformation.[70] This result emphasizes the fact that the helix−coil transition in a simple case such as AP occurs between two major states: α_R at low temperature and P_{II} at high temperature.

Weise and Weisshaar have extended their previous analysis[4] of AAMA by means of liquid crystal NMR spectroscopy (LXNMR). They have used a sample singly labeled with ^{13}C to derive four additional dipolar coupling constants[80] and determine a set of favored dihedral angles with $(\phi, \psi) = (\sim -85°, \sim +160°)$, confirming results from their earlier study.[4] A recent related study on tetraalanine by Pizzanelli *et al.* has concluded that their experimental data are compatible with the presence of a single conformation, P_{II}.[81] One issue in these experiments is whether LXNMR or related orientation experiments oversample extended conformations relative to compact ones in flexible systems.[82] For residual dipolar coupling constant measurements, there are concerns on the local environment of oriented gels as the results rely heavily on the behavior of a small percentage (on the order of 1 out of 1000) of the whole population of studied molecules. Creamer and colleagues have recently examined the structures of A_1, A_2, and A_3, each of which is flanked by several proline residues at both ends, by CD.[83] They find that the P_{II} structure is present in each of the Ala residues in the series, contradicting theoretical predictions by Scheraga's group[84] that we discuss below.

2.2. Simulations of Short Peptides

Short unfolded alanine peptides have offered minimal models for theoretical analysis of peptide conformation and solvation for many years. While there are significant differences among the force fields and water models that are in use, a number of recent calculations offer support for the experimental result that the P_{II} helix is the preferred conformation in unfolded short alanine peptides. Garcia has used a reparametrized AMBER force field to conclude that alanine peptides of 8 or less have a dominantly P_{II} structure.[66] Mu *et al.* predict that solvated trialanine has a predominantly (∼80%) extended conformation including both β and P_{II} using the GROMOS96 force field.[85] However, using a new version of the OPLS force field combined with a TIP5 water model, the numbers shift slightly to population probabilities of 65% for P_{II} and 12% for β.[86] Kentsis *et al.* find that polyalanine peptides exist as a segmented P_{II} helix in the unfolded state with each segment comprising two to seven residues.[68] Mezei *et al.* also predict that the P_{II} helix is the preferred conformation for unfolded polyalanine peptides in water,[87] in agreement with Kentsis *et al.*[68]

2.3. Debates on the Quantitative Level of P_{II}

Scheraga and his colleagues have recently published a theoretical study[84] to explore the P_{II}-forming capability of nonproline residues in a prolyl background, modeling the host−guest experiments of Creamer's group.[54] They argue that substituting non-proline residues such as Ala and Gly in oligo-proline peptides results in significant reduction of the CD signal near 220 nm, reflecting significant loss in P_{II}-helix content compared to that of an all-proline peptide. Their theoretical results using the ECEPP force field are in agreement with this analysis, leading them to conclude that previous estimates exaggerate the P_{II} content in alanine peptides, including the XAO peptide studied by Shi *et al.*[8,23] In particular, they reinterpret the CD data on XAO as lending support to their conclusion.[84] Recently, Scheraga's group has remeasured the CD and NMR properties of XAO, concluding that there cannot in fact be a high level of P_{II} in this molecule at all.[91] These results are discussed below.

We find these conclusions hard to accept for the following reasons. First, they contradict the weight of independent spectroscopic data on short alanine peptides[4−7,52,72−75,88,89] that demonstrate P_{II} is a major conformation. The evidence is discussed in section 2.1 above as well as in our earlier review.[23] Second, estimating P_{II} helical content from CD values alone requires caution, as demonstrated in a recent paper by Sreerama and Woody.[90] Figure 2 of that paper[90] shows that the P_{II} CD spectrum derived by deconvoluting CD data from a reference set of globular proteins is considerably different in magnitude for each band compared with a reference P_{II} CD spectrum. The reference CD spectra for peptides composed of different amino acids (even the same amino acid at different conditions such as differences in capping ends, pH, and neighboring residue environments, etc.) need not be identical in any case if they occupy slightly different (ϕ, ψ) basins in the Ramachandran plot; that is, CD is exceptionally sensitive to small deviations in dihedral angles from the canonical P_{II} structure. One line of potential evidence comes from a recent study by Schweitzer-Stenner and co-workers,[89] which shows that cationic and zwitterionic proline peptides have totally different CD spectra (Figure 6 of ref 89). We believe that the difference may represent distinctively different backbone dihedral angles (Table 2 of ref 89), with neither being in classical P_{II}. This point is discussed further in section 6.4.

As further evidence to support their conclusion, Scheraga and colleagues pointed to an experimental study using isothermal titration calorimetry by Ferreon and Hilser that indicates the probability of Ala forming P_{II} in the denatured state of an SH3 domain binding peptide is ∼30%,[47,48] in agreement with their theoretical predictions from the ECEPP force field.[84] However, there is some uncertainty regarding Ferreon and Hilser's definition of P_{II}, as those conformations are binding-competent to the SH3 domain, regardless of ϕ and ψ angles. It is possible that only a subset of existing P_{II} conformations are detected by the method, a point noted by

the authors themselves.[47,48] Furthermore, their experimental results are based on the assumption that the intrinsic binding energy for different peptides (wild type or Ala/Gly mutants) is the same once the peptides adopt the P_{II} conformations, which is not necessarily true. Strictly speaking, their results report the P_{II} bias of the peptide as a whole, rather than the position-specific P_{II} bias, as claimed.[47,48] We expect that there can be effects on the conformation of neighboring residues of the peptide (in either the bound or unbound states) resulting from the substitutions. Nevertheless, the study by Ferreon and Hilser is a fundamental step toward our deciphering the conformations of the backbone in unfolded proteins.

Scheraga and colleagues have recently re-examined the XAO peptide in detail, using a combination of NMR and CD spectroscopy and theoretical calculations.[91] They conclude from this new study that XAO samples a diverse set of conformational states clustered about three vertical strips in the Ramachandran plot with ϕ angles of $-160°$, $-70°$, and $+60°$, respectively, in disagreement with our previous findings.[8,23] This interesting study does not consider the strong temperature dependence of ^3J coupling constants and CD values as an indication of conformational transition. Instead they interpret the transition simply as a shift in the populations among the different ϕ regions. Since the transition might be broad and the one we observed is not complete, it is too early to dismiss this possibility, although the apparent enthalpy seems high for such a redistribution. Another fundamental difference concerns the interpretation of the CD spectra. We consider a CD spectrum with a strong negative band at ~198 nm to correspond to a situation in which P_{II} dominates, regardless of whether the value at ~218 nm is below or above zero. This takes us back to an old argument: as Krimm[92] and Schellman[93] and others asked long ago, how can the CD spectra of unfolded proteins and XAO be so intense, if a blend of random conformations is involved? It is hard to reconcile these spectra with any conformational blend that includes significant β or α populations. Scheraga's own CD spectra thus might be taken to support our view that there is a dominant conformation present in XAO, independently of NMR results. The complexity of the CD spectra at 220 nm can be explained by salt effects on the charges at the ends.

The new results weight heavily on the observation of ROE cross-peaks: they detect weak ROE signals between the amide protons of X1−X2, A6−A7, and A9−O10. The discrepancy might come from our difference in judgment of cross-peak intensity; it is possible that there are some very weak NOEs that we consider insignificant. There might also be differences between ROESY and NOESY, and be tween mixing times as well as solvent suppression pulses. At one point they argue that the observed ROEs suggest that "*the XAO peptide cannot exist in the P_{II} structure because no $H_i^N-H_{i+1}^N$ ROE signals should be observed in the P_{II} structure*"; elsewhere they argue that "*the ratio of the $H_i^N-H_{i+1}^N$ NOE intensity computed for an all-P_{II} conformation to that computed for an α-helical conformation is 1:4, which suggests that NOEs should be observed if the whole chain were to adopt the P_{II} conformation, even though such NOEs are weak*". The interactions in the Ala region of the chain are noncooperative and local, and therefore, the whole chain would never adopt P_{II} at a given time as we anticipated. Observing the weak ROE cross-peaks is hardly decisive evidence for or against the presence of a P_{II} helix. Admittedly, we have used the argument that the absence of any measurable NOEs between successive amides in the chain indicates that no measurable α helix or other α conformation is present.

From Figure 5 of their paper, conformations calculated for XAO are most sensitive to restraints derived from ^3J coupling constants. As we pointed out in our paper, there is an intrinsic limitation on NMR data in that NMR alone cannot resolve whether there is an averaging of different backbone conformations in XAO.[8] However, other spectroscopic techniques such as 2D-IR,[34,94] ROA,[71] and UVRR[70] that are capable of resolving this issue consistently report that short Ala peptides sample predominantly the P_{II} conformation (see section 2.1). We believe Scheraga's arguments concern the size of populations that are present more than anything else, and there is still some uncertainty regarding these and how they change with temperature. Additional work will be needed to settle all these questions. It is our view that their new paper still supports our earlier interpretation that there is a preponderance of P_{II} at low temperature in XAO and that, on heating, the population of β structure increases. Our data say nothing about the presence of a small population of β turns in the molecule that may or may not affect the overall dimensions.

3. P_{II} Propensity: Is the P_{II} Structure Present in Chains Composed of Non-alanine Non-proline Residues?

As detailed in the above section, independent lines of evidence indicate that oligomers of Ala in solution assume a predominantly P_{II} local conformation. Rucker and Creamer[9] have shown that K_7 is predominantly P_{II}. Additional studies on ubiquitin fragments by NMR and CD, and alanine-rich peptides containing QQQ, SSS, FFF, and VVV (a series of 11-mers) by CD indicate that they all contain a significant amount of P_{II}, in equilibrium with β structures (see Figure 3 for their CD spectra).[95] In a much longer naturally unfolded peptide consisting of a chain of 52 amino acids, 20 of which are alanines, each alanine was shown to have P_{II} conformation at 40 °C.[96] At lower temperatures, a short α helix nucleates near the C terminus of this molecule. Taken together, these results suggest that longer oligomers contain a significant amount of P_{II}. Proteins unfolded by GuHCl or urea also have a high content of P_{II} conformation, as we noted in our earlier review.[23] Indeed, high urea or GuHCl concentrations favor P_{II}.[97,98] The evidence from ubiquitin fragments suggests further that the occurrence of P_{II} structure is general for nonrepeating sequences and amino acids other than Ala.[95]

In line with the results presented above, Schweitzer-Stenner and his colleagues[88,89] have reported that K_3, D_3, and E_3 as well as P_3 are predominantly P_{II}—or somewhat distorted P_{II}—in D_2O at acidic pD. However, they find that V_3 and S_3 adopt a predominantly β-like structure.[73] Tamburro and his group[99−103] have studied the solution structures of a series of short exonic peptides dissected from human tropoelastin and elastin and find significant P_{II} structure in all of them. These peptides consist mainly of G, A, and L amino acids (for example, the sequence of Exon 3 from human tropoelastin is GAGLGALGG). Following their structural characterization of three short peptides corresponding to sequences in titin,[104] Ma *et al.*[105] have recently presented a detailed conformational analysis of these peptides and obtained results confirming their previous findings:[104] these peptides adopt three conformational states—P_{II}, β turn, and unordered spacer

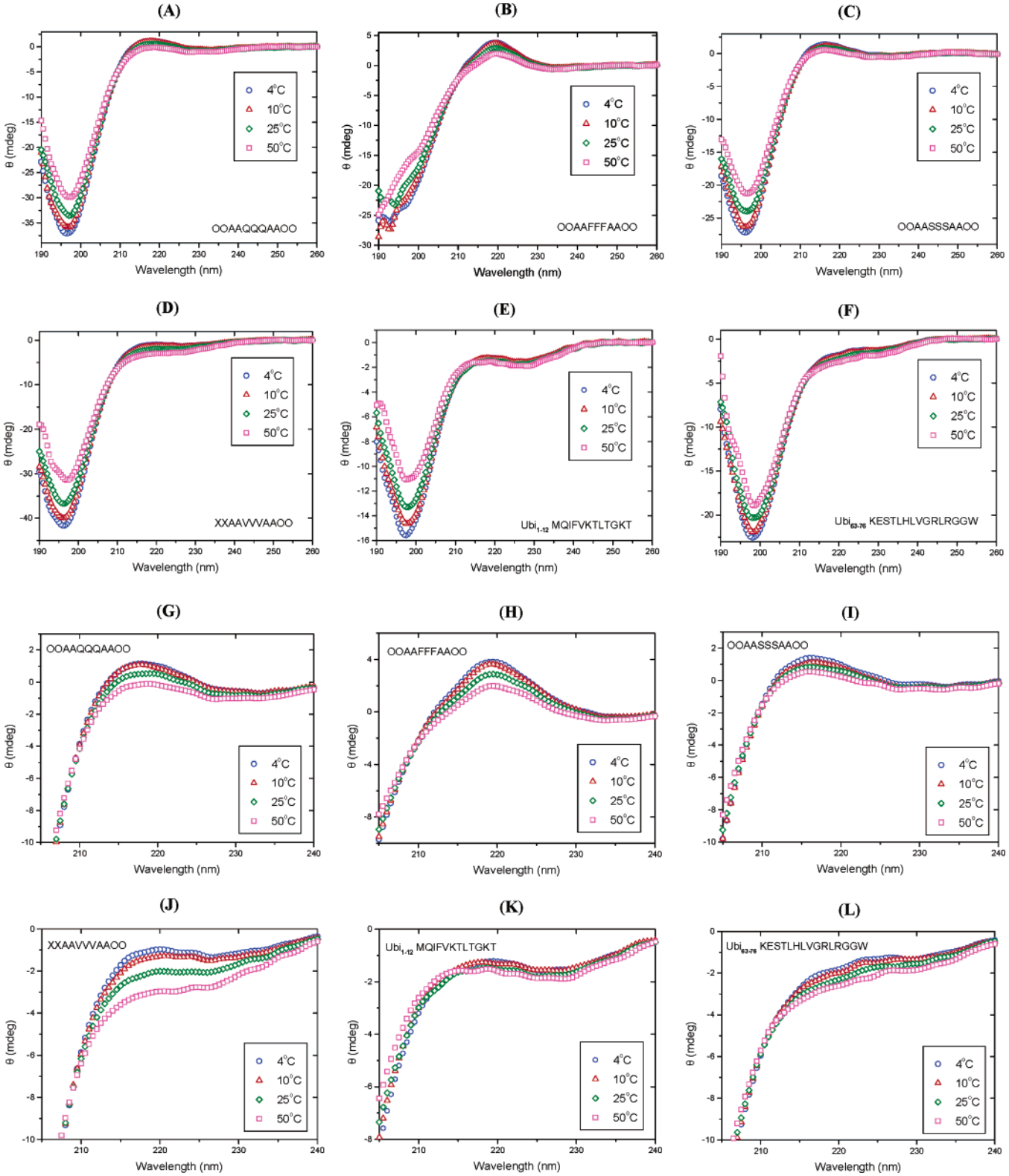


Figure 3. UV CD spectra of short model peptides and ubiquitin fragments. CD spectra of QQQ (A), FFF (B), SSS (C), VVV (D), Ubi_{1-12} (E), and Ubi_{63-76} (F) at (○) 4 °C, (△) 10 °C, (◇) 25 °C, and (□) 50 °C. Parts G−L are the enlargements of the corresponding CD spectra: QQQ (G), FFF (H), SSS (I), VVV (J), Ubi_{1-12} (K), and Ubi_{63-76} (L). (Reprinted with permission from ref 95. Copyright 2006 John Wiley & Sons, Inc.)

