Properties of Polyproline II, a Secondary Structure Element Implicated in Protein–Protein Interactions

M. V. Cubellis,[1]* F. Caillez,[2] T. L. Blundell,[3] and S. C. Lovell[4]
[1]*Dipartimento di Chimica Biologica, Napoli, Italia*
[2]*Institut de Biologie Physico-Chimique, Paris, France*
[3]*Biochemistry Department, University of Cambridge, Cambridge, United Kingdom*
[4]*School of Biological Sciences, University of Manchester, Manchester, United Kingdom*

ABSTRACT **The polyproline II (PPII) conformation of protein backbone is an important secondary structure type. It is unusual in that, due to steric constraints, its main-chain hydrogen-bond donors and acceptors cannot easily be satisfied. It is unable to make local hydrogen bonds, in a manner similar to that of α-helices, and it cannot easily satisfy the hydrogen-bonding potential of neighboring residues in polyproline conformation in a manner analogous to β-strands. Here we describe an analysis of polyproline conformations using the HOMSTRAD database of structurally aligned proteins. This allows us not only to determine amino acid propensities from a much larger database than previously but also to investigate conservation of amino acids in polyproline conformations, and the conservation of the conformation itself. Although proline is common in polyproline helices, helices without proline represent 46% of the total. No other amino acid appears to be greatly preferred; glycine and aromatic amino acids have low propensities for PPII. Accordingly, the hydrogen-bonding potential of PPII main-chain is mainly satisfied by water molecules and by other parts of the main-chain. Side-chain to main-chain interactions are mostly nonlocal. Interestingly, the increased number of nonsatisfied H-bond donors and acceptors (as compared with α-helices and β-strands) makes PPII conformers well suited to take part in protein–protein interactions. Proteins 2005;58:880–892.** © 2005 Wiley-Liss, Inc.

Key words: polyproline II; protein–protein interaction

INTRODUCTION

Secondary structure in proteins is produced by regular repeating main-chain conformation. For the classic secondary structure types, α-helices and β-strands, these repetitive structures allow the formation of main-chain hydrogen bonds within helices and between strands. There are, however, other repetitive conformations that do not allow satisfaction of main-chain hydrogen-bond donors and acceptors by other residues in a similar conformation. The polyproline type II (PPII) helix is one of these conformations.

PPII helix is a left-handed helical structure with an overall shape resembling a triangular prism. It is quite extended, with a helical pitch of 9.3 Å/turn, 3 residues per turn, and ϕ and ψ angles centered around $-75°$ and 145°, respectively. The lack of characteristic local main-chain hydrogen-bonding patterns means that they are harder to detect directly in NMR spectra than α-helices and β-sheets. Also they cannot be identified in protein structures by hydrogen-bonding patterns, as is commonly done for other secondary structure types.[1,2] As a consequence, residues in PPII conformations are often inappropriately assigned to "disordered" or "random" conformation classes.

There has recently been an increase of interest in PPII conformations. They have been implicated in amyloid formation[3,4] and nucleic acid binding.[5] However, the realization that it is a common conformation in the unfolded state has led to a large number of studies. Interestingly, this was first suggested in 1968 by Tiffany and Krimm,[6] based on the similarities between circular dichroism spectra of unfolded proteins and proline polymers. Shi et al.[7] have found that a peptide made of 7 alanines (too short to form an α-helix) has a predominantly PPII conformation. Similarly, a 21-residue peptide that forms an α-helix under native conditions has been shown to melt to a PPII conformation.[8] In an elegant set of experiments on short peptides, Ferreon and Hilser[9] showed that Ala and Gly spend approximately 30% and 10% of the time, respectively, in PPII conformations. This is a smaller proportion than found by Shi et al.,[7] but this probably represents the differences in the experimental systems used.

These experiments are complemented by a number of theoretical studies, all of which suggest that solvents, specifically, peptide–solvent interactions, are the major determinants of PPII conformation.[10–13] Most intriguingly, PPII has been shown to disrupt water organization less than β-sheet and α-helical conformations, and is therefore entropically favored.[11,12] Kentsis et al.[11] argue that the hydrophobic effect is therefore not only the major driving force in protein folding but also is responsible for

Grant sponsor: Wellcome Trust Fellowship in Mathematical Biology (to S. C. Lovell). Grant sponsor: EMBO short-term fellowship (to M. V. Cubellis).

*Correspondence to: M. V. Cubellis, Dipartimento do Chimica Biologica, Via Mezzocannone 16, 80134, Napoli, Italia. E-mail: cubellis@unina.it

Received 8 May 2004; Accepted 5 August 2004

Published online 18 January 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20327

preorganizing unfolded peptides. This is in accordance with host–guest experiments, which allow the ranking of PPII helix forming propensity in the solvent-exposed state.[14–16] Hydrophobic residues, particularly β-branched hydrophobic residues, are particularly poor at forming PPII in these systems. These residues are precisely those most expected to disrupt main-chain to solvent interactions and, according to the arguments of Kentsis et al.,[11] would be expected to have a low propensity for PPII.

All of these studies concern the either short, highly solvated peptides or the unfolded state. How do these results correlate with folded proteins? Adzhubei and Sternberg[17] found 96 PPII helices in a database of 80 proteins. This was thought to be unusually common but still only constituted 5% of residues. They found that these PPII helices were unusually solvent-exposed and tended to have high crystallographic temperature factors. They suggested that PPII helices are often stabilized by main-chain–water hydrogen bonds (in the absence of main-chain–main-chain H-bonds), and tend to have a regular pattern of hydrogen bonds with water. They are, however, still much less solvent-accessible than experimentally studied peptides. Stapley and Creamer[18] additionally suggested that local side-chain to main-chain hydrogen bonds are important in stabilizing PPII helices. They found, however, that specific rotamer preferences were weak.

In this study, we have undertaken a survey of PPII helices in known protein structures. We use a large data set of 3268 proteins structurally aligned in 978 homologous families, taken from the HOMSTRAD database.[19,20] This choice allows us to identify not only unusual amino acid propensities in PPII conformations but also unusual conservation and substitution patterns. We find that PPII helices are stabilized by nonlocal interactions. They do not tend to have strong sequence propensities, in contrast with other extended conformations, such as β-strands. The nonlocal stabilization of hydrogen-bond donors and acceptors does, however, result in PPII conformations being well suited for participating in protein–protein interactions.

MATERIALS AND METHODS

Proteins were taken from the HOMSTRAD[19,20] database of structurally aligned homologous proteins (www-cryst.bioc.cam.ac.uk/~homstrad/). HOMSTRAD families have, on the average, more than 30% sequence identity, and even when the lowest pairwise percentage sequence identity within a family is less than 20, there is a "bridge" protein with more than 20% sequence identity between them. In some cases, a subset of HOMSTRAD was used that comprised only sequences of better than 1.8-Å resolution. Where this subset was used, we have indicated so in the text. The average number of proteins per family was 3.4.

Secondary structure was assigned using SEGNO, a program developed by Cubellis, Caillez and Lovell (manuscript in preparation) that uses geometric parameters to assign secondary structure.

Definition of PPII Stretches

To define PPII helices, we measured 4 dihedral angles: ϕ, ψ, diheco [the dihedral angle between O(i − 1), C(i − 1), C(i), O(i), where i represents the residue number] and diheco2 [the dihedral angle between O(i − 1), C(i − 1), C(i + 1), O(i + 1)]. The 2 dihedral angles, diheco and diheco2, represent the angles between the planes of successive peptide bonds, separated by 1 and 2 residues, respectively.

We temporarily assign a residue to a PPII conformation if it has not been previously assigned to the β-strand by SEGNO and if

1. ϕ = −80° ± 45°
2. ψ = 142.5° ± 42.5°
3. diheco = 240° ± 60°
4. diheco2 = 120° ± 40°

That is, if ϕ and ψ have appropriate values, and if there is approximate 3-fold symmetry in the PPII helix.

We maintain the assignment only for those stretches of residues that have

1. an average diheco in the range 220–270°
2. an average diheco2 in the range 100–140°.

Less strict ϕ and ψ restraints are used for the residues at the end of PPII stretches: ψ for the last residue in a PPII helix is allowed to be 142.5° ± 52.5°, whereas ϕ for the first residue of PPII helix is allowed to be −85° ± 60°. Deviations of the first and the last dihedral angles in the stretch, in fact, do not influence the left-handed helical structure, with the overall shape resembling a triangular prism. PPII helices have a minimal length of 3.

