Poly(Pro)II Helices in Globular Proteins: Identification and Circular Dichroic Analysis†

Narasimha Sreerama and Robert W. Woody*

Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523

Received March 18, 1994; Revised Manuscript Received June 14, 1994

ABSTRACT: A method to identify poly(l-proline)-type (P) conformation in crystal structures of globular proteins is presented. Short segments of P conformation were identified in globular protein structures, and these form a significant fraction of the residues which are not assigned to α-helix, β-sheet, and β-turns. The fractions of α-helix, β-sheet, β-turns, P, and unordered, identified in conjunction with the Kabsch and Sander method (1983) Biopolymers 22, 2577, were incorporated in the analysis of circular dichroism (CD) spectra of proteins. The separation of P fraction from the fraction of residues not assigned to α-helix or β-sheet or -turns resulted in a distinctive P CD spectrum and an unusual CD spectrum corresponding to the residual unordered structures. The quality of prediction of P fraction from CD spectra of proteins was comparable to that of β-sheet and -turns.

The three major types of secondary structures recognized in globular proteins are α-helices, β-sheets, and β-turns. The α-helices and β-sheets are identified by the existence of at least one hydrogen bond between the backbone C=O and N-H groups, while the β-turns may have at most one such hydrogen bond (Pauling et al., 1951; Venkatachalam, 1968; Richardson, 1981). These are also defined, in idealized geometries, by the φ and ψ angles making up the structure (Pauling et al., 1951; Venkatachalam, 1968; Cantor & Schimmel, 1980; Richardson, 1981). The secondary structure elements in X-ray-derived structures deviate from the ideal geometry, and algorithms to identify the secondary structure elements in globular proteins have been developed (Levitt & Greer, 1977; Kabsch & Sander, 1983). Up to 25% of the residues in globular protein structures remain unassigned. These residues have been referred to by different terms by different investigators: random coil (Perczel et al., 1991; Böhm et al., 1992); unordered conformation (Chang et al., 1978; Perczel et al., 1991; Sreerama & Woody, 1993); irregular regions (Bolotina et al., 1980); other structures (Hennessey & Johnson, 1981; van Stokkom et al., 1991; Pancoska & Keiderling, 1991); and remainder (Provencher & Glöckner, 1981; Venyaminov et al., 1991). We continue to use the term “unordered” for those residues that are not assigned to a well-defined secondary structure. This term does not imply that these unassigned residues are dynamically unordered or that their conformation varies from one individual protein molecule to another.

There is evidence that some of the unassigned residues show at least short-range order, short segments of poly(Pro)II helix type (P) structure (Adzhubei et al., 1987a,b; Adzhubei & Sternberg, 1993; Woody, 1992). The P structure is a left-handed extended helix with three residues per turn, has the backbone C=O and N-H groups projecting outward, and is favored in proline-rich polypeptides due to the limited defined secondary structure. This term does not imply that Schimmel, 1980; Woody, 1992).

The geometry, and algorithms to identify the secondary structure

† Supported by NIH Grant GM22994.
* Address correspondence to this author. Ph: (303) 491-6214; Fax: (303) 491-0494; E-mail: rww@lamar.colostate.edu.
1 Abbreviations: P, poly(Pro)II-helix type; CD, circular dichroism; r, correlation coefficient; b, root-mean-square deviation.

METHODS

Identification of the P Conformation. We identify the P structure using geometric features defined by the angles τ and ζ, which are as follows: τ = virtual bond angle |C(i−1)−C(i)−C(i+1)|; ζ = virtual dihedral angle |O(i−1)=C-
The regular secondary structures with repeating φ, ψ angles have typical τ and ζ: Right-handed helical structures have positive ζ; the left-handed structures have negative ζ, and the β-structure has ζ close to 180°. The virtual angle ζ gives the geometric relation between the successive carbonyl groups and indicates the handedness of the propagating chain in a shorter stretch of residues. The τ and ζ angles in regular structures identified from X-ray coordinates, such as α-helices and β-sheets, deviate from their ideal values because of deviations from the ideal geometry. The criteria for identifying PII structure were derived from the values of these virtual angles in idealized structures and their deviations in X-ray structures.