regions; the relative content of P_{II} increases with decreasing temperature and increasing the polarity of the solvent; the content of β turn decreases with increasing temperature and the polarity of the solvent, while the content of "unordered coil" increases with increasing temperature and the polarity of the solvent.

Given that the occurrence of P_{II} structure is not exclusive to Ala or Pro and may be general, an obvious question arises. Is there a reproducible P_{II} propensity scale for each amino acid? Several groups have addressed this question recently. By combining results from a host−guest study of AXA peptides with data from related molecules, Schweitzer-

Stenner and his colleagues conclude that amino acids can be grouped according to their structural preference in aqueous solution.[64] They find that K, D, E, Y, and P share a strong preference for P_{II}, while G, M, L, and A sample P_{II} and β strand conformations about equally. On the other hand F, V, H, W, and S strongly prefer a β strand conformation. These results contradict the prediction from random coil models, although the charged ends may influence the assignment of polar or charged side chains in this series. This scale agrees imperfectly with data from Creamer's group[54] or $AcGGXGGNH_2$ peptides,[10] as well as with the results from the survey of coil libraries.[106]

Creamer and his colleagues have determined a different scale of P_{II}-forming propensity for the amino acids, except Tyr and Trp, in a longer proline-based peptide system.[54] Their P_{II} propensity scale is based on the relative intensity of the positive CD band near 220 nm in a series of host−guest peptides with the sequence $AcPPPXPPPNH_2$. Creamer's scale has some features in common with that of Schweitzer-Stenner although the two scales are determined in entirely different ways from different model systems. Specifically, in both scales P, D, E, and K have a high propensity to form P_{II}; G, L, and M are intermediate; and V and H have a low propensity. On the other hand, the side chains A, Y, F, and S differ significantly in their propensity to form P_{II} according to the two scales. One obvious cause of differences stems from the steric constraint imposed on a guest side chain by flanking proline residues in addition to the charged end effects.

Kallenbach's group[10] has studied 19 amino acids in the context of the host−guest peptide model $AcGGXGGNH_2$, with X = G excepted. The CD signals for all $AcGGXGGNH_2$ peptides except those with ring side chains show P_{II} character as well as a progressive increase in the band around 190−200 nm and a decrease at the band around 210−220 nm with temperature. Assuming the unfolded model peptides sample only two basins (P_{II} and β) in the Ramachandran plot, the authors have derived a P_{II} scale for all amino acids based on their measured $^3J_{\alpha N}$ coupling constants (Figure 4).[10]

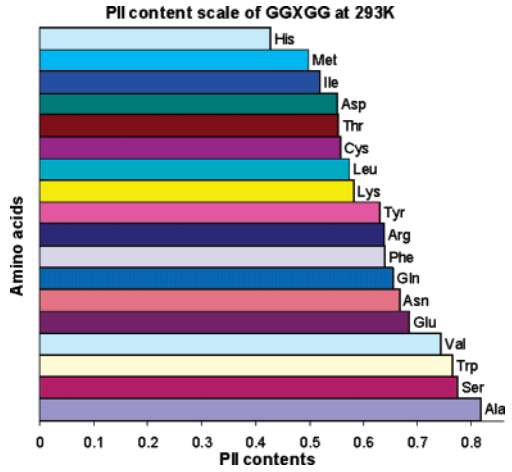


Figure 4. Derived P_{II} content scale of GGXGG at 293 K. (Reprinted with permission from ref 10. Copyright 2005 National Academy of Sciences, U.S.A.)

It is clear that not all amino acids (e.g. His) favor P_{II}. This is still consistent with the conclusion that P_{II} predominates as a major backbone conformation in the unfolded states of proteins. Several spectroscopic probes of the backbone dihedral angles in short unfolded Ala-based peptides point to P_{II} as the major structure. As noted above, Ala is a reasonable representation of the backbone in all amino acids excluding Gly and Pro. One might consider the preference of conformations other than P_{II} (β structure, for example) in some amino acids as a deviation from P_{II} due to perturbation by side chain specific solvation effects as in Baldwin's analysis,[107] steric clashing, and/or specific side chain−backbone interactions.[108] Comparison of three P_{II} scales derived from totally different systems suggests that the P_{II} scale is likely to be sequence and context dependent.

4. Coil Library Surveys

4.1. Introduction to Coil Libraries and Early Library Surveys

Despite its origin in the native states of proteins, the PDB has provided a valuable resource for analysis of protein conformation in unfolded states. Surveys of individual amino acids in the PDB suggest that each possesses its own intrinsic distribution in specific ϕ and ψ basins of the Ramachandran plot. Chou and Fasman initiated the systematic use of a set of protein X-ray structures (at first with ~20 proteins) for α helix and β strand secondary structure prediction.[109] The assumption underlying use of these libraries is that near-neighbor effects and long range interactions should cancel given sufficient sample size, so that intrinsic conformational preferences for individual side chains can emerge. Generally speaking, "full" libraries sample all the residues in the protein database, whereas "coil" libraries sample only residues that lie outside recognizable secondary structure regions. Both full and coil libraries have been used in efforts to predict protein structure from its sequence as well as to assess the properties of unfolded proteins and peptides.

In the early 1990s Garnier and co-workers investigated the local effect on the conformation of a given residue by surveying 61 Brookhaven proteins.[38] Thornton and co-workers examined residues from 463 3D protein structures to acquire ϕ and ψ distribution data excluding Pro and Gly,[39] intending to evaluate the quality of protein X-ray structures available at the time. Lee and co-workers attempted to study neighbor effects in predicting peptide secondary structure using a set of 55 high-resolution structures.[40] While none of these studies intended to find P_{II}, the P_{II} region is nevertheless densely occupied in their statistical Ramachandran plots (Figure 1 in ref 38, Figure 2 in ref 40, and Figure 5 in ref 39). At the time, the P_{II} structure was classified within the extended β region associated with "coil" and escaped special attention.

Serrano and colleagues distinguished the P_{II} region from β in their statistical analysis of α helix and β strand probabilities.[41] Their results derived from 279 protein structures correlated well with the experimental scales of α and β propensities after excluding P_{II}. Serrano's P_{II} basin is centered at $(\phi, \psi) = (-72° \pm 15°, +144° \pm 15°)$ (Figure 1 in ref 41).

4.2. Coil Libraries and P_{II}

4.2.1. Problems with the Full Libraries

To calibrate NMR parameters for unfolded proteins, Dobson and co-workers applied the PDB to random coil structure analysis.[110] At least two problems affect the use of

full libraries to predict properties of unfolded proteins: (i) The cooperativity of α helix in folded structures will tend to bias the prediction to α structure. (ii) The dielectric environmental difference between the folded core and the corresponding domain in a denatured protein might be associated with a large enthalpy difference (~10 kcal/mol) based on analysis of solvation of peptide bonds.[87,111] Dobson and co-workers noticed that the $^3J_{\alpha N}$ coupling constants predicted from the coil library are systematically larger by 0.4−0.5 Hz than those from the full library, reflecting the over-representation of α helical conformation in the full library compared to the coil library.[110]

4.2.2. Early Coil Libraries

Swindells, MacArthur, and Thornton distinguished the coil region residues from other regions of regular secondary structures using a data set of 85 high-resolution protein structures.[42] They excluded residues from α helix and β strand but included β turns. Their objective was to acquire intrinsic probabilities for α and β structures using the coil library. In fact P_{II} emerged as the dominant structure. At least one-third of the coil residues clustered into their "p" basin (ϕ, ψ) = (−71° ± 15°, +144° ± 15°) (Figure 1 in ref 42).

Thornton, Dobson, and co-workers calculated residue specific $^3J_{\alpha N}$ coupling constants from the distribution of ϕ angles in both a full library, "All", and a "COIL" version. They compared the results with experimentally determined $^3J_{\alpha N}$ coupling constants derived from a series of 10−20 residue peptides. Their COIL dataset (R = 0.92) correlates better than the All dataset (R = 0.81).[43] In both datasets residues were classified into two groups, core α and core β, the distributions of which are 45% α and 40% β in the All dataset and 28% α and 44% β in the COIL dataset. In fact, a large number of P_{II} residues in the COIL dataset were classified into the core β area. The coil library was successful in explaining the $^3J_{\alpha N}$ coupling constants of a 17 residue peptide corresponding to the C-helix of hen lysozyme dissolved in water (Figure 2 in ref 112). Interestingly, the fact that P_{II} structure dominates the coil library was not pointed out or widely accepted at that time.

Almost at the same time as the work of Swindells and Thornton, Serrano studied the ϕ angle distribution from 279 protein structures.[44] He noticed the relatively large increase in P_{II} population and the dramatic decrease of α helix in the coil library and the possibility of P_{II} conformation contributing to the random coil state. In his view, random coil structure should include P_{II} as well as β.

We should emphasize here that an earlier analysis[113,114] by Adzhubei and Sternberg has crucial implications for the role of P_{II} conformation in unfolded proteins. They observed that short segments of P_{II} helices commonly occur in the crystal structures of a set of 80 globular proteins while many of these segments contain no Pro residue. Moreover, they noted that P_{II} helices tend to lie on the surface of the proteins with few main-chain hydrogen bonds to the rest of the residues. They suggested that these P_{II} side chains are possibly stabilized by solvation, as revealed by the H-bonding of water molecules to the peptide NH and C═O groups. These observations have been confirmed in a more recent survey by Stapley and Creamer.[115]

4.2.3. More Recent Coil Libraries

Hermans and co-workers compared dipeptide simulations with database statistics for the Ala, Asn, Asp, Gly, and Val side chains in 109 proteins.[45] Their coil library distribution includes three regions β, αR, and αL for all residues except Gly. They find two minimal energy basins in the β area, as discussed in the paper (Figures 3−7 in ref 45), one assigned as extended β strand and the other as type II β turn. The formation of a standard type II β turn needs two continuous residues with dihedral angles (ϕ_2, ψ_2) = (−60°, +120°) and (ϕ_3, ψ_3) = (+90°, 0°).[116] The first residue of this β turn structure is close to P_{II}. A survey of the coil database for Asp revealed that one of the deepest minima in the β region of the Ramachandran plot is located at (ϕ, ψ) = (−70°, +140°), which is P_{II}.

Ohlson and co-workers reported the largest coil library survey up to 2002, derived from analysis of 1042 protein subunits. About 55% of the dihedral angles of the coil residues lie in the upper right corner of the β region,[46] corresponding to P_{II}. This high P_{II} distribution is common for all natural amino acids except Gly (Figure 9 in ref 46). A still larger and more comprehensive library has been constructed by Jha *et al.*[106] Their final coil library is based on over 100 000 residues excluding potential capping residues within α and β structures, as well as turns. The result based on these extremely stringent selection rules is quite striking: P_{II} dominates the coil library relative to all other conformations. While Pro is most represented by default, Ala is next, followed by Leu.

4.3. Potential Limitations of Coil Libraries

A conclusion that emerges from using coil libraries to decipher the major backbone conformation of unfolded polypeptide is that P_{II} is a major structure in the unfolded state (at least 55%) while there is lower but significant presence of other structures, including α and β. There are still several cautions in applying coil libraries to predict the structure of unfolded polypeptides. First of all, the long range interactions present in folded protein structures may not cancel completely as assumed. For example, tertiary constraints in globular proteins might tend to favor turn conformations in the coil library. Use of coil libraries from which all turns are excluded might or might not improve the accuracy, as seen in Jha *et al.*'s study.[106] The question is how much structural information in the PDB should be retained or thrown out in constructing a coil library. Second, the partial hydration of coil residues may not correspond to that in highly solvated peptides and unfolded proteins. This is the same concern we pointed out in section 4.2.1 in comparing use of full libraries vs coil libraries. The side chains of hydrophobic residues tend to be buried inside the protein rather than exposed, while hydrophilic side chains tend to populate the protein surface. Indeed Jha *et al.*[106] interpret their coil library data to imply that hydration is not involved in stabilizing P_{II}, in contrast to data we discuss below. In unfolded proteins, more residues would obviously be expected to be solvent exposed than in any regions within native proteins.