SEGNO can also be used to define isolated β-strands, because it uses a geometric rather than hydrogen-bonding approach. For this, it uses diheco and diheco2 values of 180° ± 40° (full details will be published elsewhere). Thus, isolated β-strands and PPII helices are quite distinct conformations, and it is appropriate to treat them separately.

SEGNO can also define "mixed isolated strands" and "mixed strands in sheet." These stretches of amino acids are not entirely structurally distinct from either PPII helices or β-strands, and so have not been used here except where indicated.

Residue Propensities

The propensity of a given residue with a particular attribute is defined as the ratio of the actual number of observations to the expected number of observations, where the expected number of observations is given by $n^*(n_{Xaa}/n_{total})$, where n_{Xaa} is the total number of residue type Xaa in the data set, n_{total} is the total number of all residues in the data set, and n^* is the total number of residues in the data set with the attribute of interest. For tests of statistical significance a T test was used, with propensity = 1 being the null hypothesis.

For the assessment of the proline content of PPII helices (Table I), expected values of occurrences are calculated

TABLE I. Length Distribution and Proline Content of Polyproline II Helices

Length	3	4	5	6	7	8	9	10	11	12	13	14	Total
Helices	3759	1226	454	140	67	25	10	7	3	1	1	1	5700
Helices without Pro	1943	488	172	29	7	1	1	1	0	0	0	0	2642
Helices/total (%)	65.9	21.5	8.0	2.5	1.2	0.4	0.2	0.1	<0.1	<0.1	<0.1	<0.1	100
Helices without Pro/helices (%)	52	40	38	21	10	4	10	14	0	0	0	0	46
Helices without Pro expected	2016	534	161	40	16	5	2	1	0	0	0	0	2775

TABLE II. Constrained Substitution Tables for Residues in Polyproline II Helices or in Coils

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	**6**/6																			
C	4/-3	**17**/19																		
D	-2/-2	-3/-5	**9**/6																	
E	-1/-1	-5/-6	2/1	**7**/6																
F	-3/-2	-11/-3	-7/-5	-6/-4	**12**/10															
G	2/-2	-5/-6	-3/-2	-3/-2	-10/-5	**12**/6														
H	-8/-2	-5/-8	-2/-2	-1/-1	2/0	-5/-3	**14**/11													
I	-3/-1	-7/-6	-8/-6	-5/-3	0/1	-6/-6	-8/-4	**8**/8												
K	-1/-1	-7/-5	1/-1	2/1	-5/-4	0/-2	1/0	-6/-3	**6**/6											
L	-3/-2	-12/-7	-8/-5	-5/-3	2/2	-6/-6	-5/-2	3/4	-7/-2	**7**/8										
M	-4/-1	-14/-5	-2/-4	-7/-2	0/1	-6/-4	-4/-2	2/3	-4/-2	3/4	**11**/10									
N	-1/-2	-5/-5	3/1	0/0	-2/-3	-1/-1	4/1	-7/-4	-1/0	-7/-4	0/-3	**9**/6								
P	-1/0	-13/-6	-4/-3	-2/-1	-6/-3	-3/-4	-4/-3	-2/-2	-1/-1	-4/-3	-3/-3	-5/-3	**5**/8							
Q	0/0	-6/-5	0/-1	2/2	-5/-2	0/2	1/1	-5/-3	2/2	-2/-2	-2/0	1/0	-2/-1	**7**/6						
R	-3/-1	-5/-5	-4/-2	0/0	-3/-3	-4/-3	-2/0	-5/-3	2/3	-5/-2	-5/-1	-2/-1	-3/-2	1/1	**8**/8					
S	1/1	-4/-6	0/0	-1/0	-7/-3	1/-2	-8/-2	-5/-3	-1/-1	-6/-3	-4/-2	1/0	-2/-1	-1/0	-3/-1	**7**/5				
T	-1/0	-5/-5	-2/-1	-1/-1	-3/-3	-2/-3	0/-2	-3/-2	0/0	-3/-2	-4/-1	0/0	-2/-2	-1/0	-1/-1	3/2	**7**/6			
V	0/1	-14/-6	-4/-5	-3/-2	-2/-1	-2/-5	-3/-3	4/5	-2/-2	1/2	-1/2	0/-3	-3/-2	-1/-2	-3/-3	-2/-2	0/0	**7**/7		
W	-2/-3	-5/-2	-2/-4	-6/-3	6/4	-4/-4	-2/-2	-1/0	-8/-5	4/-1	0/0	-1/-5	-3/-4	-6/-1	-5/-3	-20/-3	-3/-4	-3/-2	**16**/15	
Y	-2/-2	-11/-4	-5/-4	-6/-3	6/5	-5/-4	2/1	-3/-1	-5/-2	1/0	3/0	-4/-3	-4/-3	-1/-2	-1/-2	-10/-2	-3/-3	-2/-1	3/3	**12**/10

Scores for PPII aer in bold. Numbers represent log-odds scores for substitution frequencies (see Materials and Methods section). Cells where the score for PPII is higher than that for coils by 2 or more units are highlighted in blue; those where it is lower by 2 or more units are highlighted in green.

from the fractional proline content using the assumption that the probability of proline occurring at one position is independent of its occurrence elsewhere in the PPII helix. Thus, the expected number of PPII helices containing n prolines $C(n)$ is given by a binomial distribution:

$$C(n) = \Sigma_i^N \frac{L_i!}{n!(L_i - n)!} (f_{PRO})^n (1 - f_{PRO})^{L-n},$$

where N is the total number of PPII helices (5700), L_i is the length of PPII helix i, and f_{PRO} is the fractional content of proline in PPII helices. The summation is over all PPII helices and the elements of the summation are simply the fraction or probability that each PPII helix i of length L_i has exactly n prolines.

Substitution Tables

Environment-specific amino acid substitution tables were calculated using SUBST version 1.13.[21] Substitution probabilities are defined as follows: Each column (j) in Table II represents the probability distribution for the likelihood of acceptance of a mutational event by a residue type j in a particular structural environment, in our case, PPII helix assigned by SEGNO, leading to any other residue type (i). Since this table is the constrained type, residues j and i are in the same environment (if it were the unconstrained type, i could be in any environment).

$P(j(\text{PPII}),i)/\{P[j(\text{PPII})]\ ^*P[i(\text{PPII})]\}$ was calculated, with $P[j(\text{PPII}),i]$ being the probability of finding j in the PPII helix aligned against i, $P[j(\text{PPII})]$ being that of finding j in the PPII helix, and $P[i(\text{PPII})]$ being that of finding i in the PPII helix. Log odds ratios were multiplied by 3/log(2) and rounded to the nearest integer (i.e., the log-odds scores were expressed in "1/3 bit units").

Structural Analysis

The Ramachandran plot for residues in β-sheets, isolated β-strands, and PPII helices were contoured using kin3Dcont[22] and displayed using MAGE version 5.73.[23,24]

Residue accessibilities were calculated using PSA version 2.0 (L. Chen, unpublished), which uses the rolling probe algorithm.[25]

The stabilizing role of hydrogen bonds for PPII helices was assessed with the program PROBE version 2.1.6[26] after addition of all hydrogen atoms by REDUCE version 2.13.2.[27] Asparagine, glutamine, and histidine residues were "flipped" (i.e., the positions of side-chain hydrogen-bond donors and acceptors were exchanged) by REDUCE where this was required to improve the hydrogen-bonding network. Atoms were counted as being hydrogen-bonded by PROBE if they were a hydrogen-bond donor–acceptor pair and their van der Waals shells overlapped.

Protein models were produced with MODELLER,[28] and structures were superimposed with WHATIF.[29]

RESULTS AND DISCUSSION

Characterization of PPII

The HOMSTRAD database of structurally aligned homologous proteins[19,20] (www-cryst.bioc.cam.ac.uk/~homstrad/) was used. At the time of use, the database contained 3268 protein structures in 978 families. A subset of the database was also used: This contained only those structures solved at 1.8 Å or better and consisted of 337 structures. Where this subset was used, we have indicated so in the text.

The database was searched, and 5700 PPII helices of 3 residues or longer were found. We chose this length threshold because SEGNO, as well as DSSP[1] and STRIDE,[2] assign other extended conformations, such as β-strands, as short as 3 residues. The PPII helices represent 20,192 residues of the 685,372 in the data set (3%). Most PPII helices are quite short, with 3759 being 3 residues long. The percentage of amino acids in PPII conformation drops to 1.3% of those in the data set when only PPII helices longer than 3 residues are considered. Although PPII helices represent a small fraction of the total residues, 1965 (60%) of the protein chains contain one or more PPII helices, and 813 (83%) families contain at least one PPII helix. Despite this generally short length, 13 PPII helices were found to be longer than 9 residues. The distribution of length is shown in Table I.

