

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/236059767>

Polyproline-II Helix in Proteins: Structure and Function

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MARCH 2013

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2013.03.018 · Source: PubMed

CITATIONS

60

READS

562

3 AUTHORS, INCLUDING:



Alexei A Adzhubei

Engelhardt Institute of Molecular Biology (E...)

23 PUBLICATIONS 1,622 CITATIONS

[SEE PROFILE](#)

Polyproline-II Helix in Proteins: Structure and Function

Alexei A. Adzhubei¹, Michael J. E. Sternberg² and Alexander A. Makarov¹

1 - Engelhard Institute of Molecular Biology, Russian Academy of Sciences, Vavilov Street 32, Moscow 119991, Russia

2 - Structural Bioinformatics Group, Division of Molecular Biosciences, Imperial College London, London SW7 2AX, UK

Correspondence to Alexei A. Adzhubei and Michael J. E. Sternberg: alexei.adzhubei@eimb.ru;

m.sternberg@imperial.ac.uk

<http://dx.doi.org/10.1016/j.jmb.2013.03.018>

Edited by A. Panchenko

Abstract

The poly-L-proline type II (PPII) helix in recent years has emerged clearly as a structural class not only of fibrillar proteins (in collagen, PPII is a dominant conformation) but also of the folded and unfolded proteins. Although much less abundant in folded proteins than the α -helix and β -structure, the left-handed, extended PPII helix represents the only frequently occurring regular structure apart from these two structure classes. Natively unfolded proteins have a high content of the PPII helices identified by spectroscopic methods. Apart from the structural function, PPII is favorable for protein–protein and protein–nucleic acid interactions and plays a major role in signal transduction and protein complex assembly, as this structure is often found in binding sites, specifically binding sites of widely spread SH3 domains. PPII helices do not necessarily contain proline, but proline has high PPII propensity. Commonly occurring proline-rich regions, serving as recognition sites, are likely to have PPII structure. PPII helices are involved in transcription, cell motility, self-assembly, elasticity, and bacterial and viral pathogenesis, and has an important structural role in amyloidogenic proteins. However, PPII helices are not always assigned in experimentally solved structures, and they are rarely used in protein structure modeling. We aim to give an overview of this structural class and of the place it holds in our current understanding of protein structure and function. This review is subdivided into three main parts: the first part covers PPII helices in unfolded peptides and proteins, the second part includes studies of the PPII helices in folded proteins, and the third part discusses the functional role of the PPII.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The history of the discovery of the poly-L-proline type II (polyproline-II or PPII) helix is strikingly different from the two major structures of folded (globular) proteins, the α -helix and the β -structure (Fig. 1). This left-handed helix has been long known as one of the collagen triple-helix structures.⁹ However, it took a long time for the PPII helix to be recognized as one of the major conformations of polypeptide chain in folded proteins. It has been shown at the very early stages of protein structure analysis that the PPII helix can be adopted by polypeptides that are not necessarily dominated by proline, or even do not contain proline.¹⁰ Finally, at the end of the 1980s, the PPII helix has been shown to occur frequently in natural polypeptides and globular proteins.^{6,11,12} Protein Data Bank (PDB)¹³

analysis showed that the PPII conformation of individual amino acid residues occurs at rates comparable with the β -strand conformation.⁶ PPII helices form predominantly short stretches of regular structure characterized by the repetitive pairs of Ramachandran ϕ and ψ angles.¹ They are found in most proteins and include about 2% of residues in the PDB.^{1,3,5,14} This is much lower than the occurrence of the α -helix and β -structure but comparable with the 3(10)-helix.