As pointed out by a reviewer, one basic argument in favor of using a coil library to estimate the likelihood of identifying backbone conformations in unfolded proteins is that proteins are assembled from low-energy components according to Butterfoss and Hermans,[117] implying that the backbone dihedral angles found in the coil library should correspond to those found in unfolded proteins. A problem is that backbone conformational propensities derived from X-ray structures are weighted differently in different libraries,

including those reviewed above. Thus, probabilities should be quoted as argued by Shortle.[118] Shortle also pointed out that propensities might be useful in recognizing the native structure in studies of threading peptide fragments, whereas probabilities cannot be used in the same way.[119] He makes the distinction that edge backbone conformations cannot be substituted by core conformations in threading approaches to protein structure prediction. In many coil library surveys, there is no discrimination between the residues from edge and core regions, particularly for those in P_{II} and β conformations.

5. Why Is P_{II} a Dominant Conformation in Unfolded States of Proteins?

5.1. Generally Accepted Views

It is not easy in principle to account for a strong preference of protein backbone for any one conformation within the upper left-hand corner of the Ramachandran plot. Regarding the bias of backbone conformation to P_{II}, recent studies have emphasized the roles of solvent hydration,[4,52−54] steric effects,[55−57] side chain−backbone interaction,[53,54] or a combination of these.[23] One line of evidence from coil libraries suggested that residues in P_{II} occur in highly hydrated regions of native proteins.[113,114,115] This evidence could be circumstantial, as the most extensive library analyzed to date fails to detect any such effect.[106] On the experimental side, most conclusions concerning why P_{II} is favored are derivative or based on speculation.

In a recent study, Eker *et al*. report that acetyl-Ala-Ala samples both P_{II} and β structure in D_2O, whereas it adopts a single β-like structure in DMSO.[74] This is consistent with the idea that hydration plays an important role in the stabilization of P_{II} structure. In another report, Eker *et al*. have characterized structures of tripeptides in both H_2O and D_2O by CD and find that there is an isotope effect in the P_{II} population in aqueous solution.[52] They conclude that a hydrogen-bonding network involving the peptide and water molecules plays a major role in stabilization of P_{II} conformation. However, detailed interpretation of such an isotope effect is far from simple, as we have demonstrated in α helical peptides[120] and proteins[121] although their conclusion may be correct. Based on their experimental design, the isotope effect they observe reflects not only the difference in a peptide−water hydrogen-bonding network but also that in the water−water (D_2O vs H_2O) hydrogen-bonding network as well as other water−water interactions such as van der Waals effects (D_2O vs H_2O).

Kallenbach's group has investigated the solvent effect on a small neutral model peptide $AcGGAGGNH_2$ and report that the conformation of $AcGGAGGNH_2$ changes from P_{II} to C_{7eq} upon switching the solvent from water to neat TFE.[97] These results are consistent with a direct role for solvent hydration in stabilizing P_{II} conformation. The effect on P_{II} of simple alcohols correlates best with empirical scales of solvent polarity, rather than dielectric constant *per se* (Figure 5). This is consistent with a role for water in hydrating the backbone rather than acting as a bulk solvent, although this is by no means proven. Tamburro and his colleagues have also reported solvent effects on short exonic peptides from human tropoelastin by CD and NMR.[101,103] They suggest that changing the solvent from water to TFE results in an increased presence of β turns, which might rapidly interconvert with P_{II} structure. Chellgren and Creamer[83] find that

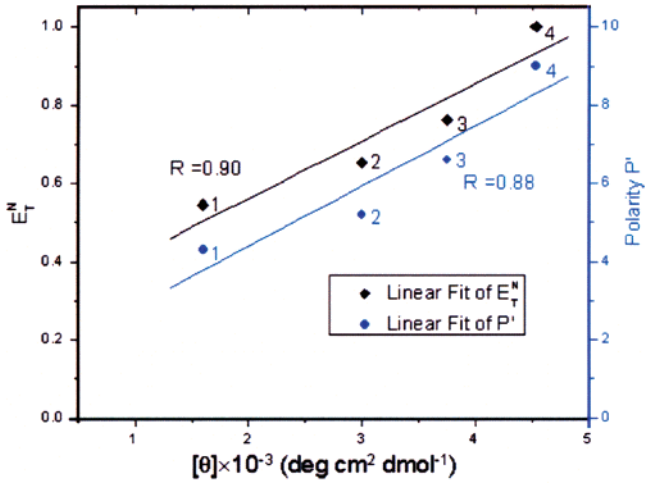


Figure 5. Correlations between the CD signal (absolute value of the minimum CD signal of the peaks) of $AcGGAGGNH_2$ and measures of solvent polarity. The subscripts 1−4 denote 2-propanol, ethanol, methanol, and water, respectively. The black line shows a linear fit to the E_T^N scale, and the blue one corresponds to the P' scale. The correlation coefficients are 0.90 (E_T^N) and 0.88 (P'). Both TFE and MeCN deviate strongly from the linear correlation for aliphatic alcohols. (Reprinted with permission from ref 97. Copyright 2004 American Chemical Society.)

D_2O stabilizes P_{II} relative to H_2O, consistent with a role for hydration in maintaining P_{II} structure. It is interesting that the stabilizing effect of D_2O is more dramatic for Val-containing peptides than for Ala-containing peptides.[83] It is worth noting that Creamer's group also studied the effect of salt bridges[122] on the stability of P_{II} helices and find that saltbridges do not stabilize P_{II}, in contrast to their role in stabilizing α helices[123−125] and proteins.[126,127]

5.2. Thermodynamics and Kinetics

The thermodynamics of P_{II} formation has been investigated by Hilser's group, using isothermal titration calorimetry to measure the interaction between peptides and an SH3 domain that selects the P_{II} conformation.[47,48] As the first careful calorimetric characterization of the thermodynamic origins of the P_{II} preference in the unfolded states of proteins and peptides, they have found that P_{II} formation is favored enthalpically by 1.7 kcal/mol per residue and is opposed by an entropy change of 0.7 kcal/mol per residue. Transition profiles in short alanine peptides were used to derive a van't Hoff enthalpy of about 3.3 kcal/mol[23] for P_{II} formation. This value and the one cited above are within the range we derive from the model peptide series $AcGGXGGNH_2$.[10] The analysis of electrostatic solvation free energy (ESF)[107] by Avbelj and Baldwin points to potentially large enthalpic contributions that are associated with hydration of peptide bonds.

Recent theoretical research has concentrated on the roles of solvation and steric effects.[128] Simulation studies by Mezei *et al*.[87] and Kentsis *et al*.[68] indicate that the solvation free energy of P_{II} is more favorable relative to other conformations because of an entropic effect, reminiscent of the hydrophobic effect. Garcia's work[66] emphasizes the formation of a delocalized water channel surrounding the groove of the P_{II} helix. He further suggests that a segment at least four residues long may be particularly conducive to water channel formation and may help to stabilize the P_{II} structure. By calculating the ESF, Avbelj and Baldwin[107] conclude that the preference of alanine for P_{II} conformation arises largely

from electrostatic interactions between peptide NH and CO dipoles and screening of these interactions by water. The screening interactions are shielded only slightly by the alanine methyl side chain but more strongly by bulky side chains. Further, they find a correlation between the calculated ESF and the observed neighboring residue effect on the backbone conformation as revealed by a coil library survey as well as the NH hydrogen exchange protection factors of Bai *et al.*[129] Avbelj and Baldwin conclude that ESF alone accounts for many aspects of backbone conformation preference. A more recent analysis of coupling constants in the series of blocked amino acids Ac-X-*N*-methylamides shows that these peptides already exhibit strong conformational preferences,[130] which correlate well with those from coil libraries as well as the GGXGG series. More interestingly, the intrinsic backbone preference in these "dipeptides" can be predicted accurately using the electrostatic screening model of Avbelj and Baldwin.

Pappu and Rose[56] point out that steric effects are critical by demonstrating that minimization of chain packing (using only a repulsive model potential) is sufficient to favor P_{II}. However, Pappu and his colleagues[57] find that peptide−solvent interactions stabilize P_{II} relative to β conformation if the sampling space is limited only to the extended conformation region. Another interesting finding by Fitzkee and Rose[65] indirectly justifies a major role for steric effects; they demonstrate that an α helix cannot be followed by a contiguous β strand due to steric collisions alone.

An estimate of the kinetics of P_{II} formation has been published[131] in a time-resolved FTIR analysis of the helix−"coil" transition in PHPG (poly-N^5-(3-hydroxypropyl)-L-glutamine), the host peptide used in Scheraga's original determination of helix formation parameters of guest amino acids. While the rate of α helix formation in PHPG following a rapid laser-induced temperature jump is of the order of hundreds of nanoseconds, typical of α helix formation rates, Petty and Volk[131] detect a much faster process on a time scale of about 10 ns (close to the limit of resolution of their instrument). They attribute this to a transition between P_{II} and some more random state in the unfolded peptide. By analogy with the 1645 cm^{-1} amide I′ band seen in poly (Glu), they assign the band at 1648 cm^{-1} in PHPG to P_{II}. This band is present at all temperatures in the spectrum of PHPG, gradually decreasing in intensity with temperature. A more recent temperature-jump/UV resonance Raman study on poly(L-lysine) by Spiro and colleagues[132] reports that the melting of α helices to P_{II} is on the order of ~250 ns, similar to the time constant reported for the melting of short Ala helices by Asher and co-workers;[133] they suggest that there is a rapid equilibrium between P_{II} and β strand on a time scale within their ~40 ns instrument resolution. These studies suggest that the rate of transitions between P_{II} and other extended conformations is on the 10 ns time scale, consistent with a highly dynamic manifold. Rates this fast should be accessible to theoretical simulations, and they offer an additional test of models.

5.3. Hyperconjugation

At the orbital level, the theory of hyperconjugation connects a variety of seemingly unrelated phenomena, including the stability of the staggered structure of ethane,[58,59] the *gauche* effect, and the anomeric effect.[134] Hyperconjugation arises from partial electron transfer from an occupied (bonding) orbital to an unoccupied (antibonding) orbital, leading to delocalization of charge and thereby additional resonance stabilization.[58,59] In many organic chemistry texts, ethane is presented as the classical example of a sinusoidally varying potential energy profile accompanying rotation around the central carbon−carbon bond (Figure 6a); the

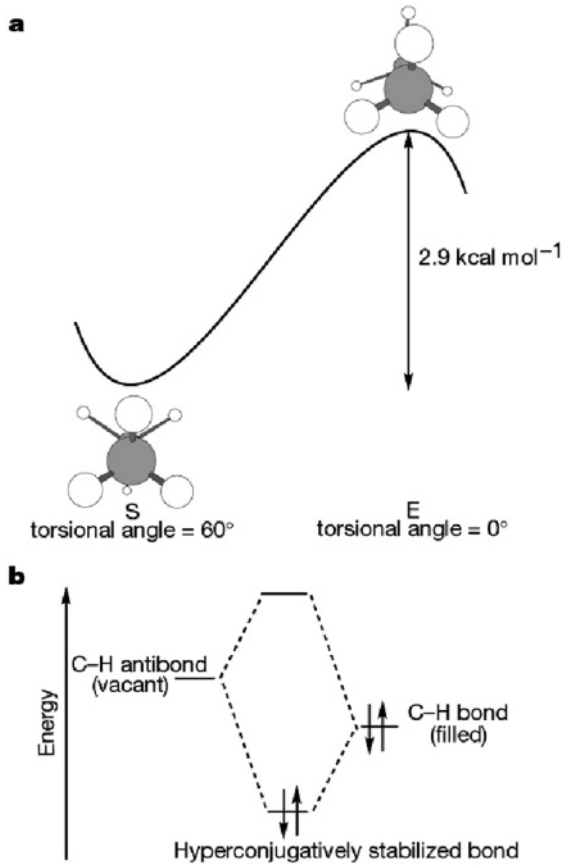


Figure 6. Energetics in ethane. (a) Ethane staggered (S) and eclipsed (E) conformers. (b) Vicinal hyperconjugative stabilization by overlap between an occupied and an unoccupied orbital on two methyl groups. (Reprinted by permission from Macmillan Publishers Ltd: *Nature* (http://www.nature.com), ref 58. Copyright 2001.)

eclipsed conformation is the unfavored high-energy state while the staggered conformation is the preferred low-energy state.[58,59] Counterintuitively, according to hyperconjugation, the staggered conformation is attributed to favorable $\sigma_{CH}-\sigma_{CH}$* orbital interactions (Figure 6b) rather than simple steric repulsion between hard-sphere-like atoms.[58,59]

It is known from quantum mechanics that each constructive (in-phase or bonding) orbital is accompanied by a corresponding destructive (out-of-phase or antibonding) orbital.[59] In ethane, each methyl group has three bonding σ_{CH} orbitals, accompanied by three antibonding σ_{CH}* orbitals. Figure 7 shows that a more favorable orbital overlap occurs in the staggered conformation and a less favorable orbital overlap occurs in the eclipsed conformation. This results in a stronger stabilizing interaction for the staggered conformation (the favored low-energy state) and a weaker interaction for the eclipsed conformation (high-energy state).