Although proline is abundant, PPII helices exist that do not contain Pro. There are slightly fewer than the expected values from the fractional content of proline in PPII helices.

Ramachandran Plot

Both PPII helices and β-strands have extended conformations. The majority of β-strands associate to make β-sheets that allow their main-chain hydrogen-bond donors and acceptors to be satisfied. However, a small number, rather than taking part in a sheet, appear to be isolated. These isolated β-sheets are analogous to PPII helices. PPII helices cannot form sheets or other oligomerizations with neighboring PPII helices, because hydrogen-bond donors and acceptors do not line up in a complementary manner, and side-chain interactions sterically hinder oligomerization. Since many secondary structure assignment programs, such as DSSP[1] and STRIDE,[2] use hydrogen-bonding patterns to assign secondary structure, isolated β-strands and PPII helices are assigned as loop or coil regions. SEGNO, with its geometric definition of secondary structure, is able to assign these conformations correctly. SEGNO assigns β-strands first on geometric features such as ϕ and ψ angle values, then distinguishes between strands in sheets and those that are isolated based on the occurrence of parallel–antiparallel elements and hydrogen bonds. ϕ and ψ for β-strands fall in a broad range, which in part includes values compatible with PPII conformation: According to SEGNO, PPII conformations must have dihedrals in the range of $-80° \pm 45°$ (ϕ) and $142.5° \pm 42.5°$ (ψ), whereas β-strands must have dihedrals in the range of $-130° \pm 60°$ (ϕ) and $122.5° \pm 62.5°$ (ψ). Therefore, we consider whether the population of PPII helices and β-strands (in a sheet or isolated) is distinct or represents a continuum of conformations. For this purpose, the Ramachandran plots for isolated β-strands and PPII helices were calculated from the high-resolution subset (1.8 Å or better) of HOMSTRAD.

In Figure 1, the maximum for PPII occurrence is 65° (ϕ), 145° (ψ), clearly distinct from the maximum for β-strands in sheet at $-120°$ (ϕ), 130° (ψ). The contour for isolated β-strands is more complex, with maxima at $-70°$ (ϕ), 140° (ψ), which may represent a population more similar to PPII helices, and at $-135°$ (ϕ), 150° (ψ), which might represent a population similar to the strands in sheet. A secondary maximum is observed in a tail at $-85°$ (ϕ), 70° (ψ).

Amino Acid Propensities

Table III shows the propensities for amino acids in PPII helices containing 3 residues or more.

Since the values for the two groups do not differ significantly (correlation coefficient $R^2 = 0.97$) and scarcity of data could cause artifacts in the analysis, PPII helices equal to or longer than 3 residues were considered together. Pro is greatly favored in PPII helices, whereas Gly and Tyr are disfavored. Qualitatively, these findings are in accordance with those observed by Stapley and Creamer,[18] and the discrepancies between some values probably reflect the larger data set used in this study (5700 PPII helices) as compared with the earlier one (272 helices).

It is worth noting that although Gly disfavors the PPII helix, it is common in collagen. Collagen consists of 3 polypeptide chains, each in a PPII helix. These 3 helices in turn are supercoiled around each other. This supercoil is right-handed, and there is a repeating pattern of H-bonds between them. Close packing of the chains near the central axis imposes the requirement that Gly occupy every third position. Breaking of the repeating X-Y-Gly pattern by a Gly→Ala substitution results in a subtle alteration of the conformation, with a local untwisting of the triple helix,[30] with water molecules intervening in the H-bonds.

In host–guest studies,[14,15] insertion of an intervening Gly into a poly(proline)-based host peptide did not appear to disrupt the PPII helix in solution any more than did other apolar amino acids. Moreover, oligomerization was

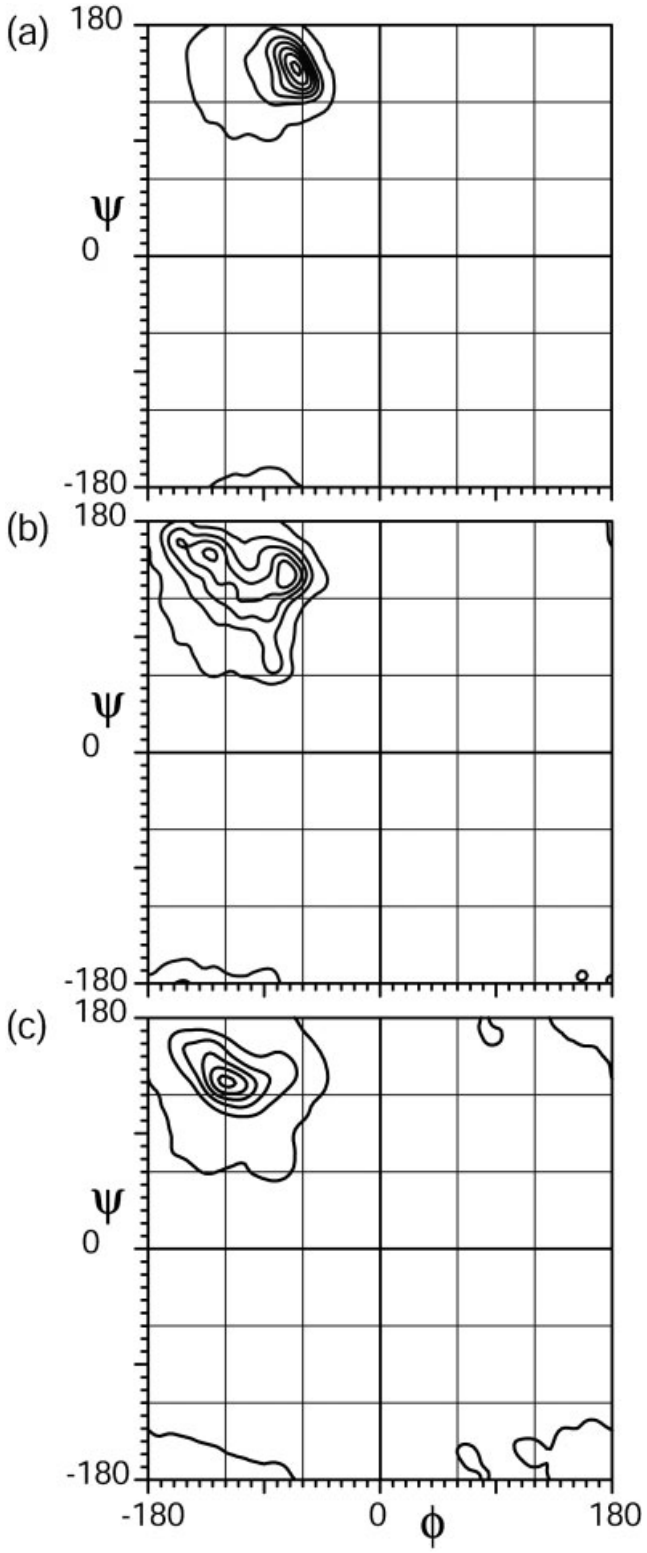

Fig. 1. The data are contoured to show the density of points for (**a**) PPII helices, (**b**) isolated β-strands, and (**c**) strands in sheets.

not induced. In contrast, using simple Monte Carlo computer simulations employing the hard sphere potential,[31] glycine residues inserted as guests into polyproline host peptides are conformationally not restricted by proline residues and do not tend to be part of the PPII helix. It seems, then, that although Gly is destabilizing for PPII, this effect is more than compensated for by hydrogen-bond formation in the backbone of collagen.

TABLE III. Amino Acid Propensities in Polyproline II Helices

	Length = 3		Length > 3	
P	**3.84***	2064	**4.15***	1722
I	**1.17**	506	**1.08**	362
M	**1.06**	260	**0.81**	153
K	**1.05**	738	**1.04**	561
T	**1.00**	669	**0.98**	506
L	**0.95**	952	**0.98**	761
S	**0.92**	645	**0.95**	513
Q	**0.91**	394	**1.04**	348
C	**0.90**	169	**1.05**	151
E	**0.90**	666	**0.98**	559
R	**0.86**	483	**1.11**	481
A	**0.86**	828	**0.90**	664
V	**0.80**	664	**0.85**	545
H	**0.76**	205	**0.68**	141
F	**0.73**	351	**0.66**	245
N	**0.73**	392	**0.65**	272
W	**0.72**	118	**0.59**	75
D	**0.71**	478	**0.76**	407
G	**0.45***	409	**0.35***	246
Y	**0.44***	286	**0.40***	203

Propensities (bold) and number of occurrences are shown. Propensities indicated by an asterisk are significant according to a T test ($P < 0.01$).