The PPII helix is an extended (3.1 Å per residue compared to 1.5 Å in the α -helix) left-handed helix defined by the ϕ , ψ torsional angle cluster with the distribution maximum at -75° , 145° . In the ideal form, it has three residues per turn ($n = -3$), has 3-fold rotational symmetry, and has the shape of a triangular prism. Thus, extended structures in proteins include the PPII helix, the β -structure, and a

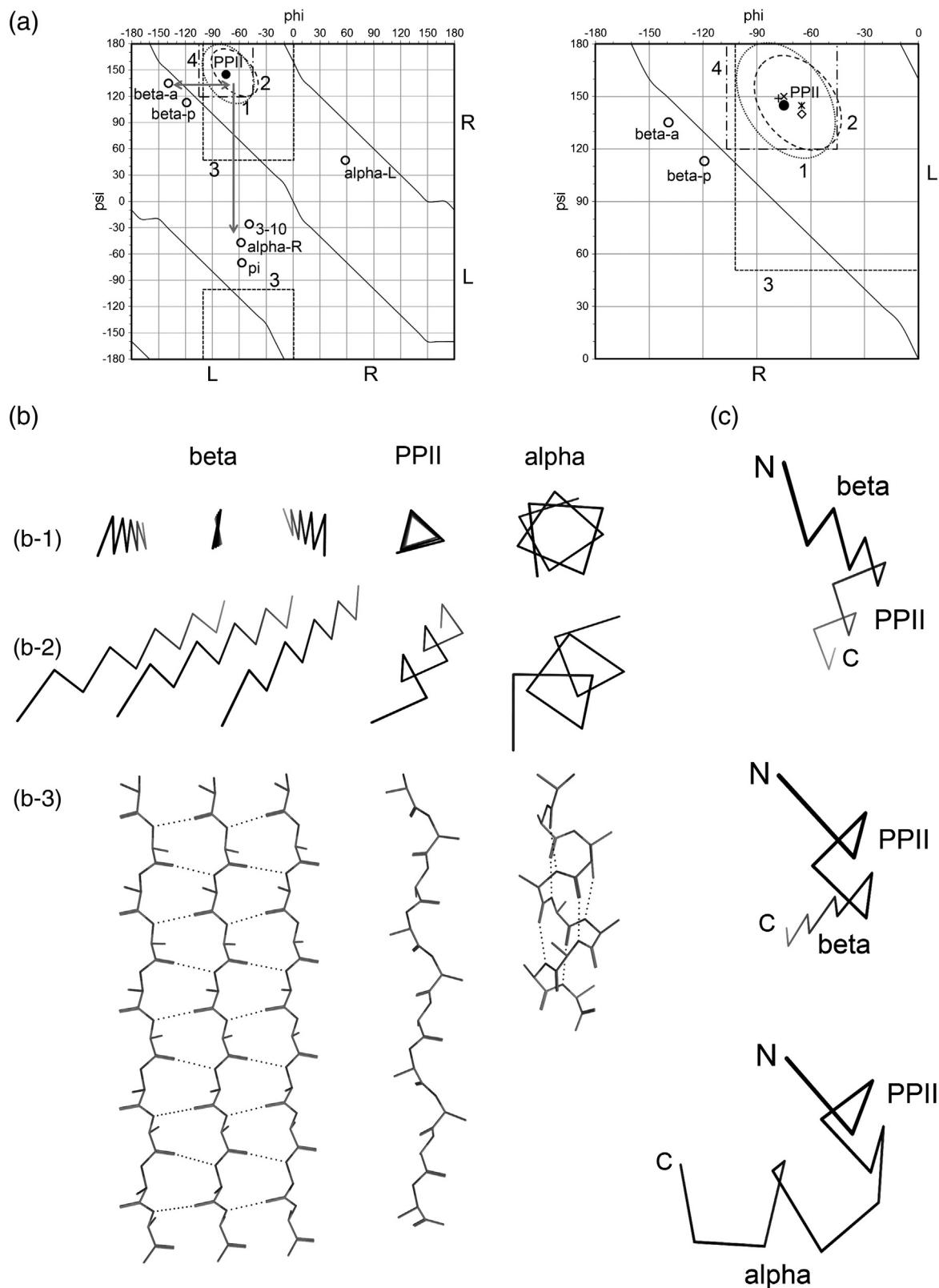


Fig. 1 (legend on next page)