How might hyperconjugation be related to P_{II}? We start with a brief introduction of a study on collagen peptides carried out several years ago. Raines and his colleagues incorporated 4(*R*)-fluoro-L-proline (Flp) into a collagen model system and found that $(ProFlpGly)_{10}$ is dramatically more stable than either $(ProProGly)_{10}$ or $(ProHypGly)_{10}$ (Hyp, 4(*R*)-hydroxy-L-proline)[60] (see Figure 8). It has been known for a

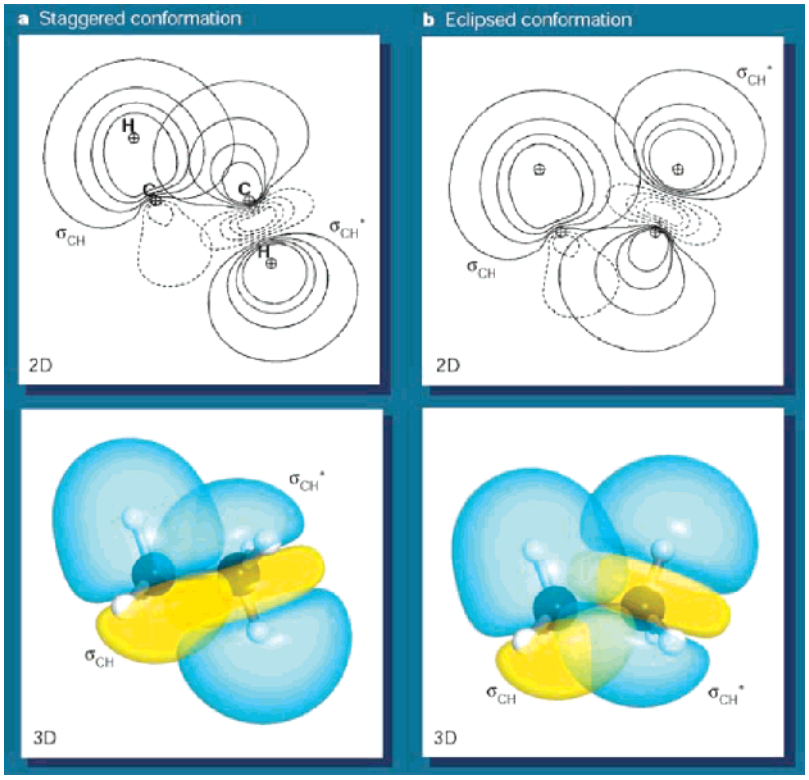


Figure 7. Contour plots (two-dimensional, 2D) and surface plots (3D) of adjacent carbon−hydrogen bond orbitals (σ_{CH} and σ_{CH}*) of ethane. (a) In the staggered conformation the more favorable orbital overlap leads to a stronger (more stabilizing) interaction. (b) There is a less favorable interaction in the eclipsed conformation, leading to a higher-energy state. (Reprinted by permission from Macmillan Publishers Ltd: *Nature* (http://www.nature.com), ref 59. Copyright 2001.)

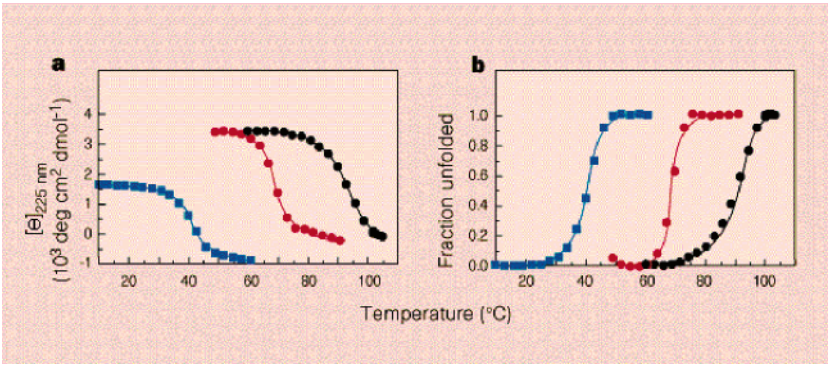


Figure 8. Thermal denaturation of collagen-related triple helices: (a) raw data; (b) transformed data. The values of T_m (°C), which is the temperature at the midpoint of the thermal transition curve, are as follows: $(ProProGly)_{10}$ (blue), 41 ± 1; $(ProHypGly)_{10}$ (red), 69 ± 1; $(ProFlpGly)_{10}$ (black), 91 ± 1. (Reprinted by permission from Macmillan Publishers Ltd: *Nature* (http://www.nature.com), ref 60. Copyright 1998.)

long time that $(ProHypGly)_{10}$ is more stable than $(ProProGly)_{10}$, an effect that had been attributed to stabilization of a network of hydrogen bonds mediated by bridging water molecules in $(ProHypGly)_{10}$.[135] The fluorine in Flp is more electronegative than the hydroxyl group in Hyp, and both are more electronegative than the corresponding Pro residue. However, the fluorine in Flp does not form hydrogen bonds. Raines and colleagues[60,136] argued that it must be a previously unappreciated inductive effect of the electron-withdrawing group rather than water bridges that enhances the stability of $(ProFlpGly)_{10}$ and $(ProHypGly)_{10}$ relative to $(ProProGly)_{10}$. The thermal stability increases in the order $(ProProGly)_{10}$ < $(ProHypGly)_{10}$ < $(ProFlpGly)_{10}$ with Flp having the strongest, Hyp the next strongest, and Pro the weakest inductive effect. Raines and his colleagues reported[137] that the pK_a's (reflecting inductive effects) of the nitrogen atom in the corresponding parent amino acids decrease in the order Pro (10.8) > Hyp (9.68) > Flp (9.23), which correlates with the prolyl peptide bond isomerization equilibrium constants and the equilibrium concentrations of the *trans* isomers: Pro derivative < Hyp derivative < Flp derivative. At the time, they hypothesized that the stability of the collagen triple helix is enhanced through inductive effects[137] that favor the requisite *trans* conformation of the peptide bond.

However, if the inductive effect accounts for the increase in collagen stability by favoring the *trans* conformation of the peptide bond, then an equal stabilization of collagen should result if the Hyp is located at either the Y or X positions (X and Y here refer to the repetitive collagen peptide sequence X-Y-Gly), which contradicts the experimental observation that Hyp at a Y position stabilizes, whereas Hyp at an X position destabilizes, the triple helix. Zagari and co-workers have recently determined a high-resolution (1.3 Å) collagen-type structure with the (ProPro-

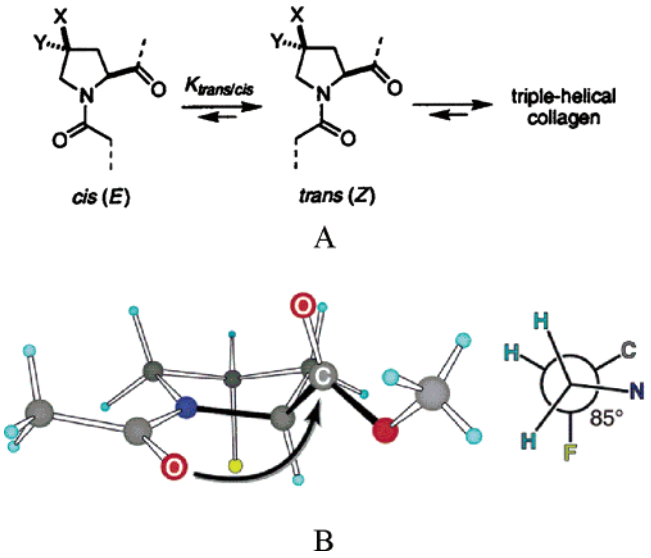


Figure 9. (A) Relationship between *cis*−*trans* prolyl peptide bond isomerization and the formation of a collagen triple helix, which contains only *trans* peptide bonds. Pro: X = H, Y = H. Hyp: X = H, Y = OH. hyp: X = OH, Y = H. Flp: X = H, Y = F. flp: X = F, Y = H. (B) (Left) Structure of crystalline AcFlpOMe. $r_{O_0 \cdots C_1}$ = 2.76 Å; $\angle O_0 \cdots C_1{=}O_1 = 98°$. The ψ dihedral angle (solid black bonds) is 141°. (Right) Newman projection depicting the *gauche* effect. The $N_1{-}C_1^{\delta}\text{-}C_1^{\gamma}\text{-}F_1^{\delta_1}$ dihedral angle is 85°. As pointed out by a reviewer, the reader should note that substitutions of different functional groups as denoted by X and Y in the chemical structures in this figure have nothing to do with the X and Y positions discussed in the text throughout section 5.3, which refer to the general repetitive sequence of collagen peptides, X-Y-Gly. (Reprinted with permission from ref 141. Copyright 2001 American Chemical Society.)

$(Gly)_{10}$ sequence.[138] This high-resolution structure reveals that Pro residues located at different positions (X or Y) have clearly distinctive torsion angles and ring conformations. Pro residues at X exhibit an average $\phi = -75°$ (±3°) and adopt a puckered down conformation whereas those at Y have an average $\phi = -60°$ (±2°) with the puckered up conformation. Polyproline peptides[139] tend to adopt an average (up and down) ring conformation, whereas polyhydroxyproline peptides[140] prefer the puckered up conformation. This suggests that Hyp residues at Y positions preorganize the main chain torsion angle to the requisite puckered conformation and stabilize the collagen triple helix, whereas Hyp residues at X positions prevent the main chain from forming the desired conformation and therefore destabilize collagen triple helices.

Following this line of reasoning, Raines and his colleagues reconsidered their hypothesis and invoke hyperconjugation as an explanation both for the *gauche* effect and $n \rightarrow \pi^*$ interactions.[62,141,142] The *gauche* effect determines the pyrrolidine ring pucker, which in turn preorganizes the main chain torsion angles ϕ and ψ; furthermore, the $n \rightarrow \pi^*$ interaction stabilizes not only the ideal ψ angle but also the requisite *trans* conformation ($\omega = 180°$) of the peptide bond. The resulting effect is then an interplay of the pyrrolidine ring pucker and the ϕ/ψ torsion angles and peptide *trans*/*cis* isomerization equilibrium of substituted proline residues (see Figure 9).

Recently, Raines and his group have analyzed the $n \rightarrow \pi^*$ interaction as a function of different regions in the Ramachandran plot and find that optimal $n \rightarrow \pi^*$ interactions can exist in P_{II}, α_R, and α_L but not β (Figure 10).[61] Interestingly, they find that, for P_{II} helices, there is not only an $n \rightarrow \pi^*$ interaction from $O_{i-1}(n)$ to $C_i' = O_i(\pi^*)$ but also one from $O_{i-1}(n)$ to $C_i'{=}N_{i+1}^{+}(\pi^*)$. Since the resonance structure $C_i'(O_i^{-}){=}N_{i+1}^{+}$ can account for as much as 40% of the population of an amide, $n \rightarrow \pi^*$ interactions might indeed play a role in P_{II} structure. They estimate a stabilization energy of about 0.7 kcal/mol ($=RT \ln 3$) by comparing the *trans*/*cis* equilibrium constants of the following *cis* ↔ *trans* isomerization reactions (Scheme 1 for the formamide derivative and Scheme 2 for the acetamide derivative).

In summary, hyperconjugation offers a novel explanation for why the backbone conformation in unfolded states of proteins is biased to P_{II}.[61] Hyperconjugation offers one more explanation to add to previous ones, namely side chain−backbone interactions, solvent hydration of the backbone, sterics, and dipolar interactions, for the prevalence of P_{II}. Quantitative tests of the relative magnitude of these effects involve substitutions in natural side chains, as well as variation of solvents.[143] In particular, determining how hyperconjugation varies with dielectric constant might help

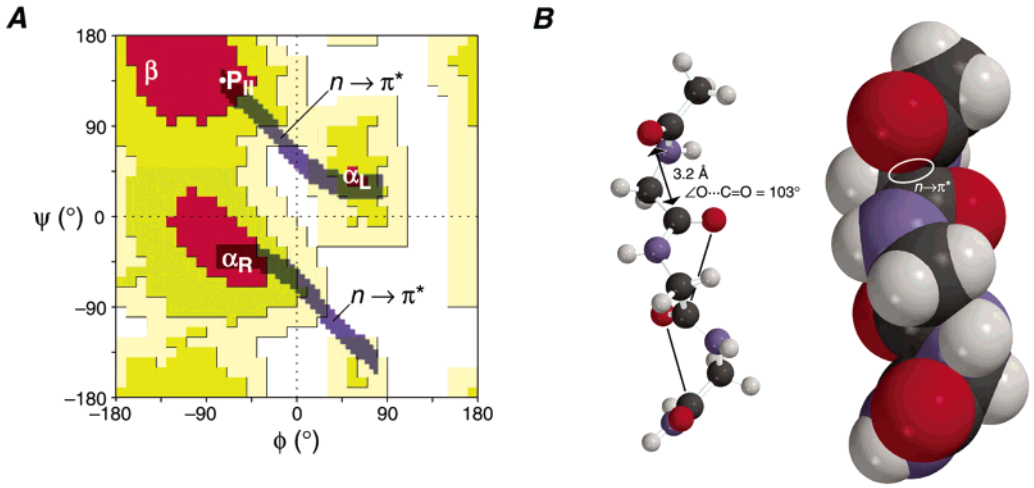
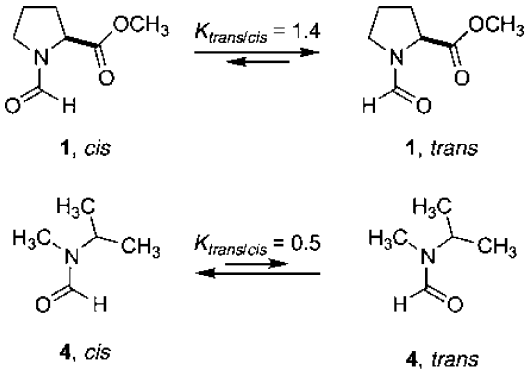
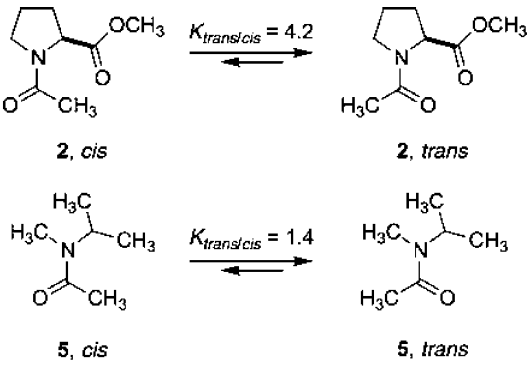