A position-specific analysis of amino acid propensities is shown in Table IV. The occurrence of each amino acid is reported and propensities are in bold. Here residue preferences at each position in the PPII helix and in the proximal regions are calculated, and trends can be recognized. Positions were labeled so that for a 3-residue PPII helix, $n1$ would be the first position, mid the second, and $c1$ the third, whereas, as an example, for a 7-residue PPII helix, $n1$ would be the first position, $n2$ would be the second, mid would group positions between the second and the penultimate, $c2$ would be the penultimate, and $c1$ the last. $n-1$, $n-2$, $n-3$ $c+1$, $c+2$, and $c+3$, indicate in any case the 3 positions preceding and following a PPII strand.

Pro is favored, whereas Gly and Tyr are disfavored at all positions. Gly becomes the preferred amino acid immediately before and after PII helices, at position $n+1$,$c+1$.

The low propensity of aromatic amino acids for PPII seems to contrast with the finding that Tyr and Phe are in the consensus sequences of the peptide ligands of class I WW domains and EVH1 domains, respectively.[32] YAP65[33] and dystrofin,[34] both belonging to class I WW domains, have been crystallized in the presence of peptide ligands; the Tyr residue of the consensus is assigned by SEGNO as the last residue of a helix in PPII conformation. Mena[35] and Homer,[36] which are both EVH1 domains, have been crystallized in the presence of peptide ligands, showing the Phe residue of the consensus target sequence adjacent to the PPII helix, but outside its boundaries assigned by

TABLE IV. Position-Dependent Propensities for Polyproline II Helices

	$n-3$		$n-2$		$n-1$		$n1$		$n2$		mid		$c2$		$c1$		$c+1$		$c+2$		$c+3$	
P	**1.45**	368	**1.49**	380	**0.78**	206	**2.05***	538	**4.85***	433	**5.16***	1166	**4.71***	421	**4.68***	1228	**2.33***	610	**0.72**	185	**1.10**	282
A	**0.85**	384	**0.86**	392	**0.67**	315	**0.93**	435	**0.92**	147	**0.85**	341	**1.01**	161	**0.87**	408	**0.79**	370	**0.92**	425	**0.82**	375
M	**0.68**	79	**0.78**	91	**0.93**	112	**1.12**	134	**0.49**	20	**1.30**	134	**1.32**	54	**0.59**	71	**0.60**	72	**0.64**	75	**0.68**	79
L	**0.99**	469	**0.71**	339	**0.69**	340	**1.13**	554	**0.81**	136	**1.03**	437	**1.32**	221	**0.74**	365	**0.69**	337	**0.63**	302	**0.78**	375
V	**0.95**	371	**0.77**	303	**0.58**	234	**1.09**	443	**0.88**	122	**0.85**	296	**0.96**	133	**0.53**	215	**0.56**	226	**0.77**	305	**0.95**	375
I	**1.24**	252	**1.15**	237	**0.92**	195	**1.49**	314	**1.14**	82	**1.21**	220	**1.22**	88	**0.78**	164	**0.83**	175	**0.98**	203	**1.27**	261
C	**0.93**	82	**1.02**	91	**0.60**	55	**1.21**	110	**1.16**	36	**1.02**	80	**0.93**	29	**0.71**	65	**0.72**	66	**0.85**	76	**0.85**	76
K	**0.94**	312	**0.94**	312	**0.95**	325	**1.14**	389	**1.35**	157	**1.26**	372	**0.93**	108	**0.80**	273	**0.80**	273	**0.78**	262	**0.77**	257
R	**1.07**	283	**0.94**	252	**1.07**	292	**1.25**	342	**1.20**	112	**0.87**	206	**0.91**	85	**0.80**	219	**0.78**	212	**0.74**	199	**0.90**	241
Q	**1.06**	215	**1.01**	208	**0.91**	192	**1.16**	244	**1.06**	76	**1.10**	200	**0.92**	66	**0.74**	156	**0.77**	162	**1.16**	241	**1.14**	234
E	**0.88**	306	**0.87**	304	**0.84**	303	**1.00**	361	**1.28**	157	**0.88**	273	**0.93**	114	**0.89**	320	**0.98**	353	**1.48**	522	**1.35**	473
T	**0.86**	271	**1.20**	380	**1.03**	335	**1.21**	393	**0.80**	89	**0.86**	243	**0.68**	75	**1.15**	375	**0.93**	303	**1.36**	436	**1.25**	396
S	**1.29**	426	**1.18**	392	**1.05**	359	**0.84**	286	**0.68**	79	**0.66**	195	**0.57**	66	**1.55**	532	**1.05**	359	**1.55**	523	**1.08**	359
D	**1.12**	371	**1.07**	357	**1.02**	350	**0.54**	184	**0.63**	74	**0.57**	169	**0.59**	69	**1.14**	389	**1.03**	351	**1.40**	473	**1.22**	406
N	**1.13**	286	**1.47**	375	**1.24**	326	**0.72**	190	**0.34***	30	**0.54**	121	**0.73**	65	**0.98**	258	**1.06**	279	**1.16**	300	**0.97**	249
H	**0.96**	121	**1.10**	141	**0.97**	128	**0.88**	116	**0.54**	24	**0.66**	75	**0.74**	33	**0.75**	98	**0.62**	81	**1.13**	146	**0.96**	123
W	**0.91**	70	**1.09**	85	**0.80**	64	**0.84**	67	**0.51**	14	**0.77**	53	**0.96**	26	**0.41**	33	**0.65**	52	**1.18**	93	**1.37**	107
F	**0.91**	205	**0.87**	198	**0.90**	211	**1.07**	251	**0.59**	47	**0.56**	113	**0.60**	48	**0.59**	137	**0.72**	169	**0.93**	215	**0.96**	220
Y	**0.73**	226	**0.56**	173	**0.63**	202	**0.53**	168	**0.48**	52	**0.34***	93	**0.32**	35	**0.44***	141	**0.40***	128	**0.71**	224	**0.72**	224
G	**0.95**	409	**1.26**	547	**2.61***	1164	**0.41**	181	**0.36***	54	**0.32***	123	**0.29***	44	**0.57***	253	**2.52***	1118	**0.93**	407	**1.04**	450
Total		5470		5557		5708		5700		1941		4910		1941		5700		5696		5612		5562

Propensities (bold) and numbers of occurrences. Propensities indicated by an asterisk are significantly significant according to a T test ($P < 0.01$).

SEGNO. These flanking aromatic residues in the peptide ligand make extensive hydrophobic contact with the respective receptor. Given the rarity of aromatic residues following PPII helices (Table IV), it is possible that they are in part responsible for specificity.

Both Asn and Asp are disfavored except at the C-terminal position. In contrast, Glu, Gln, and aliphatic hydrophobic amino acids (ILMCV) are less favorable at the C-termini than at the N-termini.

Substitution Tables

To build substitution tables, we first identified residues in PPII helices using the program SEGNO and then searched HOMSTRAD. Each amino acid was assigned a local structural environment, and for each position in the multiple alignment, substitutions were counted. Tables were calculated using Subst.[21]

The environment-constrained substitution table for PPII defined by SEGNO was compared to the constrained substitution table calculated for coils defined by DSSP.[1] Since DSSP does not assign PPII helices, residues that SEGNO assigns to this class would be merged with coil residues. A direct comparison is possible since, in Table II, scores for both environments are reported. Scores for PPII are in bold.

The constrained tables derived for PPII helices (SEGNO) and coils (DSSP) are clearly different: The correlation coefficient $R^2 = 0.64$ should be compared with $R^2 = 0.61$ between tables for coil and α-helix (DSSP assignment), and with $R^2 = 0.82$ between tables for coil and β-strand (DSSP assignment). To highlight this difference, in Table II, cells where the score for PPII is higher than that for coils by 2 units or more (i.e., residues in PPII are more conserved than in coil) are highlighted in blue; those where it is lower by 2 units or more (residues in PPII are less conserved than in coil) are highlighted in green.