A residue is assigned to the PII structure if the virtual angles τ and ζ do not deviate more than 15° and 25°, respectively, from their ideal values of 120° and -115°, respectively (i.e., 105° < τ < 135°; -140° < ζ < -90°). The criteria are relaxed to extend the PII structure as follows: (a) if the previous residue is in PII structure, then the allowed deviations for τ and ζ are relaxed to ±20° and ±35°, respectively; (b) a residue flanked by residues in PII structure is assigned to the PII structure if 100° < τ < 140° (20° deviation from its ideal value) and ζ < 0° (left-handed propagation).

**Proteins.** The X-ray structures of the following 16 proteins, which formed our basis set for CD analysis, were taken from the Protein Data Bank (Bernstein et al., 1977). The proteins and the X-ray structures used (PDB code in parenthesis) are as follows: Bene-Jones protein (1rei), prealbumin (2pab), rubredoxin (3rxn), α-chymotrypsin (5cha), elastase (3est), papain (9pap), thermolysin (3thl), lysozyme (7lyz), subtilisin BPN' (1bst), glyceraldehyde-3-phosphate dehydrogenase (3gpd), flavodoxin (1fxl), lactate dehydrogenase (51dh), triosephosphate isomerase (1tim), cytchrome c (3cyt), hemoglobin (2mbh), and myoglobin (4mbn). Of these, the first five are ββ proteins, the next three are α/β proteins, the next five are α/αβ proteins, and the last three are αα proteins (Levitt & Chothia, 1976).

**Analysis of CD Spectra.** The CD spectra associated with various types of secondary structures were deconvoluted from CD spectra of the basis set proteins. Our basis set consisted of 16 proteins and poly(l-Glu), an all-α polypeptide, which was similar to the one used in our previous study (Sreerama & Woody, 1993). The CD spectra of these proteins and poly(l-Glu) were kindly provided by Dr. W. C. Johnson, Jr. The method followed to obtain secondary structure CD spectra was similar to that followed by Compton and Johnson (1986). The matrix containing the basis CD data, **C**, is expressed as a product of three matrices using the singular value decomposition algorithm (Forsythe et al., 1977), **C** = **USV**<sup>T</sup>, where **U** and **V** are unitary matrices and **S** is a diagonal matrix. This is incorporated in the matrix equation relating the CD spectra to the secondary structure data matrix, **F** = **XC**. The multiplicative inverse of **X**, which is **FVS**<sup>T</sup><sup>U</sup><sup>T</sup>, gives the spectra corresponding to the unknown secondary structure considered in constructing **F**.

The CD spectrum of the protein analyzed for secondary structure was removed from our basis set, and the secondary structure fractions were predicted using the other members of the basis set, following the self-consistent method (Sreerama & Woody, 1993). In the self-consistent method the spectrum of the protein analyzed is included in the matrix of CD spectral data, and an initial guess, the structure of the protein having the CD spectrum most similar to that of the protein analyzed, is made for the unknown secondary structure. The matrix equation relating the CD spectra to the secondary structure, **F** = **XC**, is solved by the singular value decomposition algorithm (Forsythe et al., 1977). The solution obtained replaces the initial guess, and the process is repeated until self-consistency is reached.

The performance of the analysis is expressed as root-mean-square deviations (δ) and correlation coefficients (r) between the X-ray and CD estimates of secondary structure fractions for different secondary structure assignments. δ and r were calculated using the equations:

\[
\delta = \left( \frac{1}{N} \sum (f_i^x - f_i^y)^2 \right)^{1/2}
\]

\[
r = \frac{N \sum f_i^x f_i^y - \left( \sum f_i^x \right) \left( \sum f_i^y \right)}{\left[ N \sum (f_i^x)^2 - \left( \sum f_i^x \right)^2 \right] \left[ N \sum (f_i^y)^2 - \left( \sum f_i^y \right)^2 \right]^{1/2}}
\]

where **f**<sup>x</sup> and **f**<sup>y</sup> are the X-ray and CD estimates of secondary structure fractions of **N** samples.