Figure 10. Some implications of hyperconjugation involving $n \rightarrow \pi^*$ interactions between O_{i-1} and C'_i. (A) Ramachandran plot showing the two "$n \rightarrow \pi^*$" regions of the *trans* isomer of $AcGlyNH_2$. In these regions, the $O_{i-1} \cdots C'_i$ distance is $\delta_{BD} \leq 3.2$ Å and the $\angle O_{i-1} \cdots C'_i{=}O_i$ angle is $99° \leq \tau_{BD} \leq 119°$. The white dot indicates the ϕ and ψ angles for an ideal polyproline II helix (B). (B) Energy-minimized structure of $AcGly_3NH_2$ in the conformation of a polyproline II helix with $\phi = -75°$ and $\psi = +145°$. The structure is depicted as a ball-and-stick (left) or space-filling (right) model. The $O_{i-1} \cdots C'_i$ distance ($\delta_{BD} = 3.2$ Å) and $\angle O_{i-1} \cdots C'_i{=}O_i$ angle ($\tau_{BD} = 103°$) are indicated in the ball-and-stick model. (Reprinted with permission from ref 61. Copyright 2003 Cold Spring Harbor Press.)

Scheme 1. The $n \rightarrow \pi^*$ Interaction Determined from a Formamide Derivative as a Ratio of the above Two Equilibrium Constants (~3) (Reprinted with Permission from Ref 61. Copyright 2003 Cold Spring Harbor Press.)



Scheme 2. The $n \rightarrow \pi^*$ Interaction Determined from an Acetamide Derivative as a Ratio of the above Two Equilibrium Constants (~3) (Reprinted with Permission from Ref 61. Copyright 2003 Cold Spring Harbor Press.)



explain the strong solvent dependence observed for P_{II} structure in short model peptides. It should be mentioned that designing any unequivocal test of the hyperconjuation model is not easy.

6. Significance of P_{II} as a Major Backbone Structure in Unfolded Proteins

6.1. Implication for Protein Folding

The presence of defined localized structure in the backbone conformation of unfolded proteins has far-reaching implications for the protein folding problem. It suggests that the structure of unfolded proteins is less heterogeneous than has been thought previously and that the backbone entropy in unfolded states is much lower than that implied by the random coil model, even for the cases where there is only a weak bias to a certain structure in the unfolded ensemble of protein conformations as emphasized by Hilser.[47,48] As one consequence, Levinthal's paradox[144] with respect to the peptide backbone can be resolved by accepting the idea that unfolded proteins have well-defined local structures. The immobilization of side chains upon folding still contributes significantly, of course. Nevertheless, concepts[19] such as folding funnels, kinetic traps, and frustration may prove to be misleading.[11]

Tanford[1] relied on hydrodynamic data to demonstrate that the dimensions of most chemically denatured proteins scale with the numbers of residues in the chain according to a power-law relationship predicted from the random coil model. Today, experimental results from small-angle X-ray scattering (SAXS)[27] experiments have improved these measurements enormously and confirm that it is rare for any denatured protein to deviate from a power-law scaling relationship connecting R_g and chain length. Fitzkee and Rose have recently demonstrated that consistency with this power law behavior is remarkably insensitive to the presence of well-defined local structure(s) in an otherwise flexible chain.[11] They generated ensembles of segmentally rigid chains from proteins with known structure from the PDB by varying the backbone torsion angles at random for ~8% of residues, maintaining the remaining ~92% of residues with their native torsion angles. Strikingly, they find that ensembles generated by imposing these bizarre constraints have characteristics (including end-to-end distance, mean radii of gyration, and simulated Kratky plots) consistent with random coil expectations in 30 cases among a set of 33 proteins.

Fitzkee and Rose cite one SAXS study that provides strong support for their simulation results. In this study the denatured state of hen lysozyme in 40% trifluoroethanol (TFE) is found to have an R_g value comparable to that of the same protein denatured in 4 M GuHCl.[145] Another study on cytochrome *c* finds that the denatured state in methanol has an R_g of 31.7 Å, which is very close to that for the acid denatured state (30.1 Å) as well as that for the urea denatured state (32.1 Å).[146] TFE and methanol both stabilize helical structure,[147] as revealed by their effect on the CD spectra of these two proteins.[145,146] Thus, the observation that R_g values conform to the mean dimension of a random coil ensemble does not provide evidence for the random coil models or exclude well-defined local secondary structures in unfolded proteins. This point has also been made by Sosnick and colleagues on the basis of their recent SAXS study.[148]

Fitzkee and Rose[11] suggest that the random coil might indeed be better considered as a conceptual obstacle that has impeded alternative explanations than a useful concept. With mounting evidence demonstrating the failure of the random coil model, locally determined residue specific backbone conformation preferences within a primary sequence might be expected to influence early and subsequent events in folding. Rose's LINUS program among others postulates that folding is locally determined and hierarchical. Thus, one may approach the protein folding problem by trying to answer the following question: how is the conformation of a residue determined by its own chemical nature and its local environment, that is, the conformations of its nearest neighboring residues?

Given that there is a distinct structural propensity for each amino acid together with strong local context effects, deciphering the intrinsic structural propensity for each amino acid together with a set of rules governing local context effects might allow one to improve the prediction of structures of proteins, folded or unfolded, as well as provide insight into folding pathways.

6.2. Impact on Force Field Development and Refinement

6.2.1. Reliability of Force Fields

Molecular dynamic simulations provide an essential tool in efforts to understand the high-resolution structure and dynamics of complex biological macromolecules.[149] The evidence showing that models such as AAMA and trialanine

have well-defined P_{II} structure[4−7,72] invites re-examination of some force fields in current use.[67,69,86]

At present, the most widely used molecular mechanics force fields include AMBER,[150] CHARMM,[151] GROMOS,[152] and OPLS.[153] While these force fields share a common form of empirical potential energy function, they differ in their values of a large set of associated parameters. Accurate quantum mechanical calculations can only be carried out for small systems with a limited amount of atoms presently. As a result, empirical force fields are calibrated by fitting results on small building blocks of biopolymers through *ab initio* and density functional theory (DFT) calculations. Even for short peptide fragments, it is impractical to conduct a high level quantum mechanical calculation in the presence of water. As a compromise, most force fields are mainly developed through fitting to the results from gas-phase calculations.

One essential test of a force field is to compare simulated results with observable properties of a number of well-characterized small biomolecular benchmarks including AAMA and trialanine. A potential energy distribution diagram, mapping the energy as a function of the ϕ and ψ dihedral angles in the corresponding short peptides, is widely used as a criterion. Almost all force fields identify a similar geometry (C_{7eq}) as the lowest energy structure *in vacuo*. However, in the presence of water, each force field yields a significantly different distribution of structures. Only some[154−157] reproduce the featured P_{II} structure, although there is confusion in the literature concerning the definition of β conformation and a number of publications include P_{II} within the broader β structure basin.

Given this situation, several groups have undertaken a detailed analysis of current widely used force fields.[67,69,85,86] Recently, three groups have independently compared the performance of several of these force fields.[67,69,86] Common findings in these studies include the following:

(i) Accuracy is a major challenge for the development and refinement of empirical force fields. Currently, the uncertainty is still well above $RT = 0.6$ kcal/mol, a number that is thought to be required to reproduce adequately experimental observations on both small peptides and larger polypeptides or proteins.

(ii) Water models and the way they treat polarization effects make a significant difference in calculating population probabilities of different basins.

(iii) Prediction results from AMBER 94 and CHARMM tend to oversample α_R state; AMBER 96 and GROMOS predict equal populations of β (~40%) and P_{II} (~40%) and a small population of α_R (~20%); thus, both oversample β region and underestimate P_{II} as compared to experimental results; OPLS predicts a majority of extended conformation (>80%), without distinguishing P_{II} from β basin in the Ramachandran plot.

6.2.2. Refining Force Fields

Seemingly minor changes in parameters can exert significant effects in terms of the population probabilities for different basins, as demonstrated by the different results from two versions of AMBER. On the positive side, this means that force fields can be fine-tuned so they can duplicate experimental observations on both small peptides and longer polypeptides and proteins. One demonstration is by Garcia and Sanbonmatsu,[158] who find that a modification of the AMBER force field replicates experimental results for both short and long peptides quite well, simply by flattening the torsion angle H-bond potential in AMBER 96.[69,158] Duan *et al*. have recently reported development of a third generation of AMBER relying on an increased level of QM calculations.[159] They predict a favorable probability for P_{II} that is in better agreement with experimental data. Recently, MacKerell and colleagues[160] have refined the treatment of protein backbones in CHARMM22 by introducing a grid-based correction to the full ϕ/ψ two-dimensional conformational energy surface. Their refined model also shows improved consistency with recent experimental results.

Improved water models together with adjustment of parameters of existing force fields can and should improve the performance of predictions significantly, as demonstrated in a recent report from Mu *et al*.[86] Using a new version of the OPLS force field combined with a TIP5 water model on trialanine, they predict population probabilities of 65% for P_{II}, 12% for β, and 18% for α conformation, respectively,[86] much closer to what is observed experimentally. With the availability of more powerful computational resources, another key may lie in carrying out additional quantum mechanical calculations. For example, Hu *et al*. have recently employed a fast combined QM/MM force field to simulate Ala and Gly dipeptides.[67] They report that the combined QM/MM force field outperforms any existing MM force fields. Significantly, basin distribution results for both Ala and Gly dipeptides by the QM/MM simulation show closer agreement with recently reported distributions in high-resolution protein structures[161] than those from any other simulations using MM force fields. Their Ala result shows little P_{II}, however.

Theoretical analysis of the helix−"coil" transition in simple peptides has been based on different sampling procedures, force fields, and solvent treatments. The results vary a great deal, and the extent to which the unfolded conformation includes P_{II} ranges from negligible (Ohkubo and Brooks[162]) to substantial (Garcia[66] and Durani[163]). Simulations by van Gunsteren's group[152] revealed the unexpected fact that the conformational manifold of short unfolded peptides must be restricted, since they could see transitions in both directions between helices and "coil". This observation is in complete accord with the presence of structure in the unfolded state. But how the effect varies with chain length is puzzling: evidently, as the chain length increases, one expects effects to scale with length.

6.3. P_{II} Structure in Natively Unfolded Proteins and Fibril Formation

Barron's group has recently investigated a number of amyloid forming proteins including human lysozyme,[164] α-synuclein, and tau proteins[49] as well as several natively unfolded proteins[49,50] using ROA. In all cases, they find a significant amount of P_{II} structure suggesting a connection between P_{II} and the fibril formation mechanism. A similar analysis based on ultraviolet resonance Raman (UVRR) spectroscopy finds P_{II} in assembled tau protein.[165] The structure of α-synuclein also has P_{II} structure, based on a Raman study.[166] This research has broad implications for structures of natively unfolded proteins as well, which may account for about one-third of all proteins.[167−173] The prion protein PrP is found to have P_{II} structure at its N terminus,[174] and two groups have reported that unassembled Aβ peptides have P_{II} conformation.[175,176] The glycoprotein human salivary mucin contains a tandem repeat that is bacteriocidally active and also has P_{II} conformation.[177] As we have mentioned,

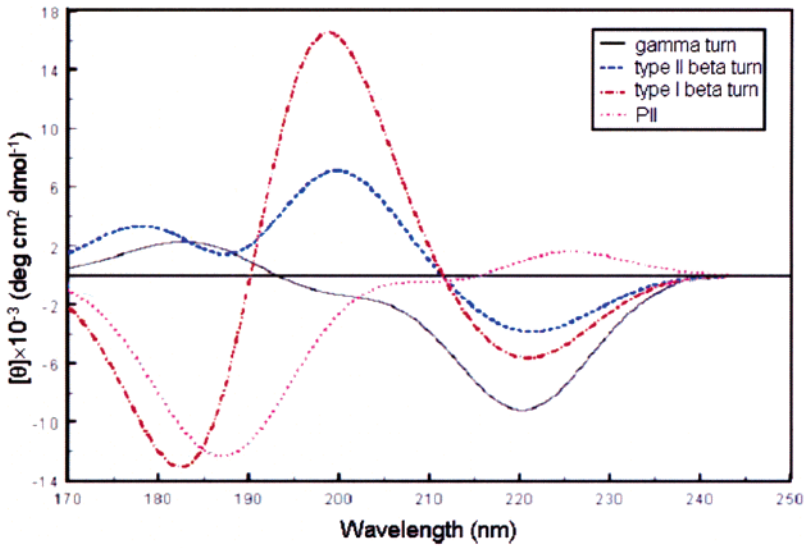


Figure 11. Calculated CD spectra of $AcGGAGGNH_2$ in γ and β turn conformations: P_{II} (ϕ, ψ) = (−60°, +170°); γ turn canonical (ϕ, ψ) = (−78°, +65°); type I β turn canonical (ϕ, ψ) = (−60°, −30°); the type II β turn canonical (ϕ, ψ) used in this paper are from Venkatachalam's conformations 8 and 10,[183] in which the first C_α of the turn has (ϕ, ψ) = (−60, +90°) vs (−60, +120°). (Reprinted with permission from ref 97. Copyright 2004 American Chemical Society.)