This analysis of the substitution propensities suggests that PPII should not be merged with coil residues, and PPII helices should be considered an environment different from coil as β-strands and α-helices. It also suggests that the use of the PPII conformation as a separate environment is likely to aid fold recognition, comparative modeling, and other applications where environment-specific substitution tables are used, for example, in FUGUE[21] and CODA.[37]

The matrix element A(ij) in Table II is positive if the amino acid j is substituted more frequently by the amino acid i than expected from the background frequency of occurrence (see Materials and Methods section), whereas it is negative if it is substituted less frequently. If j and i are amino acids with similar physical properties, then it is likely that they will frequently substitute for each other, and A(ij) will be positive. In Table II, we observe that positive elements (common substitutions) are mainly in blue (PPII residues more conserved), and negative elements (uncommon substitutions) are mainly in green (PPII residues less conserved). This indicates a higher conservation of similar amino acids in PPII than in coils. Values on the diagonal line indicate how well a specific amino acid is conserved. These values are predominantly blue, meaning that they are more conserved in PPII than they are in coil. A noticeable exception is Pro, which although highly conserved in PPII, has an even higher degree of conservation in coil. This is probably because a proline in a PPII helix has the same stabilizing effect wherever it is positioned, whereas a proline in a loop

region is often critical for stabilizing a β-turn, and its role is position-sensitive.

Examination of the Stabilizing Interactions in PPII Helices

To assess the extent to which hydrogen bonding contributes to the stability of PPII helices, we looked for partners of backbone H-bond donors (NH) and acceptors (CO). We used the program PROBE[26] and the subset of the structures in the HOMSTRAD database with resolution equal to or better than 1.8 Å. In this subset, 261 structures out of 337 contained at least one PPII helix. The number of residues in this subset of well-resolved PPII helices was 3862, 780 of which were Pro and so have no donor NH. Potential donors unsatisfied were 612, whereas potential acceptors unsatisfied were 906.

In order to check whether intermolecular interaction had been greatly underestimated, crystallographic symmetry mates were calculated for a subset of the structures analyzed for hydrogen bonds in PPII helices. Less than 2% unsatisfied donors or acceptor found their partners once symmetry-related molecules were generated.

It has been hypothesized that backbone solvation is a major driving force in the formation of PPII helices by nonproline residues; Adzhubei and Sternberg[17] and Stapley and Creamer[18] found that residues in known protein structures were significantly more accessible to solvent than average. In this analysis, unsatisfied potential donors and acceptors in PPII helices can probably be ascribed to water molecules not identified by crystallographers. Buried residues (<7% exposed area, when normalized against an isolated residue that is counted as 100% exposed) were identified among those in PPII helices with unsatisfied main-chain donors or acceptors. Their percentages, 19.1% and 18.4%, respectively, should be compared to those measured for buried residues in coils with unsatisfied main-chain donors and acceptors, 12.3% and 11.8%, respectively.

Table V represents a summary of the groups satisfying potential donors or acceptors in PPII backbone.

In some cases, PROBE finds two possible partners for each donor or acceptor. The number of water molecules that bind CO groups in PPII backbone is much higher than the number of water molecules that bind NH groups, whereas main-chain–main-chain interactions are comparable.

That many more water molecules are hydrogen-bonded to carbonyls in PPII helices than to NH groups in PPII helices would suggest that backbone carbonyls point out from the body of the protein, ready to interact with ligands or solvent. Moreover, it would suggest that PPIIs are mainly anchored to the body of the protein through their main-chain NH. We observed that all Ns of prolines in PPII are buried, whereas 61.4% and 50.1% of main-chain potential donors and acceptors are buried, respectively.

Side-chain to main-chain interactions can be either local (from within the same PPII helix or its close proximities) or nonlocal. In Table V, we show the proportion of interactions between residue *i* in a PPII helix and a residue within $i \pm 3$, whereas the latter (in bold) represents total interactions. It is worth noting that side-chain interactions with CO groups of the PPII backbone are mainly nonlocal (only 18% local interactions), with Ser and Thr representing an exception. A higher proportion of local interactions is observed between NH groups of PPII backbone (51% local interactions) and various side-chains such as those of Gln, Glu, Asp, Ser, and Thr. Gln and Glu local interactions are mainly between Glx *i* OE→res $i - 3$ (NH in PPII). Many interactions are also observed between Glu *i* OE→Glu *i* (NH in PPII).

TABLE V. Hydrogen Bonds Stabilizing PPII Main-Chain Donors and Acceptors

	Acceptor satisfying	Donor satisfying
HOH	**2158**	**991**
CO mc	—	**952**
NH mc	**975**	—
ASN sc	5/**78**	12/**63**
ASP sc	1/**1**	78/**125**
ARG sc	22/**258**	—
CYS sc	2/**4**	2/**6**
GLN sc	9/**81**	40/**75**
GLU sc	—	57/**101**
HIS sc	14/**34**	8/**15**
LYS sc	13/**50**	—
MET sc	—	5/**7**
SER sc	36/**64**	25/**37**
THR sc	31/**58**	34/**53**
TRP sc	0/**31**	—
TYR sc	0/**56**	0/**31**
HET	**3**	**16**

Total figures (bold) and figures referring to contacts between residue *i* in a PPII helix and residues within $i \pm 3$; "mc" indicates main-chain, and "sc," side-chain.

The relatively high proportion of main-chain to main-chain hydrogen bonds and nonlocal interactions between side-chains and main-chain donors and acceptors in PPII helices demonstrate that PPII helices can often be stabilized as a result of tertiary interactions rather than as a result of intrinsic propensities of amino acids. This is likely to be the case for PPII helices that have no prolines, as proline is the only residue with a significant propensity for this conformation. The potential for identifying PPII helices from sequence patterns will probably be limited by false-negative results.

The special case of stabilizing hydrogen bonds between the side-chain of residue *i* and the main-chain of residue $i + 1$ is worth considering in detail. It was proposed that Gln has a relatively high propensity for PPII for the ability of its side-chain to be hydrogen-bonded to the main-chain carbonyl of the following amino acid. Stapley and Creamer[18] give 4 examples of this stabilizing effect of Gln from *Rhizomucor miehei* tryacyl-glycerol lipase [Protein Data Bank (PDB) code: 3tg1), from *Clostridium* molybdenum-iron protein (PDB code: 1mio), from bacteriophage PHIX174 capsid protein (PDB code: 2bpa), and from bovine carboxypeptidase A (PDB code: 2ctc). In our analysis, we found only one case of a hydrogen bond between Gln

side-chain NE2 and the carbonyl of the following residue in PPII conformation, namely, that observed in bovine carboxypeptidase A. The other structures, with PDB codes 3tg1, 1mio, and 2bpa, did not belong to the set analyzed with PROBE, since they were not solved at a resolution of 1.8 Å or better. Even if we do not observe the hydrogen bond between Gln i OE and res $i + 1$ CO interaction, we do observe a high propensity of Gln in the first positions of PPII helices (Table IV).

Positively charged amino acids, which are frequently seen in the consensus sequences for SH3 ligands,[32] have a relatively high propensity to occur at the beginning of PPII helices and interact, more frequently than Gln, with the carbonyl of residue $i + 1$. The frequencies are as follows: 8 times out of 22 local hydrogen bonds for Arg, and 5 times out of 13 local hydrogen bonds for Lys. Moreover, Asp interacts with NH of res $i + 1$ in 20 out of 78 local hydrogen bonds, although it has not been recognized commonly in the consensus sequences for polyproline-binding proteins. Positively charged residues both local and nonlocal to the PPII helices stabilize the PPII conformation by satisfying the H-bond donor potential of the main-chain carbonyl. Arginine in particular tends to make H-bonds to the main-chain carbonyls that are partly buried, stabilizing the conformation of both the PPII helix and the Arg residue.

We agree with Stapely and Creamer,[18] who claim that still too little is known on PPII conformation: The availability of larger sets of structures, especially those solved at high resolution, can modify our understanding of PPII stabilization. The role of glutamine appears now to be reduced, but still too few examples of other local side-chain–main-chain interactions are found to propose other candidates.

Consensus Sequences and Conservation of PPII

Of PPII helices identified in 978 HOMSTRAD families, 5700 were searched for the 24 consensus sequences that bind to specific protein interaction modules.[32] Only the consensus PPxY for class I WW domains was found in the other families: for β-*N*-acetylhexosaminidase from *Streptomyces plicatus* (PDB code: 1hp5) in the family of glycosyl hydrolase 20 catalytic domain family, and for DNA topoisomerase I (N-terminal fragment) from *Streptomyces cerevisiae* (PDB code: 1ois) in the Eukaryotic DNA topoisomerase family.