much less frequent single β -strand^{1,15} (Fig. 1). The PPII helix does not support regular patterns of intrachain hydrogen bonds because of its extended character; it neither forms regular interchain hydrogen bonds, although a limited number of interchain hydrogen bonds and H-bonds with water, as well as side-chain–main-chain hydrogen bonds, can be formed.^{5,7,16} Though the conformational freedom of PPII can be partially limited when it includes Pro due to its pyrrolidine ring, the extended PPII helix is markedly more flexible in comparison with the α -helix and β -sheet. Because of the conformational properties of proline, it has been suggested that PPII is also a dominant conformation in the structure of the proline-rich regions (PRRs) in proteins.¹⁷ In general, it provides for the necessary structural flexibility and mobility and often appears at the edges of α -helices,¹ in the domain linker regions,^{18,19} interactions interfaces, and protein terminal regions.¹⁷

From the time of the first discovery in folded proteins, a substantial volume of data has been collected showing the importance of the PPII helix for the functional and structural machinery of peptides and proteins.^{15,17,20,21} PPII helices play a major role in protein–protein^{15,22} as well as protein–nucleic acid²³ interactions. In free peptides in solution and proteins, PPII conformation can be observed by CD^{12,24–26} vibrational circular dichroism (VCD),²⁷ and Raman optical activity (ROA).^{28,29} NMR spectroscopy has been successfully used to assess PPII fraction in peptides.³⁰ These methods provide information about overall structural content, but not the residue-by-residue structure of the polypeptide

chain. In folded proteins, PPII helices have been identified by analyses of experimentally determined structures.^{1,3,5,14}

PPII Structure in Unfolded Peptides and Proteins

PPII helix early studies: PPII and “random coil”

Tiffany and Krimm examined the CD spectra of poly-L-glutamic acid (PGA) in solution, where they observed α -helical conformation at low pH and a distinctive two-band spectrum similar to that of the poly-L-proline II (198 nm negative and 218 nm positive) at high pH (7–12) and low salt concentration, resulting in the charged form of PGA (Fig. 2). They suggested that this spectrum, as well as the similar spectrum of charged poly-L-lysine (PL), indicated the peptides that are not fully unstructured but adopt a regular extended helical conformation close to the threefold PPII left-handed helix, short stretches of which are interspersed with turns.³² They also argued that earlier similar CD patterns were misinterpreted as a maximally unordered random coil.³³ Similar CD spectra were observed later in poly(Glu-Ala), poly(Ala-Gly-Gly), poly(Pro-Ala-Gly), poly(Ala-Ala-Gly), and poly(Pro-Ser-Gly) (for references, see Krimm and Tiffany³³). Tiffany and Krimm concluded that the unstructured, also defined as unordered, conformation of heat-denatured PGA and proteins is a distinct state that can be