Tamburro and his colleagues find that the Exon 5 peptide dissected from human tropoelastin has P_{II} structure and, in addition, is capable of self-assembly into fibrils, as revealed by transmission electron microscopy measurements.[103] Recently, Armen *et al.*[178] offered the provocative suggestion that intermediate states in amyloid formation have a unique α pleated sheet structure, originally proposed by Pauling and Corey, in which alternating residues have dihedral angles corresponding to α_R and α_L. In simulations of transitions between helix and β strand, they regularly detect a high occupancy of both P_{II} and α_L, despite the absence of the latter structure in coil libraries and the additional problem that a significant presence of α_L should lead to cancellation of the CD and ROA signals from such intermediates. The possibility that α_L conformations are involved in transition states between α and β basins is interesting, since α_L is also stabilized by hydration in AAMA.[57]

6.4. CD Assignments of P_{II} and Other Nonstandard Conformations

Available evidence summarized above indicates that there is a substantial amount of P_{II} conformation in unfolded proteins and peptides. These lines of evidence provide strong support for assigning the reference CD spectrum of P_{II}:[12,17,23] a very weak negative band near 235 nm, a weak positive band near 220 nm, and a very strong negative band near 195 nm (for non-proline and nonaromatic residues).

The literature to date still assigns this spectrum as "random coil". The CD spectrum of "random coils" was calculated by Ronish and Krimm more than 30 years ago.[24,25] They showed that the CD of a random polypeptide chain including α and β structure should have a positive band around 196 nm and a negative band centered around 215 nm with a crossover point at about 205 nm. While the results are uncertain because of the unknown compositional coefficients and chain length dependence for each conformation in the blend, the spectrum they predicted is clearly unrelated to that of any unfolded peptides and proteins. At this stage, it is our opinion that it is safer to say in many cases the CD of unfolded proteins corresponds more closely to that of P_{II} rather than any blend of conformations.

We have noticed that P_{II} CD spectra for different types of amino acids are slightly different in terms of band location and magnitude because each amino acid probably samples slightly different basins within the P_{II} region, as well as distinctive local minima in the broader β basin that flanks P_{II}. Such preferences are likely to be sensitive also to local sequence context, different end charge states, and solution conditions. Given the important role of P_{II} structure in unfolded proteins, it is necessary to systematically and carefully investigate these effects since they will make estimation of P_{II} contents from experimental CD data uncertain though they may not affect the estimation of α and β contents in native proteins as much.

Assignments of CD spectra of other nonstandard conformations are becoming an urgent, though challenging, task since other nonstandard conformations presumably also play pivotal roles in the structure of unfolded proteins. Recent efforts by Woody have addressed this problem. Woody[179,180] has recently calculated the CD spectra of γ turn, type I β turn, and type II β turn as well as P_{II} conformations, and presented the results in a recent study[97] (see Figure 11). It is worth noting that CD spectra can be evaluated through *g*-factor analysis[181,182] in which the CD and UV spectra are converted to dimensionless *g*-factor spectra by dividing the CD by the UV signal at each wavelength. This has the advantage that data can be analyzed without information on the protein molecular weight, sample concentration, or sample path length. It holds some promise for secondary structure deconvolution including nonstandard conformations, which may occur in insoluble proteins such as prions or amyloid.

7. Comparison of Different Models for the Structure in Unfolded Proteins

In this review and a previous one, we have attempted to summarize the evidence that unfolded peptides and proteins have a strong tendency to be P_{II} locally while still conforming to the overall dimensions of a statistical coil. We refer to this as the P_{II} model, originally proposed by Tiffany and Krimm[12,26] and supported by the extensive evidence discussed above. Several alternative models are in vogue for unfolded

proteins. One that we refer to here as the native topology model has been proposed by Shortle and his colleagues.[184] It is based on a series of results from studies of unfolded SNase, including a truncated version of the protein, Δ131Δ, which is unfolded under native conditions as well as the wild-type SNase in a variety of denaturing conditions. A second model we refer to as the long-range structured random coil is advocated by Dobson, Schwalbe, and their colleagues.[185,186] Based on a mutational analysis of reduced hen lysozyme in denaturing conditions, they hypothesize that unfolded lysozyme is an ensemble of random coils coexisting with some long-range structures, most likely nativelike, that are stabilized by extensive clusters of hydrophobic interactions. The third is the diffusive statistical random coil model derived from Tanford's pioneering work,[1] still the most widely accepted model today. This forms the basis for interpreting many current experimental results, for example as in the publication we cited at the beginning of this review by Eaton and his colleagues.[14]

The major difference between the P_{II} model and that of Shortle and his colleagues[184] is that they believe unfolded proteins are random coils that retain a nativelike topology. While unfolded proteins may indeed have some residual long-range structure, native or non-native, in our view this is imposed on a strong background of P_{II} backbone conformation.[23] The P_{II} conformation makes the peptide backbone locally ordered, although it must still be interspersed by many types of turns or loops and other local structures. Thus, unfolded proteins are locally ordered, yet coil-like in structure, flexible in dynamics, as well as heterogeneous and disordered in terms of long-range space occupancy.

The validity of Shortle's model depends on how one defines nativelike topology. They have reached the conclusion by characterizing the unfolded SNase using an impressive variety of spectroscopic probes including CD, ^3J coupling constants, NOE, NMR chemical shifts, and residual dipolar coupling constants (RDCs) as well as paramagnetic relaxation enhancement.[184,187−191] Very recently, they reported some surprising results from measurements of residual dipole coupling constants in several variants of SNase including the protein with 30 N-terminal residues deleted, a mutant with the deletion of the final 30 C-terminal residues, and a mutant with all 10 hydrophobic residues replaced by polar substitutions.[192,193] Under denaturing conditions, they observe a robust correlation in residual dipolar coupling constants among all these mutants as well as with those of the unfolded wild-type enzyme.[192] They also report a robust correlation in residual dipolar coupling constants for corresponding residues of the protein denatured in different experimental conditions, regardless of whether the protein unfolds by deletion in the sequence without denaturants, by different concentrations of urea, or by decreasing pH, which introduces approximately 20 additional positive charges into the molecule.[193]

Both urea and GuHCl have been found to favor P_{II} in proline-containing as well as non-proline peptides.[97,98] An increase in extended P_{II} structure can explain the observation that denatured proteins expand slightly at higher concentrations of denaturants. One recent set of measurements suggests that changes in R_g as a function of [GuHCl] are minor.[27] At higher concentrations of urea or GuHCl, denatured proteins might populate longer segments of P_{II} that occur in natively unfolded proteins or unstable mutants at lower denaturant concentrations. Thus, the robustness of the correlation among

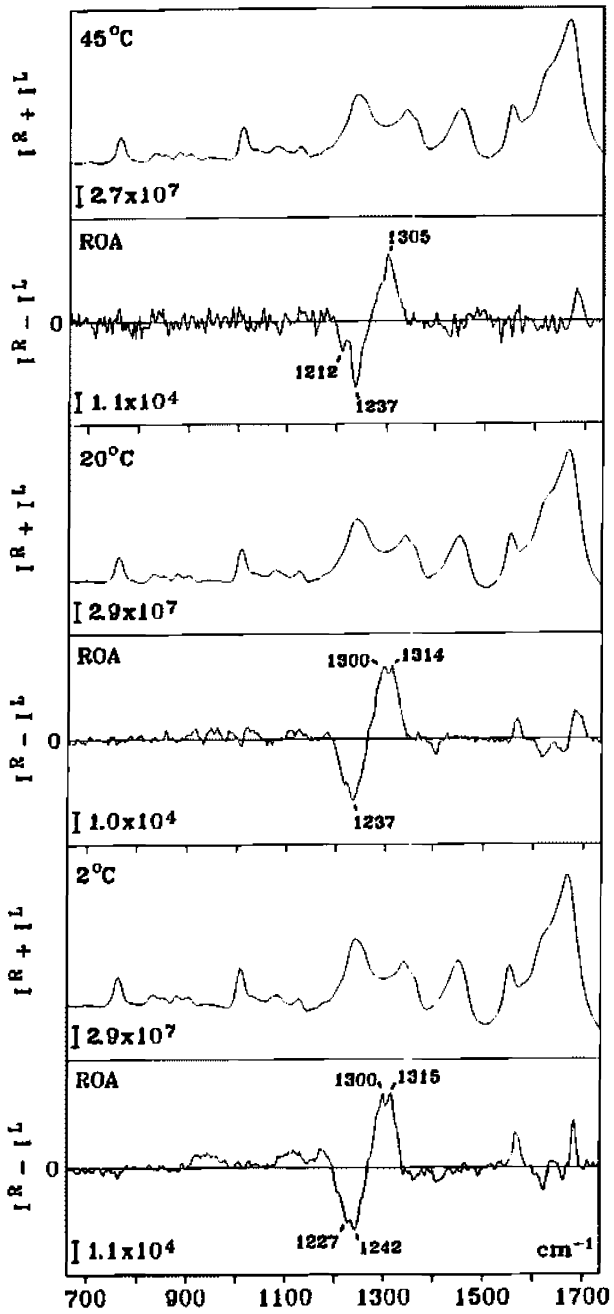


Figure 12. Backscattered Raman and ROA spectra of reduced hen egg white lysozyme in citrate buffer, pH 2.0, at 45 °C (top pair), 20 °C (middle pair), and 2 °C (bottom pair). (Reprinted with permission from ref 18. Copyright 1996 American Chemical Society.)

different unfolded protein molecules with such dramatic modifications in sequence may reflect the persistence of local P_{II} structure rather than the persistence of nativelike topology, which can also explain why no correlation is found between native and unfolded states[184] while the correlation among all denatured states is strong.[192]

Dobson, Schwalbe, and their colleagues based their model for denatured proteins on extensive NMR data including NOE, ^3J coupling constants, and ^{15}N relaxation measurements on hen lysozyme and a variety of mutants denatured by urea or by reducing and methylating all disulfide bonds.[185,186] Their observations indicate that there are extensive clusters involving distinct regions of the sequence which can be disrupted, for example, by a single point mutation that replaces Trp62 with Gly. They fit their ^{15}N relaxation rates

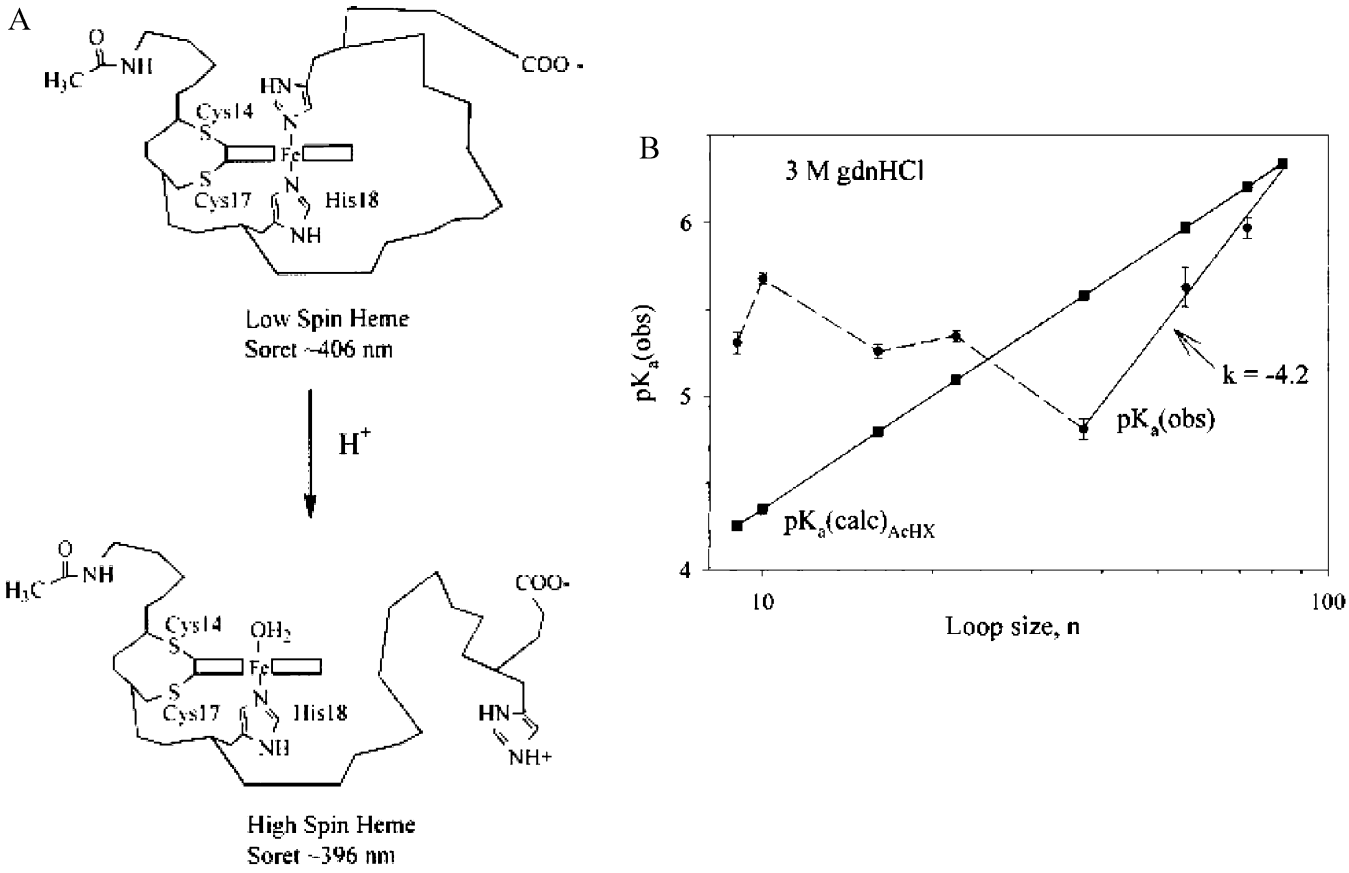


Figure 13. (A) Representation of the pH-dependent equilibrium for histidine/heme loop formation under denaturing conditions with an acetylated N terminus. (B) Comparison of pK_a(obs) to pK_a(calc)$_{AcHX}$ for the AcHX variants. The pK_a(obs) (filled circles) and pK_a(calc)$_{AcHX}$ (filled squares) values are plotted against loop size. The loop size axis is given on a log scale. Error bars are shown for all pK_a(obs) data points. A continuous line connects the pK_a(calc)$_{AcHX}$ values. The continuous line connecting the last four pK_a(obs) data points is a linear least-squares fit to the data. The other pK_a(obs) data points are simply connected with broken lines. (Reprinted with permission from ref 30. Copyright 2001 Elsevier.)