Adzhubei and Sternberg[17] found that the most important PPII segments were conserved when they analyzed 3 protein families: serine proteinases, aspartic proteinases, and immunoglobulin constant domains. They suggested that PPII helices, in addition to the other 2 secondary structure classes, should be identified as part of structurally conserved regions in protein. They also demonstrated the improvement in model building in 2 test studies. We extended their analysis and looked for conserved PPII helices in the 11 HOMSTRAD families that contain at least 1 member with a PPII helix longer than 9 residues.

The longest segment belongs to the benzoylformate decarboxylase (BFD) from *Pseudomonas putida* (PDB code: 1bfd) in the family of thiamine pyrophosphate enzymes. The same family includes pyruvate decarboxylase (PD) isozyme 1 from *S. cerevisiae* (PDB code: 1pvd), pyruvate oxidase from *Lactobacillus plantarum* (PDB code: 1pox), pyruvate decarboxylase from *Zymomonas mobilis* (PDB code: 1zpd). Figure 2 shows the alignment from HOMSTRAD with β-sheets, isolated β-strands, and PPII helices highlighted. Four out of 6 possible PPII helices longer than 3 residues are conserved in at least 2 structures of the family. The longest PPII helix is conserved in the structures with PDB codes 1bfd and 1bvd. It is worth noting that the HOMSTRAD alignment, which was built with no consideration of the occurrence of PPII helices, shows a gap in the sequence of 1bfd right in correspondence with the long PPII helix.

We modeled BFD from *P. putida* using as a template the structure with PDB code 1bvd. We aimed to identify the difference in the model if we used the alignment stored in HOMSTRAD (that shown in Fig. 2) or a hand-modified alignment in which the gap in correspondence with the long PPII helix of the template 1bvd is avoided. Differences in the alignments are shown in Figure 3; the region where they occur is bold in Figure 2. We produced two homology models, 1bfd_homs (HOMSTRAD alignment) and 1bfd_hand (hand-modified alignment), and the models were superimposed on the crystal structure with PDB code 1bfd.

The overall root-mean-square deviation (RMSD) of 1bfd_hand from 1bfd in the PDB was slightly better than the RMSD of 1bfd_homs from 1bfd in the PDB. Then we superimposed the 14-residue-long peptide RQLPTAAPEPAKVD of each model on the same peptide in the experimental structure: The modified alignment, which takes into account the occurrence of a PPII helix in the template and accordingly disfavors gaps, produces locally a better model (RMSD 1.253) than the homstrad alignment (RMSD 2.664).

Figure 3 shows the HOMSTRAD and manually modified local alignment (this being the only alteration of the alignment shown in Fig. 2) and the superimposed peptides.

In the experimentally determined structure (PDB code: 1bfd), the long PPII helix mediates the main contact between tetramers in the extremely well-ordered, stable crystals described by Hasson et al.[38] The same authors suggest that since, in vivo, BFD is part of a multienzyme complex that includes other members of the mandelate pathway,[39] the long, regular PPII helix, exposed on the surface of the structure 1bfd, could be a convenient "handle" by which BFD in held in the complex.

Protein–Protein Interactions

Given that stabilizing hydrogen bonds in PPII conformations are nonlocal, and that PPII conformations are usually highly solvent-exposed,[18] it is possible that PPII helices are involved in protein–protein interactions.

Many protein interactions occur via proline-rich sequences that in turn are assumed to have high propensity to form PPII. For example, proline-rich sequences are the

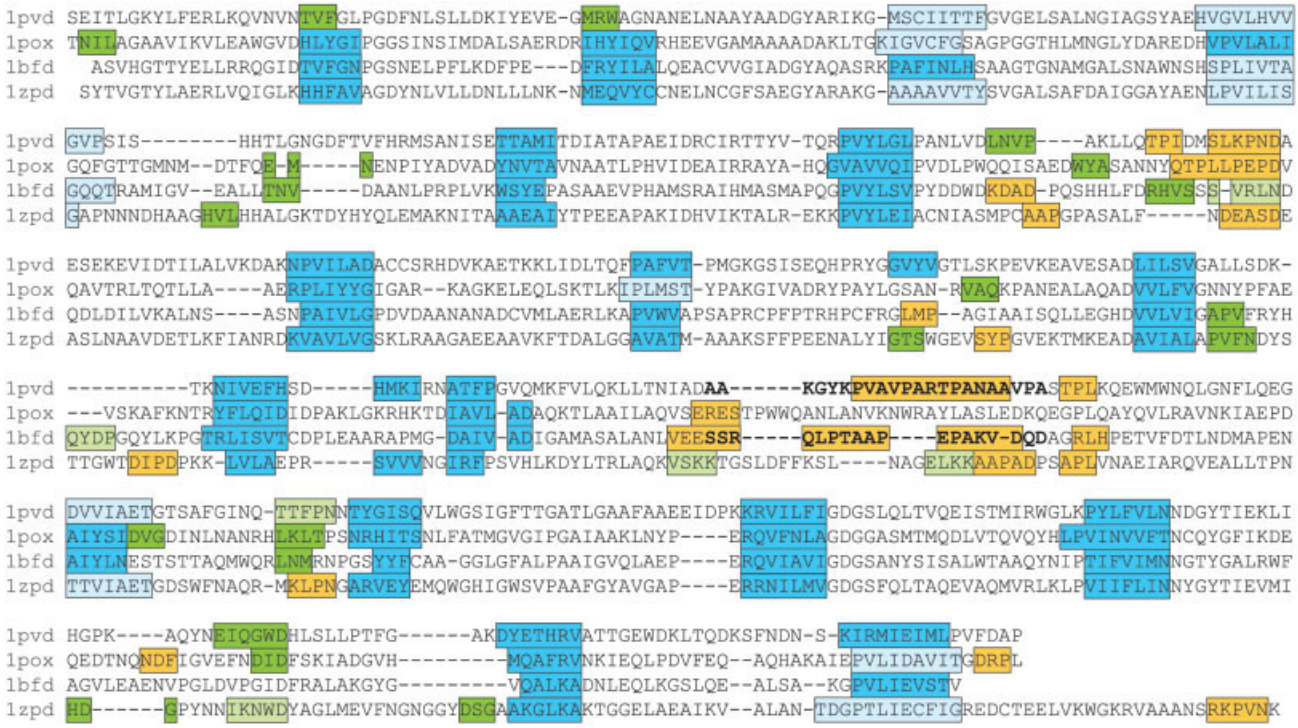

Fig. 2. Alignment for the family of thiamine pyrophosphate enzymes from the HOMSTRAD database. Colors indicate secondary structures: β-strands in sheets (dark blue), β-strands in sheets flanked by PPII, isolated β-strands (dark green), isolated β-strands flanked by PPII (pale green), and PPII helices (yellow). Bold text indicates the region used for modeling in Figure 3. Secondary structures were assigned with SEGNO.

ligands of SH3 (Src homology 3),[40,41] WW (named for a conserved Trp-Trp motif),[42] Enabled/VASP homology (EVH1, also known as WASP homology 1 or WH1),[43] and GYF domains,[44] as well as of profilin proteins.[45]

To perform a systematic search for interactions of PPII containing proteins, we used BINDBlast, a local instance of the BLAST tool adapted to the Biomolecular Interaction Network Database (BIND).[46] BIND contains information about interactions from the literature, interaction data from the PDB, and information derived from a range of large-scale interaction mapping experiments. Interactions can be submitted by users. BIND inevitably represents the same biases toward model organisms as the literature.

Initially, the sequences of BFD and PD were used as queries, with the aim of identifying interactions for molecules similar in sequence to the query sequences. Blastp was run without low-complexity filters and with *E*-value threshold 0.0001.

Results are summarized in Table VI, including results for the other HOMSTRAD families that contain at least one member with a PPII helix longer than 9 residues. We report the number of PPII helices longer than 3 residues that are conserved in at least 2 members of each family. The alignment from which these data were obtained is not shown, but as in the case of the alignment of thiamine pyrophosphate enzymes in Figure 2, generally conserved PPII helices are not the same length; sometimes very short PPII helices substitute long ones. The longest PPII helices tend to be the best conserved.