Fig. 1. (a) Ramachandran plot representation of the PPII helix and other regular periodic structures in proteins. Regions enclosing right-handed and left-handed conformations are marked R and L, respectively. PPII, left-handed PPII helix; beta-a and beta-p, antiparallel and parallel β -structure; alpha-R, right-handed α -helix; alpha-L, left-handed α -helix; 3-10, 3(10)-helix; pi, π -helix. Enclosed areas marked 1 (Adzhubei and Sternberg¹) and 2 (Hollingsworth *et al.*²) show clusters of PPII helices of three and more residues; 3 represents the PPII basin of Jha *et al.*³ that included single residues and continuous segments, and for a separate coil library, only stretches of four and more consecutive residues; 4 is the area of the mesostates identifying PPII in the Protein Coil Library database (Fitzkee *et al.*⁴) that included fragments of at least two continuous residues. The upper left quadrant of the Ramachandran plot on the right details the positions of the PPII clusters observed in protein structure surveys. The filled circle (●) indicates the PPII distribution maximum according to surveys of Adzhubei and Sternberg¹ and Stapley and Creamer⁵, diamond (◊), Adzhubei *et al.*⁶; x, Jha *et al.*³; *, Cubellis *et al.*⁷ and Hollingsworth *et al.*² + shows ϕ , ψ of the poly-L-proline II crystal structure (Arnott and Dover⁸). Although definitions of the PPII area bins differ considerably in surveys, the resulting clusters center close together on a narrow area, revealing regular character of the PPII structure, independent of the bin definition parameters. Arrows show the ϕ , ψ transitions resulting in the single-residue structure switches in proteins: left-handed PPII helix \rightarrow right-handed helix, and PPII helix \leftrightarrow β -structure, as shown in (c). (b) Idealized major periodic structures: β -structure, α -helix, and PPII helix, modeled as the CA-trace helical axis projection (b-1) and perspective projection (b-2), and the 10-Ala polypeptide chain (b-3). PPII helix with $n = -3$ and $d = 3.1 \text{ \AA}$ is a left-handed narrow and extended helix, the most extended helical structure occurring in proteins, and only slightly less extended than the β -structure. For α -helix, $n = +3.6$ and $d = 1.5 \text{ \AA}$; thus, PPII helix covers twice its length per residue. It does not form any regular pattern of local intra- or interchain hydrogen bonds. The CA-trace projection along the helical axis shows the PPII characteristic shape of a triangular prism (b-1). (c) Idealized models of the single-residue structure switches involving PPII helices observed in proteins¹ and shown as the CA-trace perspective projection. The PPII- α -helix motif is formed by the PPII \rightarrow α -R switch involving a single PPII boundary residue, where the ϕ torsional angle remains in the PPII conformation, while the ψ angle shifts to the α -R conformation, changing the chain chirality and resulting in a sharp turn (bottom). Similar structure switches PPII \rightarrow β and $\beta \rightarrow$ PPII produce less pronounced bends in the PPII- β -strand and β -strand-PPII structural motifs (middle and top). [Graphical abstract, section “ ϕ , ψ plot” is Copyright © 1987 Elsevier. Adzhubei, A.A., *et al.* (1987). *Biochem. Biophys. Res. Commun.* 146, 934–938. Reproduced with permission of Elsevier.]