**RESULTS AND DISCUSSION**

**Identification of the PII Conformation.** The secondary structure fractions obtained from our method in conjunction with Kabsch and Sander (1983) method, for the proteins in our basis set, are given in **Table 1**. We start with the Kabsch and Sander (1983) assignments of secondary structures (KS), which use hydrogen bond patterns, and assign seven types of secondary structures (α-helix, 3_10-helix, π-helix, β-sheet, turns, β-bridge, and bends). The residues assigned to helix and β-sheet were eliminated, and the rest were examined for the bridges are isolated β-bridges and were considered in the unordered fraction. In effect, we consider PII to be a higher order structure than turns and reassigned some residues assigned to turns, bends, and bridges to PII.

Approximately 10% of the residues were assigned to the PII structure among the proteins in our basis set, and the PII fraction in these proteins varied from 0.024 (51dh) to 0.231 (3rxn) (Table 1). Generally, proteins with higher α-helix content had less PII structure, and a negative correlation was
found between the fractions of α-helix and PII (−0.533, Table 1). Among the residues assigned to the PII structure, approximately 50% were isolated residues. Two residues in PII structure would result in three successive C=O groups arranged as in one turn of a PII helix, which has implications for exciton interactions and the resulting CD spectra. However, whether isolated residues assigned to the PII structure need to be considered as PII or left unassigned is not clear. A single residue in the PII conformation cannot be considered as a PII helix, but even in an isolated PII residue two successive amide groups are oriented so that the exciton interaction expected in a PII helix is possible. We evaluated both these possibilities by considering all residues in the PII structure (SW1 assignments) and only those in chains of PII structure with two or more residues (SW2 assignments), and using the resulting fractions of secondary structures in the analysis of CD spectra.

We examined the possibility that turns are higher order structures than PII because of the existence of a main-chain hydrogen bond in turns identified from the KS method (results not presented). While more than 85% of the residues in PII belonged to unassigned residues from KS, only 0.5% residues were classified both as turns and as PII, and these were either at the beginning or at the end of a PII helix. These PII assignments differed only a little from those obtained by giving precedence to PII.

Secondary Structure CD Spectra. The CD spectra associated with various types of secondary structures were deconvoluted from the basis CD spectra associated with various types of secondary structures were deconvoluted from the basis CD spectra and are given in Figure 1. These are similar in several cases to the CD spectra of synthetic polypeptides in the corresponding conformations (Figure 1A). The CD spectrum of the α-helix shows characteristic bands at 190, 208, and 220 nm, and that of β-sheet shows bands at 190 and 220 nm. The turns, with different sets of φ and ψ, give different classes of CD spectra, and the spectrum we calculate from the protein CD spectra corresponds to class C' (Woody, 1985), with an inverted α-helix-like spectrum. Similar β-turn CD spectra were obtained by Chang et al. (1978), Bolotina et al. (1980), Compton and Johnson (1986), and van Stokkum et al. (1990).

Neither the type I (III) nor type II β-turns, which are prevalent in proteins, are expected to give such a spectrum (Woody, 1985). For reasons which are unclear, aromatic contributions may segregate with the β-turns in this procedure. The CD spectra corresponding to the PII fraction (Figure 1B), obtained by considering either PII segments of at least two residues (curve 4a) or all residues in PII (curve 4), have a strong negative band at 195 nm and a weak positive band at 215 nm, which are characteristic of the poly(Pro)II helix (Woody, 1992). These bands were observed in the CD spectra of unordered polypeptides but with smaller amplitudes (Woody, 1985, 1992). On the other hand, the CD spectrum we calculate for the unordered fraction, after defining the PII structure, has a weak positive band at 198 nm and a weak negative band at 219 nm and qualitatively resembles the β-sheet CD spectrum. Our definition of the PII structure has resulted in a distinctive PII CD spectrum with features consistent with model polypeptide spectra.