for all measurable backbone residues with a random coil model that undergoes segmental motion and derive an intrinsic relaxation rate for random coils with the number 7 (in residues) for the persistence length of the chain for different denatured proteins. It can be argued that this value of the persistence length is not inconsistent with a locally ordered yet long-range disordered P_{II} model of unfolded proteins.[23] Furthermore, the ROA spectra for reduced hen lysozyme (Figure 12) reported by Barron and his colleagues[18] several years ago show clear indications of P_{II} structure with additional helix and turns based on their revised interpretation of the 1320 cm^{-1} band.[37] These observations are more consistent with the P_{II} model than the model proposed by Dobson's group.

We have discussed the fact that a random coil is not inconsistent with the presence of local order in unfolded proteins. We examine here some recent studies by Eaton and his colleagues, who have experimentally investigated the free energy surface for protein folding using single molecule fluorescence spectroscopy.[14] They report that there is no excess broadening in the E_{app} distribution for an unfolded protein (hyperthermophilic CspTm) beyond that observed in a reference oligomer $(Pro)_{20}$; E_{app} denotes the mean and width of the measured FRET efficiency. They then assume a Gaussian chain model and derive a reconfiguration time of less than 25 μs to explain the result. However, if unfolded states of proteins have some kind of well-defined structure, there is no need to construct *ad hoc* models to explain the experimental observation: the fact that the E_{app} distribution for the unfolded protein matches that observed for an extended P_{II} molecule such as $(Pro)_{20}$ may not be surprising.

There are two additional minor but nevertheless relevant problems in these experiments.[14] In Figure 2d of their paper, Eaton and his colleagues report an E_{app} value slightly above 50% for $(Pro)_{20}$. In theory, a value of 50% for E_{app} corresponds to a P_{II} helix length of 5.4 nm; the expected length for $(Pro)_{20}$ is 6.2 nm. The E_{app} for $(Pro)_{20}$ should thus be slightly less than 50%. They reconcile the discrepancy by arguing that long flexible linkers attached to the dyes "*allow the dyes to approach each other during the fluorescence lifetime of the donor*". However, the long flexible linkers of the dyes should also allow them to escape from each other during the fluorescence lifetime of the donor. We have noticed that had they not attributed the additional maximum at $E_{app} \approx 0$ as due to chemically altered or donor labeled chains, the distribution in $(Pro)_{20}$ would shift to a value slightly below 50%, which is more consistent with what would be expected. Another observation that is hard to explain is that, in Figure 2c of their paper, there is a negligible distribution at $E_{app} \approx 0$ for $(Pro)_6$ whereas the corresponding distribution at $E_{app} \approx 0$ for $(Pro)_{20}$ is comparable to that at $E_{app} \approx 0.5$ although one finds no significant difference between parts a and b of Figure 2 of the paper. Apparently, chemically altered or donor labeled chains are present only in $(Pro)_{20}$. In contrast to providing support for the diffusive statistical random coil model as they claim, it can be argued that their experimental results actually support an opposing P_{II} model.

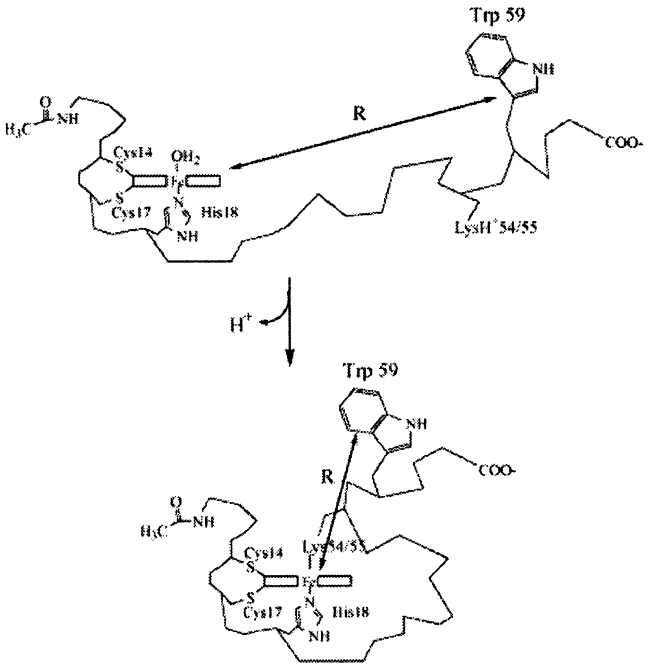


Figure 14. Schematic of the effect of binding Lys 54/55 to the heme of denatured AcTM on the Trp 59−heme distance. The double-headed arrows show the Trp 59−heme distance, *R*, for the low pH water bound state of the heme and for the high pH Lys 54/55 bound state of the heme. (Reprinted with permission from ref 31. Copyright 2003 American Chemical Society.)

A series of loop closure experiments on unfolded iso-1-cytochrome *c* by Bowler and his colleagues also addresses this issue.[28−31] They introduced a series of His substitutions in cytochrome *c* and determined the histidine/heme loop-forming equilibrium constant by means of His pK_a shifts (see Figure 13A).[30] Their results deviate significantly from predictions for any random coil model of the denatured state. Specifically, for different variants of mutants corresponding to different loop sizes, they report the pK_a values shown as filled circles in Figure 13B.[30] Assuming a random coil model, they calculate the pK_a values shown as filled squares in the figure. It is clear that the random coil model does not apply to cases with loop sizes of 9, 10, 16, 22, and 37 residues. For cases with larger loop sizes of 37, 56, 72, and 83, the observed pK_a values fit a straight line, but the slope −4.2 is inconsistent with the random coil model, which should show the expected Stockmayer−Jacobson value near −1.5.

In a more recent report, Bowler and his colleagues[31] have studied the denatured state of iso-1-cytochrome *c* by monitoring changes in Trp59−heme fluorescence quenching (Figure 14). They could derive a heme−Trp59 distance in a denatured mutant of iso-1-cytochrome *c* (AcH54I52) from Förster energy transfer theory and find that the derived value of ∼26 Å at pH 4 and 10, while His54 is not bound to the heme, is much shorter than the 56 Å that would be predicted from the random coil model. Surprisingly, when His54 binds to the heme in the denatured state at neutral pH, the heme−Trp59 distance decreases only slightly, to ∼20.7 Å. In this case, the number of residues separating the heme and Trp59 is 6. The measured distance suggests that all residues between His54 and Trp59 adopt some kind of extended conformation. The distance they measure agrees very well with that calculated by assuming the stretch of polypeptide chain in the denatured states occupies a P_{II} structure (6 × 3.2 Å = 19.2 Å as compared to ∼20.7 Å).

A direct measure of the chain dimension in the peptide containing seven alanines (XAO) reports that the radius of gyration R_g in this molecule is 7.4 ± 0.5 Å, about half of what is predicted for a fully extended P_{II} helix.[194] While this result is claimed to be a discrepancy with the analysis of the same peptide by Shi *et al*.,[8] the original article emphasized that the probability of full P_{II} occupancy of all alanines in the peptide is low, since the formation of P_{II} is noncooperative, as determined in the series $AcGGA_nGGNH_2$. The R_g value measured does not reflect an experimental discrepancy if we assume that a P_{II} helix of XAO bends at two positions (three segmented P_{II} helices); such a structure could have a radius of gyration within the value they have measured. A different problem raised by Zagrovic *et al*. is that tests of simulations with six different force fields uniformly failed to find P_{II} conformation in the peptide.[194] Instead, the simulations tend to predict appreciable levels of α helix. This is a serious discrepancy and confirms the problems encountered in previous calculations and evaluations of force fields by others.[67,69,86]

From the above comparison and analysis, we suggest that the diffusive statistical random coil model is not applicable to the structure of unfolded proteins and polypeptides on short length scales. The data can be reconciled if we assume unfolded states of proteins have well-defined local structures that are not as heterogeneous as previously thought.

8. Conclusion and Summary

Though the random coil model has a long history of interpreting experimental results on unfolded proteins, in particular those from hydrodynamic and SAXS measurements of R_g,[27] it is inconsistent with older observations from Krimm's group as well as a wealth of spectroscopic techniques that demonstrate the presence of well-defined backbone structure within the unfolded states of proteins and peptides. It is then time to question prevailing views of unfolded proteins and consider alternatives.[11] It has been known for a long time that each residue/peptide backbone has a slightly different pK_a value[195] as well as distinctive inductive and blocking effects.[129] This is consistent with the presence of residue-specific conformational preferences. Of course, important questions still remain to be answered: Why is there a localized preference for certain backbone conformations for a given residue in a certain neighboring sequence context? What physical interactions are at work? How do these effects influence the overall chain conformation? Finally, how can we apply the information gained by answering the above and related questions to improve prediction of the native and unfolded structure of proteins and to understand the mechanism of protein folding, including the mechanism of diseases relating to protein folding?

9. Acknowledgments

Preparation of this manuscript and research carried out at NYU were supported by a grant from ONR (N.R.K.). We thank Bobby Arora, Buzz Baldwin, Trevor Creamer, Angel Garcia, Alex Kentsis, Ron Raines, George Rose, Tobin Sosnick, and Robert Woody for communicating results and insightful discussions. Correspondence with Adam Liwo and Harold Scheraga helped clarify our ideas on the role of PII. Z.S. thanks Professor Joshua Wand for helpful advice and support. Z.S. is a recipient of the Margaret and Herman Sokol Award for postdoctoral research from NYU.