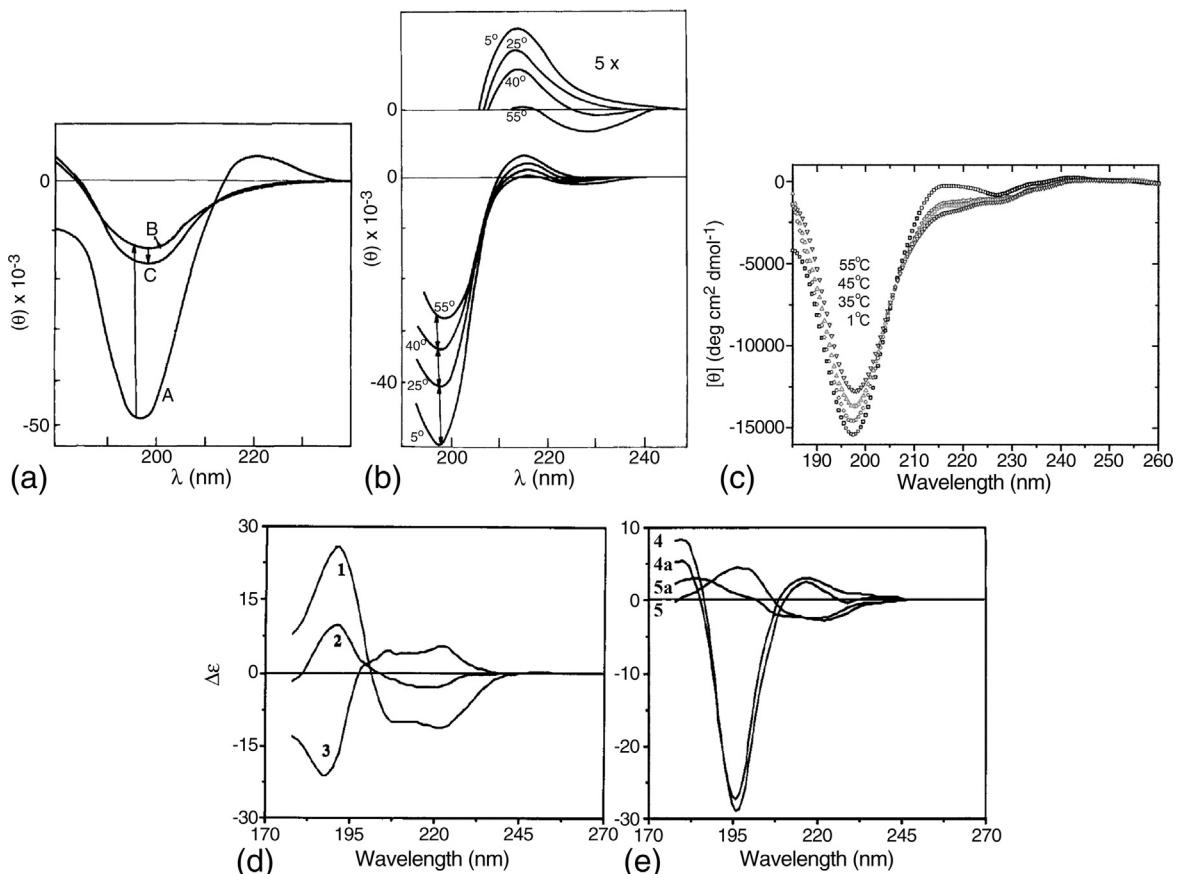


Fig. 2. CD spectra characteristic of the PPII. (a) The CD spectrum of the rat tendon collagen: A, native form; B, heat denatured; and C, heat denatured and cooled. Adapted from the classic work by Tiffany and Krimm.³¹ [Copyright © 1972 John Wiley & Sons, Inc. Tiffany, M.L. & Krimm, S. (1972). *Biopolymers* **11**, 2309–2316. This material is reproduced with permission of John Wiley & Sons, Inc.] (b) The spectrum of the charged PGA at pH 7 as a function of temperature changes with temperature increase from 5° to 55 °C. (Tiffany and Krimm³¹). (c) Changes in the conformation of the seven-alanine XAO peptide with temperature showing a similar pattern to (b). Temperature range from 1 to 55 °C corresponds to the CD spectra from bottom to top. Adapted from Shi *et al.*³⁰ [Copyright © 2002, National Academy of Sciences, USA. Shi, Z., *et al.* (2002). *Proc. Natl. Acad. Sci. USA* **99**, 9190–9195. Reproduced with permission of National Academy of Sciences, USA.] NMR data from the same study suggest PPII predominant at low temperature with an increase of the β-conformation at higher temperature. (d, e) Deconvoluted CD spectra characteristic of different classes of secondary structure in proteins, adapted from Sreerama and Woody.²⁶ [Copyright © 1994 American Chemical Society. Sreerama, N. & Woody, R.W. (1994). *Biochemistry* **33**, 10022–10025. Reproduced with permission of American Chemical Society.] (d) 1, α-helix; 2, β-sheet; 3, turns. (e) 4, PPII including single residues; 4a, PPII helices of two and more residues; 5 and 5a, corresponding unordered segments.

achieved by heating and subsequently cooling PGA in a 4.5-M solution of LiClO₄. In this case, the CD spectrum looks like a broad, weak negative band around 200 nm (± 10 nm).³⁴ The classic model of the random-coil polypeptide chain, which has been and sometimes still is considered to describe the conformation of unfolded peptides and proteins, implies that amino acid residues are distributed randomly in the sterically allowed regions of the Ramachandran plot (reviewed by Schweitzer-Stenner³⁵). However, the work of Tiffany and Krimm indicated that the conformation of the analyzed polypeptides was not an unstructured polypeptide chain and that the conformational state termed “random coil” in fact retains a degree of local order similar to the PPII

conformation albeit with small differences in the CD spectra. Tiffany and Krimm suggested that the balance of electrostatic and steric interactions of charged side chains favors formation of the extended helix (PPII) conformation in polypeptides in water with its high dielectric constant.³⁶