In the third column of Table VI, we report the names of the homologous proteins found by BINDBlast. In the fourth is the number of interactions in BIND for the homologous proteins. All the homologous proteins found belong to *S. cerevisiae*; the significance of this is unclear, although it is likely to be due to the large number of *S. cerevisiae* interactions in the database from the large-scale yeast interaction mapping projects.[47]

Typically, the dissociation constant (Kd) values of synthetic peptides binding WW domains are in the high nanomole to low micromole range: those for binding SH3 domains are 1–100 μ*M*, but the Kd values of full-length proteins that interact with SH3 domains are significantly lower than for peptides.[48] Although proline-rich peptide ligands might have a degree of flexibility in solution, the loss of entropy is small.[49] This is probably due to the preformation of the PPII conformation in highly solvated structures,[7,9–13] with a relatively small degree of extra ordering upon binding. The origin of the difference between Kd values between peptide and protein ligands is unclear.

Although in some cases it could be important for function to have a weak interaction, PPII in proteins could produce stronger binding. Many experimental techniques are usefully employed to identify interacting proteins but cannot define the binding sites. If one of the partners contains a PPII helix, a potential site could be suggested for further investigations. Since proline-rich regions are the most common sequence motif in the *Drosophila* ge-

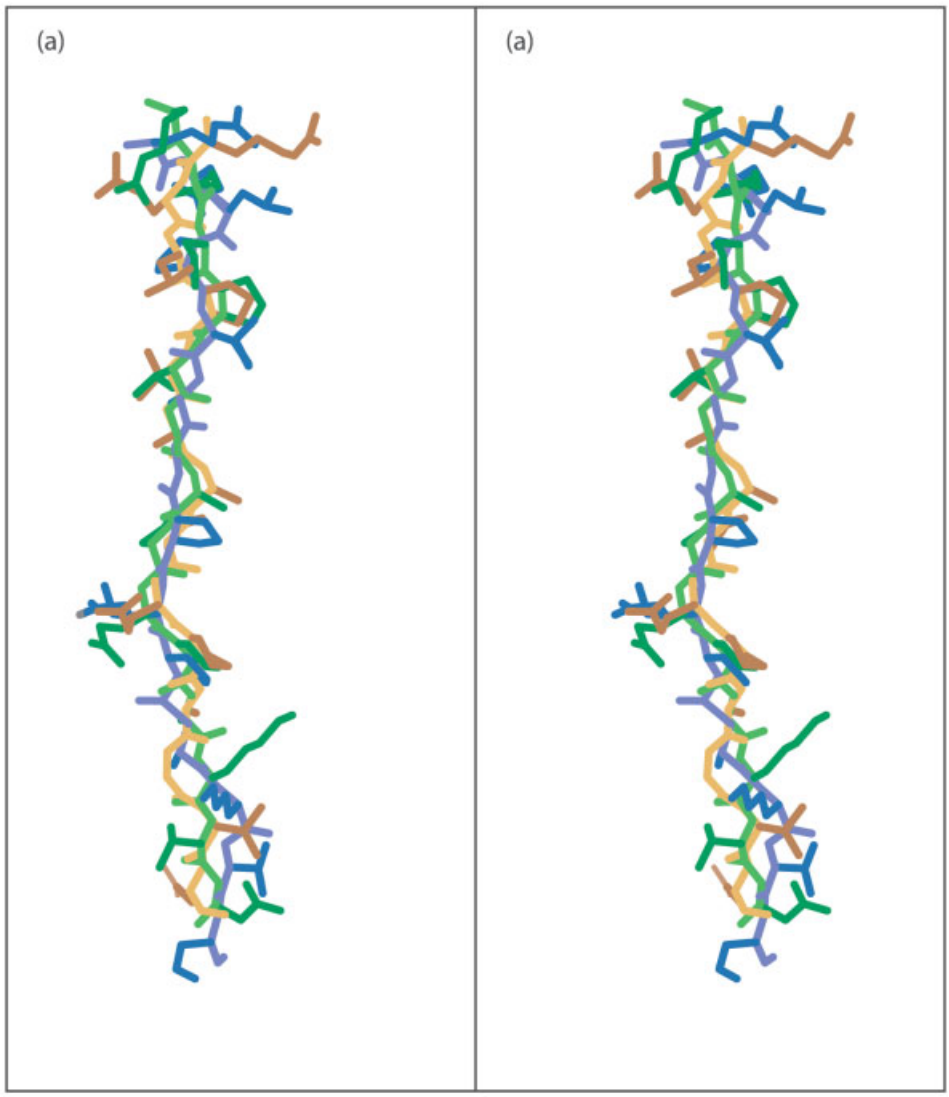

Fig. 3. (**a**) The structure of res 335–347 in PDB file 1bfd is in yellow; the structures of the same residues in the models obtained with 1pvd as a template are in different colors: blue for the model obtained using the alignment stored in HOMSTRAD, and green for the model obtained using the hand modified alignment. (**b**) Partial alignment of benzoylformate decarboxylase (BFD) from *P. putida* to pyruvate decarboxylase isozyme 1 from *S. cerevisiae* (PDB code: 1pvd): 1BFD_homs is the alignment stored in HOMSTRAD; 1BFD_hand is the hand-modified alignment; boxed in yellow are the residues in PPII conformation in 1pvd.

nome and the second most common in the *Caenorhabditis elegans* genome,[50] it is possible that any new protein–protein interactions could be discovered by this method.

CONCLUSIONS

With the advent of combinatorial peptide libraries, ligand specificity has been determined rapidly for a large number of domains that bind proline-rich sequences. Peptides may adopt PPII conformation in solution[7]; those with certain consensus sequences may adopt PPII conformation more readily.[14–16] As a result of binding, more entropy is lost[49] and the final defined PPII conformation is formed. In a folded protein, a PPII conformation can be stable even in the absence of a consensus sequence; the amino acid, with increased propensity, in PPII helices in folded proteins is Pro. Presumably this is because Pro both has allowed ϕ and ψ angles that are consistent with PPII helices, and because it lacks the NH hydrogen-bond donor potential, and thus is well suited to a conformation where hydrogen bonds are difficult to make due to steric considerations.

TABLE VI. Occurrence of Long Polyproline II Helices in HOMSTRAD Families and Their Potential Involvement in Protein–Protein Interactions

Homstrad family	PPII helices	BINDBlast hits	BIND Interactions
Thiamine pyrophosphate enzymes	4/6	Acetolactate synthase; Ilv2p	3
		pyruvate decarboxylase; Pdc1p	9
		pyruvate decarboxylase; Pdc5p	11
		pyruvate decarboxylase; Pdc6p	8
		Hypothetical open reading frame; Ye1020cp	1
		Regulatory factor for thiamin metabolism; Thi3p	2
		Aro10p	2
Glycosyl hydrolase family 14 (β-amylase)	3/4		
Eukaryotic DNA topoisomerase I	4/4	Topoisomerase I; Top1p	9
DNA photolyase	2/3	Phr1p	1
Cyclophilin	2/2	Cyclophilin peptidyl-prolyl *cis–trans* isomerase; Cpr1p	23
		Cyclophilin-3 ;Cpr3p	1
		cyclophilin related to the mammalian CyP-40; Cpr6p	14
		Human cyclophilin B protein; Cpr2p	4
		Cyclophilin D, peptidyl-prolyl cis-trans isomerase D; Cpr5p	1
		Similar to secretory pathway cyclophilin Cpr4; Cpr8p	7
Peroxidase	6/6		
α β-hydrolase	4/4		
Transaldolase	0/1	Transaldolase; Tallp	5
Photosynthetic reaction center cytochrome C subunit 1	2/3		
Phosphoribulokinase/ uridine kinase	0/1		
Capsid protein (F protein)	2/5		

However, even Pro is not essential. Hydrogen bonds can also be satisfied with other, nonlocal parts of the main-chain, or other nonlocal side-chains. We find that local side-chain to main-chain interactions are not common, which in turn leads to lack of distinct side-chain rotamer preferences[18] and lack of distinct residue preferences. PPII helices tend, therefore, not to have clear consensus sequences and are not easily identified on this basis.

Is it desirable, then, to class PPII helices as either loop conformations, or with β-strands? We think not, as they have different physical characteristics (both Ramachandran propensities and typical degree of solvent exposure[18,51]) and, crucially, different substitution propensities as compared to both of these conformations. Additionally, not only are residues within PPII helices under unusual evolutionary restraint, the conformation itself is more conserved than loop conformations. This means that PPII helices, as a secondary structure type, are likely to be of use for comparative modeling and fold recognition. Penalties for the introduction of gaps and environment-specific substitution tables that take into account the existence of PPII can help align sequences when structural information is available for one of them.