10. References

(1) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121.
(2) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley: New York, 1969.
(3) Han, W. G.; Jalkanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587.
(4) Poon, C. D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642.
(5) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316.
(6) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727.
(7) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628.
(8) Shi, Z. S.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190.
(9) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980.
(10) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17964.
(11) Fitzkee, N. C.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12497.
(12) Krimm, S.; Tiffany, M. L. *Isr. J. Chem.* **1974**, *12*, 189.
(13) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220.
(14) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743.
(15) Drake, A. F.; Siligardi, G.; Gibbons, W. A. *Biophys. Chem.* **1988**, *31*, 143.
(16) Dukor, R. K.; Keiderling, T. A. *Biopolymers* **1991**, *31*, 1747.
(17) Woody, R. W. *Adv. Biophys. Chem.* **1992**, *2*, 37.
(18) Wilson, G.; Hecht, L.; Barron, L. D. *Biochemistry* **1996**, *35*, 12518.
(19) Dill, K. A. *Protein Sci.* **1999**, *8*, 1166.
(20) Baldwin, R. L.; Zimm, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12391.
(21) Baldwin, R. L. *Adv. Protein Chem.* **2002**, *62*, 361.
(22) Rose, G. D. *Adv. Protein Chem.* **2002**, *62*, xv.
(23) Shi, Z.; Woody, R. W.; Kallenbach, N. R. *Adv. Protein Chem.* **2002**, *62*, 163.
(24) Ronish, E. W.; Krimm, S. *Biopolymers* **1972**, *11*, 1919.
(25) Ronish, E. W.; Krimm, S. *Biopolymers* **1974**, *13*, 1635.
(26) Krimm, S.; Mark, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60*, 1122.
(27) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491.
(28) Godbole, S.; Bowler, B. E. *J. Mol. Biol.* **1997**, *268*, 816.
(29) Godbole, S.; Hammack, B.; Bowler, B. E. *J. Mol. Biol.* **2000**, *296*, 217.
(30) Hammack, B. N.; Smith, C. R.; Bowler, B. E. *J. Mol. Biol.* **2001**, *311*, 1091.
(31) Smith, C. R.; Wandschneider, E.; Bowler, B. E. *Biochemistry* **2003**, *42*, 2174.
(32) Park, S. H.; Shalongo, W.; Stellwagen, E. *Protein Sci.* **1997**, *6*, 1694.
(33) Sreerama, N.; Woody, R. W. *Biochemistry* **1994**, *33*, 10022.
(34) Woutersen, S.; Hamm, P. *J. Phys.: Condens. Matter* **2002**, *14*, R1035.
(35) Keiderling, T. A.; Silva, R.; Yoder, G.; Dukor, R. K. *Bioorg. Med. Chem.* **1999**, *7*, 133.
(36) Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1.
(37) Barron, L. D.; Blanch, E. W.; Hecht, L. *Adv. Protein Chem.* **2002**, *62*, 51.
(38) Gibrat, J. F.; Robson, B.; Garnier, J. *Biochemistry* **1991**, *30*, 1578.
(39) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *Proteins* **1992**, *12*, 345.
(40) Kang, H. S.; Kurochkina, N. A.; Lee, B. *J. Mol. Biol.* **1993**, *229*, 448.
(41) Munoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301.
(42) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596.
(43) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.
(44) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322.
(45) O'Connell, T. M.; Wang, L.; Tropsha, A.; Hermans, J. *Proteins* **1999**, *36*, 407.
(46) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr., D: Biol. Crystallogr.* **2002**, *58*, 768.
(47) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12*, 447.
(48) Hamburger, J. B.; Ferreon, J. C.; Whitten, S. T.; Hilser, V. J. *Biochemistry* **2004**, *43*, 9790.
(49) Syme, C. D.; Blanch, E. W.; Holt, C.; Jakes, R.; Goedert, M.; Hecht, L.; Barron, L. D. *Eur. J. Biochem.* **2002**, *269*, 148.
(50) Blanch, E. W.; Kasarda, D. D.; Hecht, L.; Nielsen, K.; Barron, L. D. *Biochemistry* **2003**, *42*, 5665.
(51) Parrot, I.; Huang, P. C.; Khosla, C. *J. Biol. Chem.* **2002**, *277*, 45572.
(52) Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2003**, *125*, 8178.
(53) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P. *Biochemistry* **2001**, *40*, 14376.
(54) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. *Proteins* **2003**, *53*, 68.
(55) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565.
(56) Pappu, R. V.; Rose, G. D. *Protein Sci.* **2002**, *11*, 2437.
(57) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574.
(58) Pophristic, V.; Goodman, L. *Nature* **2001**, *411*, 565.
(59) Weinhold, F. *Nature* **2001**, *411*, 539.
(60) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666.
(61) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188.
(62) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 2497.
(63) Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2003**, *125*, 8092.
(64) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A.; Schweitzer-Stenner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10054.
(65) Fitzkee, N. C.; Rose, G. D. *Protein Sci.* **2004**, *13*, 633.
(66) Garcia, A. E. *Polymer* **2004**, *45*, 669.
(67) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, *50*, 451.
(68) Kentsis, A.; Mezei, M.; Gindin, T.; Osman, R. *Proteins* **2004**, *55*, 493.
(69) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693.
(70) Asher, S. A.; Mikhonin, A. V.; Bykov, S. *J. Am. Chem. Soc.* **2004**, *126*, 8433.
(71) McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D. *J. Am. Chem. Soc.* **2004**, *126*, 5076.
(72) Woutersen, S.; Pfister, R.; Hamm, P.; Mu, Y. G.; Kosov, D. S.; Stock, G. *J. Chem. Phys.* **2002**, *117*, 6833.
(73) Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2002**, *124*, 14330.
(74) Eker, F.; Cao, X. L.; Nafie, L.; Huang, Q.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2003**, *107*, 358.
(75) Schweitzer-Stenner, R.; Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A. *J. Am. Chem. Soc.* **2004**, *126*, 2768.
(76) Chen, K.; Liu, Z.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15352.
(77) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(78) Lifson, S. *J. Chem. Phys.* **1961**, *34*, 1963.
(79) Asher, S. A., Griffiths, P. R., Eds. *Handbook of Vibrational Spectroscopy*; Wiley: New York, 2002.
(80) Weise, C. F.; Weisshaar, J. C. *J. Phys. Chem. B* **2003**, *107*, 3265.
(81) Pizzanelli, S.; Monti, S.; Forte, C. *J. Phys. Chem. B* **2005**, *109*, 21102.
(82) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131.
(83) Chellgren, B. W.; Creamer, T. P. *Biochemistry* **2004**, *43*, 5864.
(84) Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A. *Biophys. J.* **2004**, *86*, 731.
(85) Mu, Y. G.; Stock, G. *J. Phys. Chem. B* **2002**, *106*, 5294.
(86) Mu, Y. G.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B* **2003**, *107*, 5064.
(87) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins* **2004**, *55*, 502.
(88) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A.; Schweitzer-Stenner, R. *Biochemistry* **2004**, *43*, 613.
(89) Schweitzer-Stenner, R.; Eker, F.; Perez, A.; Griebenow, K.; Cao, X.; Nafie, L. A. *Biopolymers* **2003**, *71*, 558.
(90) Sreerama, N.; Woody, R. W. *Protein Sci.* **2003**, *12*, 384.
(91) Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1744.
(92) Tiffany, M. L.; Krimm, S. *Biopolymers* **1973**, *12*, 575.
(93) Schellman, J. A.; Schellman, C. *The Proteins*, 2nd ed.; Academic: New York, 1964.
(94) Kim, Y. S.; Wang, J. P.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109*, 7511.
(95) Shi, Z.; Chen, K.; Liu, Z.; Sosnick, T. R.; Kallenbach, N. R. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 312.
(96) Cao, W.; Bracken, C.; Kallenbach, N. R.; Lu, M. *Protein Sci.* **2004**, *13*, 177.
(97) Liu, Z.; Chen, K.; Ng, A.; Shi, Z.; Woody, R. W.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2004**, *126*, 15141.
(98) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry* **2005**, *44*, 6269.

(99) Martino, M.; Bavoso, A.; Guantieri, V.; Coviello, A.; Tamburro, A. M. *J. Mol. Struct.* **2000**, *519*, 173.
(100) Bochicchio, B.; Tamburro, A. M. *Chirality* **2002**, *14*, 782.
(101) Tamburro, A. M.; Bochicchio, B.; Pepe, A. *Biochemistry* **2003**, *42*, 13347.
(102) Bochicchio, B.; Ait-Ali, A.; Tamburro, A. M.; Alix, A. J. *Biopolymers* **2004**, *73*, 484.
(103) Bochicchio, B.; Floquet, N.; Pepe, A.; Alix, A. J. P.; Tamburro, A. M. *Chem.—Eur. J.* **2004**, *10*, 3166.
(104) Ma, K.; Kan, L.; Wang, K. *Biochemistry* **2001**, *40*, 3427.
(105) Ma, K.; Wang, K. *Biochem. J.* **2003**, *374*, 687.
(106) Jha, A.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691.
(107) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10967.
(108) Creamer, T. P.; Campbell, M. N. *Adv. Protein Chem.* **2002**, *62*, 263.
(109) Chou, P. Y.; Fasman, G. D. *Annu. Rev. Biochem.* **1978**, *47*, 251.
(110) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. *Fold. Des.* **1996**, *1*, R95.
(111) Baldwin, R. L. *Biophys. Chem.* **2002**, *101*−*102*, 203.
(112) Bolin, K. A.; Pitkeathly, M.; Miranker, A.; Smith, L. J.; Dobson, C. M. *J. Mol. Biol.* **1996**, *261*, 443.
(113) Adzhubei, A. A.; Sternberg, M. J. *J. Mol. Biol.* **1993**, *229*, 472.
(114) Adzhubei, A. A.; Sternberg, M. J. *Protein Sci.* **1994**, *3*, 2395.
(115) Stapley, B. J.; Creamer, T. P. *Protein Sci.* **1999**, *8*, 587.
(116) Voet, D.; Voet, J. G. *Biochemistry*, 3rd ed.; Wiley: New York, 2004.
(117) Butterfoss, G. L.; Hermans, J. *Protein Sci.* **2003**, *12*, 2719.
(118) Shortle, D. *Protein Sci.* **2003**, *12*, 1298.
(119) Shortle, D. *Protein Sci.* **2002**, *11*, 18.
(120) Shi, Z. S.; Olson, C. A.; Kallenbach, N. R.; Sosnick, T. R. *J. Am. Chem. Soc.* **2002**, *124*, 13994.
(121) Shi, Z. S.; Krantz, B. A.; Kallenbach, N.; Sosnick, T. R. *Biochemistry* **2002**, *41*, 2120.
(122) Whittington, S. J.; Creamer, T. P. *Biochemistry* **2003**, *42*, 14690.
(123) Lyu, P. C.; Gans, P. J.; Kallenbach, N. R. *J. Mol. Biol.* **1992**, *223*, 343.
(124) Smith, J. S.; Scholtz, J. M. *Biochemistry* **1998**, *37*, 33.
(125) Olson, C. A.; Spek, E. J.; Shi, Z.; Vologodskii, A.; Kallenbach, N. R. *Proteins* **2001**, *44*, 123.
(126) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403.
(127) Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431.
(128) Fitzkee, N. C.; Rose, G. D. *J. Mol. Biol.* **2005**, *353*, 873.
(129) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75.
(130) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272.
(131) Petty, S. A.; Volk, M. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1022.
(132) Jiji, R. D.; Balakrishnan, G.; Hu, Y.; Spiro, T. G. *Biochemistry* **2006**, *45*, 34.
(133) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. *J. Am. Chem. Soc.* **2001**, *123*, 2388.
(134) Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1994**, *116*, 6558.
(135) Kramer, R. Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M. *J. Mol. Biol.* **1998**, *280*, 623.
(136) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6*, 63.
(137) Eberhardt, E. S.; Panasik, N.; Raines, R. T. *J. Am. Chem. Soc.* **1996**, *118*, 12261.
(138) Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A. *Biopolymers* **2001**, *58*, 459.
(139) Torchia, D. A. *Macromolecules* **1971**, *4*, 440.
(140) Torchia, D. A. *Macromolecules* **1971**, *5*, 566.
(141) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777.
(142) Jenkins, C. L.; Raines, R. T. *Nat. Prod. Rep.* **2002**, *19*, 49.
(143) Horng, J. C.; Raines, R. T. *Protein Sci.* **2006**, *15*, 74.
(144) Levinthal, C. *J. Chim. Phys. Phys.-Chim. Biol.* **1968**, *65*, 44.
(145) Hoshino, M.; Hagihara, Y.; Hamada, D.; Kataoka, M.; Goto, Y. *FEBS Lett.* **1997**, *416*, 72.
(146) Kamatari, Y. O.; Konno, T.; Kataoka, M.; Akasaka, K. *J. Mol. Biol.* **1996**, *259*, 512.
(147) Nelson, J. W.; Kallenbach, N. R. *Proteins* **1986**, *1*, 211.
(148) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369.
(149) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.
(150) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(151) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(152) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.
(153) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(154) Anderson, A. G.; Hermans, J. *Proteins* **1988**, *3*, 262.
(155) Tobias, D. J.; Brooks, C. L. *J. Phys. Chem.* **1992**, *96*, 3864.
(156) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099.
(157) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568.
(158) Gnanakaran, S.; Garcia, A. E. *J. Phys. Chem. B* **2003**, *107*, 12555.
(159) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
(160) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. *J. Am. Chem. Soc.* **2004**, *126*, 698.
(161) Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437.
(162) Ohkubo, Y. Z.; Brooks, C. L., 3rd. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13916.
(163) Ramakrishnan, V.; Ranbhor, R.; Durani, S. *J. Am. Chem. Soc.* **2004**, *126*, 16332.
(164) Blanch, E. W.; Morozova-Roche, L. A.; Cochran, D. A.; Doig, A. J.; Hecht, L.; Barron, L. D. *J. Mol. Biol.* **2000**, *301*, 553.
(165) Juszczak, L. J. *J. Biol. Chem.* **2004**, *279*, 7395.
(166) Maiti, N. C.; APetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. *J. Am. Chem. Soc.* **2004**, *126*, 2399.
(167) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, *41*, 415.
(168) Uversky, V. N. *Eur. J. Biochem.* **2002**, *269*, 2.
(169) Uversky, V. N.; Fink, A. L. *Biochim. Biophys. Acta* **2004**, *1698*, 131.
(170) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. R.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C. H.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, M.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modell.* **2001**, *19*, 26.
(171) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323*, 573.
(172) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(173) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321.
(174) Gill, A. C.; Ritchie, M. A.; Hunt, L. G.; Steane, S. E.; Davies, K. G.; Bocking, S. P.; Rhie, A. G.; Bennett, A. D.; Hope, J. *EMBO J.* **2000**, *19*, 5324.
(175) Jarvet, J.; Damberg, P.; Danielsson, J.; Johansson, I.; Eriksson, L. E.; Graslund, A. *FEBS Lett.* **2003**, *555*, 371.
(176) Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *Biochemistry* **2004**, *43*, 6893.
(177) Antonyraj, K. J.; Karunakaran, T.; Raj, P. A. *Arch. Biochem. Biophys.* **1998**, *356*, 197.
(178) Armen, R. S.; DeMarco, M. L.; Alonso, D. O.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11622.
(179) Woody, R. W. *Biophys. J.* **2004**, *86*, 617A.
(180) Woody, R. W. *Monatsh. Chem.* **2005**, *136*, 347.
(181) McPhie, P. *Anal. Biochem.* **2001**, *293*, 109.
(182) Baker, B. R.; Garrell, R. L. *Faraday Discuss.* **2004**, *126*, 209.
(183) Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 1425.
(184) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487.
(185) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977.
(186) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.
(187) Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063.
(188) Wang, Y.; Shortle, D. *Biochemistry* **1995**, *34*, 15895.
(189) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158.
(190) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170.
(191) Zhang, O.; Kay, L. E.; Shortle, D.; Forman-Kay, J. D. *J. Mol. Biol.* **1997**, *272*, 9.
(192) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 13791.
(193) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 3089.
(194) Zagrovic, B.; Sorin, E. J.; Millett, I. S.; van Gunsteren, W. F.; Doniach, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11698.
(195) Veis, A.; Nawrot, C. F. *J. Am. Chem. Soc.* **1970**, *92*, 3910.

CR040433A