The presence of PPII in the conformational state that traditionally had been considered random coil was further supported by the work of Drake *et al.* In their analysis of temperature-dependent conformational states of PL using CD spectra, low temperature favored the formation of PPII.³⁷ This conclusion was corroborated on poly-L-proline, PGA, and PL of various lengths by analysis and comparisons of VCD and electronic CD spectra at different salt

concentrations, temperature, and pH.³⁸ Although this interpretation was not immediately accepted, it was later largely verified by numerous studies (reviewed by Woody²⁴).

PPII structure in model peptides

Short peptides provide a different model of polypeptide chain conformational properties compared with the longer polypeptides, such as the charged, unfolded PL and PGA. In particular, short peptides dominated by Ala represent a system closely resembling polypeptide backbone without long-range interactions because of its size and the nature of side chains. Ala-based peptides are also important as an established model for helix-coil transition and random coil. Short peptides can better sample the allowed conformational space and thus provide insight on the random-coil-like conformations in unfolded proteins. End effects, however, can considerably alter their conformational distribution.^{39,40}

A substantial body of spectroscopic and theoretical calculations data on the PPII, mostly in unfolded peptides and proteins, has been accumulated by the beginning of 1990s, and several detailed analyses of these data were published, starting with the review by Woody.²⁴ In their 2002 review, Shi *et al.* focused on the evidence of the PPII dominance in short peptides based on the computer simulations results and experimental data obtained using NMR and other major spectroscopic methods, which emerged as a powerful tool to study polypeptide chain in solution, in particular CD, infrared (IR) vibrational spectroscopy, and VCD.²⁰ The authors concluded that, at certain conditions, such peptides adopt a distinctive structure, the PPII, contrary to the previously suggested classic random-coil model.

The reviews by Rath *et al.*²¹ and the Kallenbach group (Shi *et al.*)⁴¹ centered on the role of PPII in unordered proteins; they also considered it as a predominant conformation of a number of model peptides, including short peptides with the sequence not containing Ala and Pro. Other represented conformational states in such peptides are β -strand, β -turn, and unordered. The review by Shi *et al.* and other publications^{42,43} suggest that the PPII conformation displays weakly cooperative to non-cooperative behavior, in agreement with its non-cooperative nature proposed by Makarov *et al.*^{25,44}

In a particularly important work by Shi *et al.*, PPII was identified in the XAO peptide containing seven Ala residues and commonly used as a model of unordered state, using NMR and CD³⁰ (Fig. 2). The NMR data showed that the conformational distribution at 2 °C consisted predominantly of PPII and up to 10% β -strand. This distribution remained mostly unchanged after a temperature shift to 55 °C with only 10% increase of the β -strand content. Both

NMR and CD indicated thermal transition between conformational states. Further support for the PPII conformation of Ala peptides in aqueous solution was provided by the ROA spectra analysis of the seven-alanine peptide.²⁹

A contrasting view was offered in the work by Scheraga's group on a CD and NMR analysis, combined with theoretical calculations of the XAO peptide.^{45,46} Makowska *et al.* stated that the radius of gyration observed for XAO was too small to represent the PPII structure and concluded that the level of PPII was previously overestimated and it actually represented one of the local conformational states but not a dominant conformation in XAO and unfolded proteins.