Since the analysis of homologous protein families indicates that PPII helices are conserved, finding a good alignment with a protein that contains PPII helices would suggest that the PPII conformation is present in all members of the family. This is particularly useful given that there is no consensus sequence for PPII. In turn, a possible interaction with polyproline-binding domains may be suggested, with the PPII helix being the binding site. In PPII, both the side-chains and the backbone carbonyls point out from the helical axis into solution at regular intervals and are free to participate in intermolecular hydrogen bonds. Thus, as suggested by Zarrinpar et al.,[52] both side-chains and carbonyls can easily be "read" by interacting proteins. Thus, we hope to be able to predict the occurrence of PPII stretches in

proteins, and use this finding to gain functional information.

REFERENCES

1. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
2. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins 1995;23:566–579.
3. Blanch EW, Morozova-Roche LA, Cochran DA, Doig AJ, Hecht L, Barron LD. Is polyproline II helix the killer conformation?: a Raman optical activity study of teh amyloidogenic prefibrillar intermediate of human lysozyme. J Mol Biol 2000;301:553–563.
4. Eker F, Griebenow K, Schweitzer-Stenner R. Aβ1-28 fragment of the amyloid peptide predominantly adopts a polyproline II conformation in an acidic solution. Biochemistry 2004;43:6893–6898.
5. Hicks JM, Hsu VL. The extended left-handed helix: a simple nucleic acid-binding motif. Proteins 2004;55:330–338.
6. Tiffany ML, Krimm S. New chain conformations of poly(glutamic acid) and polylysine. Biopolymers 1968;6:1379–1382.
7. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA 2002;99:9190–9195.
8. Asher SA, Mikhonin AV, Bykov S. UV Raman demonstrates that alpha-helical polyalanine peptides melt to polyproline II conformation. J Am Chem Soc 2004;126:8433–8400.
9. Ferreon JC, Hilser VJ. The effect of the polyproline II (PPII) conformation on the denatured state entropy. Protein Sci 2003;12: 447–457.
10. Drozdov AN, Grossfield A, Pappu RV. Role of solvent in determining conformational preferences of alanine dipeptide in water. J Am Chem Soc 2004;126:2574–2581.
11. Kentsis A, Mezei M, Gindin T, Osman R. Unfolded state of polyalanine is a segmented polyproline II helix. Proteins 2004;55: 493–501.
12. Mezei M, Fleming PJ, Srinivasan R, Rose GD. Polyproline II helix is teh preferred conformation for unfolded polyalanine in water. Proteins 2004;55:502–507.
13. Sreerama N, Woody RW. Molecular dynamics simulations of polypeptide conformations in water: a comparison of alpha, beta and poly(Pro)II conformations. Proteins 1999;36:400–406.
14. Rucker AL, Pager CT, Campbell MN, Qualls JE, Creamer TP. Host-guest scale of left-handed polyproline II helix formation. Proteins 2003;53:68–73.
15. Kelly MA, Chellgren BW, Rucker AL, Troutman JM, Fried MG, Miller AF, Creamer TP. Host–guest study of left-handed polyproline II helix formation. Biochemistry 2001;40:14376–14383.
16. Chellgren BW, Creamer TP. Short sequences of non-proline residues can adopt the polyproline II helical conformation. Biochemistry 2004;43:5864–5869.
17. Adzhubei AA, Sternberg MJ. Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol 1993;229:472–493.
18. Stapley BJ, Creamer TP. A survey of left-handed polyproline II helices. Protein Sci 1999;8:587–595.
19. Mizuguchi K, Deane CM, Blundell TL, Overington JP. HOMSTRAD: a database of protein structure alignments for homologous families. Protein Sci 1998;7:2469–2471.
20. de Bakker PIW, Bateman A, Burke DF, Miguel RN, Mizuguchi K, Shi J, Shirai H, Blundell TL. HOMSTRAD: Adding sequence information structure-based alignments of homologous protein families. Bioinformatics 2001;17:748–749.
21. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence–structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310: 243–257.
22. Word JM. All-atom small-probe contact surface analysis: an information-rich description of molecular goodness-of-fit. Thesis, Duke University, Durham, NC; 2000.
23. Richardson DC, Richardson JS. The kinemage: a tool for scientific illustration. Protein Sci 1992;1:3–9.
24. Richardson DC, Richardson JS. MAGE, PROBE, and Kinemages. In: Rossmann MG, Arnold E, editors. International tables for crystallography. Vol. F. Dordrecht: Kluwer; 2001. p 727–730.
25. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. Science 1983;221:709–713.
26. Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness of fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol 1999;285:1711–1733.
27. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 1999;285:1735–1747.
28. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
29. Vriend G. WHATIF: a molecular modelling and drug design program. J Mol Graphics 1990;8:52–56.
30. Bella J, Eaton M, Brodsky B, Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 1994;266:75–81.
31. Creamer TP. Left-handed polyproline II helix formation is (very) locally driven. Proteins 1998;33:218–226.
32. Kay BK, Williamson MP, Sudol M. The importance of being proline:the interaction of proline: the interaction of proline-rich motifs in signalling proteins with their cognate domains. FASEB J 2000;14:231–241.
33. Pires JR, Taha-Nejad F, Toepert F, Ast T, Hoffmuller U, Schneider-Mergener J, Kuhne R, Macias MJ, Oschkinat H. Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPPYTVG, *N*-(*n*-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. J Mol Biol 2001;314:1147–1156.
34. Huang X, Poy F, Zhang R, Joachimiak A, Sudol M, Eck MJ. Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat Struct Biol 2000;7:634–638.
35. Fedorov AA, Fedorov E, Gertler F, Almo SC. Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. Nat Struct Biol 1999;6:661–665.
36. Beneken J, Tu JC, Xiao B, Nuriya M, Yuan JP, Worley PF, Leahy DJ. Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. Neuron 2000;26:143–154.
37. Deane CM, Blundell TL. CODA: A combined algorithm for predicting the structurally variable regions of protein models. Protein Sci 2001;10:599–612.
38. Hasson MS, Muscate A, McLeish MJ, Polovnikova LS, Gerlt JA, Kenyon GL, Petsko GA, Ringe D. The crystal structure of benzoylformate decarboxylase at 1.6 Å resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes. Biochemistry 1998;37:9918–9930.
39. Halpin RA, Hegeman GD, Kenyon GL. Carbon-13 nuclear magnetic resonance studies of mandelate metabolism in whole bacterial cells and in isolated, in vivo cross-linked enzyme complexes. Biochemistry 1981;20:1525–1533.
40. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S, Quondam M, Zucconi A, Hogue CW, Fields S, Boone C, Cesareni G. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science 2002;295: 321–324.
41. Brannetti B, Via A, Cestra G, Cesareni G, Helmer-Citterich M. SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. J Mol Biol 2000;298:313–328.
42. Macias MJ, Wiesner S, Sudol M. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. FEBS Lett 2002;513:30–37.
43. Ball LJ, Jarchau T, Oschkinat H, Walter U. EVH1 domains: structure, function and interactions. FEBS Lett 2002;513:45–52.
44. Freund C, Dotsch V, Nishizawa K, Reinherz EL, Wagner G. The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. Nat Struct Biol 1999;6: 656–660.
45. Schluter K, Jockusch BM, Rothkegel M. Profilins as regulators of actin dynamics. Biochim Biophys Acta 1997;1359:97–109.
46. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res 2003;31:248–250.
47. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu A, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. Global

mapping of the yeast genetic interaction network. Science 2004;303: 808–813.
48. Lee CH, Leung B, Lemmon MA, Zheng J, Cowburn D, Kuriyan J, Saksela K. A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. EMBO J 1995;14:5006–5015.
49. Petrella EC, Machesky LM, Kaiser DA, Pollard TD. Structural requirements and thermodynamics of the interaction of proline peptides with profilin. Biochemistry 1996;35:16535–16543.
50. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LS, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJ, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Zhong F, Zhong W, Gibbs R, Venter JC, Adams MD, Lewis S. Comparative genomics of the eukaryotes. Science 2000;287:2204–2215.
51. Adzhubei AA, Sternberg MJ. Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. Protein Sci 1994;3:2395–2410.
52. Zarrinpar A, Bhattacharyya RP, Lim WA. The structure and function of proline recognition domains. Sci STKE 2003;179: 1–10.