Table 2: Performance of CD Analyses for Different Secondary Structure Assignments

<table>
<thead>
<tr>
<th>secondary structure assignment</th>
<th>α-helix</th>
<th>β-sheet</th>
<th>turns</th>
<th>PII</th>
<th>unordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>0.084</td>
<td>0.972</td>
<td>0.074</td>
<td>0.834</td>
<td>0.048</td>
</tr>
<tr>
<td>SW1</td>
<td>0.071</td>
<td>0.973</td>
<td>0.073</td>
<td>0.824</td>
<td>0.046</td>
</tr>
<tr>
<td>SW2</td>
<td>0.068</td>
<td>0.974</td>
<td>0.074</td>
<td>0.823</td>
<td>0.044</td>
</tr>
</tbody>
</table>

*KK assignments correspond to the assignments from Kabsch and Sander (1983) method with no PII classification, resulting in four secondary structures. SW1 and SW2 assignments have five secondary structure fractions with PII assignments from this work. SW1 includes isolated PII residues in the PII fraction. SW2 does not include isolated PII residues in the PII fraction.

Figure 1: CD spectra associated with various types of secondary structures, deconvoluted from protein CD spectra. (A) CD spectra of α-helix (curve 1), β-sheet (curve 2), and turns (curve 3). (B) CD spectra of PII, considering all residues in the PII conformation (curve 4) and only those residues in PII segments of two or more residues (curve 4a), and the corresponding unordered conformation (curves 5 and 5 respectively).
fractions, \( f_{\text{max}} - f_{\text{min}} \), for \( \alpha \)-helix, \( \beta \)-sheet, turns, PII, and unordered were 0.972, 0.491, 0.310, 0.231, and 0.240 for SW1 and 0.972, 0.491, 0.333, 0.173, and 0.276 for SW2. Each RMS difference should be divided by the dynamic range of the corresponding fractions (Pancoska et al., 1992) to obtain a better comparison. For the four secondary structures, \( \alpha \)-helix, \( \beta \)-sheet, turns, and PII, and the unordered fraction, the RMS differences relative to the dynamic range using the SW1 assignments were 0.15, 0.15, 0.19, and 0.33, respectively, indicating that the prediction of PII fraction is on a par with that of \( \beta \)-sheet. Inclusion of isolated PII residues in either the unordered or the turns fraction worsens the predictions of turns and PII (Table 2, SW2). The prediction indices for the unordered fraction are worse than those obtained with KS assignments. The separation of the PII fraction from the unassigned set of KS has left only the residues with no obvious secondary structure in the unordered fraction, which has resulted in the low correlation between the predicted and the X-ray fraction of the unordered fraction.

Should the isolated PII residues be classified as PII or unordered? The PII CD obtained resembles the poly(Pro)II CD, regardless of whether or not the isolated PII residues are included in the PII fraction (Figure 1B, curves 4 and 4a). However, the CD spectrum corresponding to the unordered fraction obtained when isolated PII residues are included in the unordered fraction (Figure 1B, curve 5a) is different from that obtained by considering them as PII (Figure 1B, curve 5). The quality of prediction is improved by including the isolated PII residues in the PII fraction. This suggests that the isolated PII residues should be included in the PII fraction in these analyses.

**CONCLUSIONS**

We have developed a method to identify poly(Pro)II type structure in globular protein structures and used the resulting secondary structural fractions in the analysis of protein CD spectra. Our attempt to quantitate the PII structure as a significant fraction of the unassigned structure in proteins has been successful because of its characteristic CD spectrum. While the high-\( \alpha \) proteins are likely to have less PII structure, no significant correlation was obtained between the high-\( \beta \) proteins and PII structure. However, only a small number of structures were used in this study, and an analysis of available protein structures is in order. Most of the longer PII helices are exposed to solvent, implying stabilization from solvent molecules, and molecular dynamics simulations also point to stabilization by a hydrogen-bonded network (Sreerama & Woody, 1992). The conformational rigidity of proline more than doubles its frequency of occurrence in the PII conformation relative to its average frequency of occurrence in proteins (Adzhubei et. al., 1987a,b). The amino acid propensities toward forming PII structure should be important in sequence-based structure assignment (Chou & Fasman, 1978; Garnier et al., 1978; Holley & Karplus, 1989), and these have been calculated using longer segments of PII structure (Adzhubei & Sternberg, 1993). A survey of all available protein structures for characterizing the amino acid preferences is in order, and studies toward that goal are currently underway.

**ACKNOWLEDGMENT**

We acknowledge helpful discussions with Alexei A. Adzhubei and thank him and Mark C. Manning for reviewing the manuscript.

**REFERENCES**