In their discussion of the CD spectra observed by Scheraga's group, Shi *et al.*⁴¹ remarked that an alternative interpretation of these data is possible, pointing to the PPII as the dominant conformation in XAO, and that NMR data alone cannot be used for the final structural assignment. They also argued that spectroscopic and modeling data consistently support the assignment of predominantly PPII conformation in Ala short peptides, a view supported by Barron's group simulations of the observed ROA spectrum of XAO and other ROA data.⁴⁷

In fact, these views are complementary in a certain way. The excitonic coupling modeling of Fourier transform infrared (FTIR) spectroscopy, VCD, and isotropic and anisotropic Raman spectra of XAO showed the presence of an ensemble of conformations, including turn and β -strand, whereas PPII was dominant at about 50%.⁴⁸ It has also been proposed on the basis of small-angle X-ray scattering and molecular dynamics (MD) simulations that XAO has a flexible structure in which each residue is represented by a set of conformations fluctuating around, but rarely adopting the ideal PPII conformation.⁴⁹ Zhu *et al.* suggested that if such flexible conformation indeed exists for XAO, the uncharacteristically small radius of gyration can be explained by the above findings, while the torsion angles averaging leads to the assignment of standard PPII.⁴⁷ This model can be suggested as generally well describing the PPII helix in peptides and proteins—a flexible structure not restricted by regular pattern of hydrogen bonds and capable of fast conformational changes within the PPII ϕ , ψ boundaries.

Recently, Graf *et al.* performed a combined MD and NMR study on the Ala and Val homopeptides of different length, corroborating previous results on the PPII as predominant conformation of Ala peptides.⁵⁰ The validity of their approach was verified by Schweitzer-Stenner, who reproduced the set of *J*-coupling constants used in this work.⁵¹ The Ala tripeptide was shown by Graf *et al.* to have up to 90% fraction of PPII, 10% of β -strand, and no α -helix, and the Val tripeptide had significant fractions of PPII, β -, and α -helix conformations. No marked changes in

the conformation were detected for Ala peptides with the increase of length up to 7 residues. In the synthetic 9- and 19-residue peptides corresponding to the N-terminal part of lysozyme sequence containing central Ala tripeptide, no changes in the Ala tripeptide conformational distribution were observed in the shorter peptide; however, in the 19-residue peptide, it displayed 30–40% of the α -helix conformation. Thus, chain length and sequence context substantially influenced the conformation of the central Ala tripeptide.

The interplay between PPII, α -helix, and peptide length is manifested clearer in longer peptides that can undergo transition from a folded to an unfolded state. ROA observations showed that the increase of temperature from 0 °C to 30 °C induced a transition of the alanine-rich 21-residue peptide AK21 from an α -helical conformation to an unfolded, possibly PPII, conformation.⁵² In agreement with these results, Asher *et al.* used UV Raman spectroscopy to find that the non- α -helical conformations of Ala peptides, including short peptides and a 21-Ala-based peptide, were predominantly PPII, without any evidence for random coil.⁵³ The UV Raman spectra of the XAO peptide were largely the same as the spectra of Ala and Ala-based peptides of various lengths and displayed essentially similar temperature dependencies. The 21-Ala-based peptide underwent thermal melting from the mainly α -helical conformation at 0 °C to mainly PPII helix at above room temperature, notably without the significant presence of intermediate forms.

The pattern observed in other peptides is less clear. Conformational distributions of Ala and Val tripeptides showed that PPII was the dominant conformation of Ala at a higher level than previously reported and that Val was mostly in the β -strand conformation combined with PPII and α -helix fractions.⁵¹ CD, NMR, and MD analyses of 400 blocked dipeptides, representing combinations of amino acids with high occurrence in unfolded proteins, showed that PPII and β -strand were preferable conformations, which did not strongly depend on the side chains.⁵⁴

Host–guest studies of the PPII propensities on the Gly-based short peptides, while confirming PPII prevalence for Ala, showed conformations of the majority of other residues distributed at comparable levels between PPII and β -strand^{55,56} with these two conformational regions substantially overlapping.⁵⁶ They also indicated the absence of significant α -helical fraction. Further results of the host–guest Gly-based peptide studies showed that PPII propensities were sequence dependent and affected by pH and by the presence of charged end groups.⁴⁰

The Lys and Glu peptides are of special interest since their conformational preferences can be compared to that of the PGA and PL. Thus, Lys was reported to adopt the PPII conformation in a

heptapeptide.⁵⁷ Lys, Glu, and Asp tripeptides were shown to adopt predominantly PPII conformation at room temperature.⁵⁸ However, in a subsequent study of the H-(AAKA)-OH peptide, Lys was reported to show equal sampling of PPII, β , and helical conformations, attributed to the nearest-neighbor effect and the context-dependent nature of the PPII.⁵⁹

Overall, on the strength of spectroscopic studies, PPII emerges as the prevalent conformation in unfolded Ala-based peptides in aqueous solution, and PPII and β -strand emerge as preferable conformations in other short peptides. These findings also suggest that PPII plays an important role in the transition between α -helix and unfolded state. Model peptide studies show that PPII is not the only local structure in unfolded peptides but that it is a predominant one—a conclusion largely supported and complemented by the studies of unfolded proteins.

The role of water in stabilization of the PPII helix

The geometry of PPII leads to exposed backbone carbonyl and amide groups, which extend into the solvent, unless they are occluded by bulky side chains. Since the PPII structure does not have intrachain hydrogen bonds, there were hypotheses assuming that it could be stabilized by water^{16,60,61} by forming a regular network of hydrogen bonds with water molecules.¹ CD and IR spectroscopy of small peptide hormones ACTH and β -LPH, β -endorphin, and C-terminal fragments of histones H1 and H5 showed preference for the PPII structure in aqueous solution.^{11,25}

Numerous results point to the importance of hydration for the PPII structure. Work from the Creamer group showed that a short seven-residue lysine peptide has the PPII conformation at the virtually uncharged state, which suggests that not only electrostatic forces but also interaction with solvent can be a stabilizing factor.⁵⁷ Host–guest studies of the PPII helix suggested that propensities to form PPII by apolar residues are partly defined by interactions of their side chains with solvent.^{16,62}

The analysis of liquid crystal NMR data indicates that, in water solution, the PPII conformation of the Ala dipeptide is dominant,⁶³ forming two “double-water” bridges. Calculations of the effect of hydration on the alanine dipeptide *N*-acetyl-L-alanine *N*-methylamide yielded two conformations with equal energy minima stabilized by water, PPII, and the α -helix, where H-bonds with water were formed by both structures, while comparison of the simulation to experimental vibrational absorption, Raman, and VCD spectra showed that the simulated PPII spectra best fitted the experimental data.⁶⁴ The β -extended and α -helix conformations were found to have comparable free-energy minima in the MD simulations of the Ala

dipeptide in solution.⁶⁵ This was further confirmed by later results obtained for Ala dipeptides in aqueous solution by Monte-Carlo-based computational analysis of ROA and Raman data. PPII and α -helix were found to be the predominant conformations with explicit interactions with solvent playing an essential role.⁶⁶ MD simulations of the Ala oligopeptide (Ala)₈ show that the PPII conformation in water is approximately two times more populated compared to the β -structure due to the one-molecule water bridges that favor PPII.⁶¹

The role of water bridges in stabilizing the PPII structure has been recently challenged by the analysis of a large data set of high-resolution crystal structures and MD simulations data.⁶⁷ The authors concluded that water bridges are found less frequently around PPII and that they do not provide favorable conditions for this conformation. Solvation free energy and Monte Carlo computations of 12-residue-long polyalanine showed favorable solvation free energy for PPII compared to other conformations with no water bridges formed by the PPII.⁶⁸

An alternative view gives preference to steric forces in the stabilization of PPII in solution. Results of minimization of chain packing density associated with the increase in solvation in unfolded state, as well as calculations of the accessible surface area, might provide another explanation why the PPII conformation is preferable under such conditions based on steric forces alone.^{69,70} Similar logic was applied in the work on Monte Carlo simulations of the hydrated Ala dipeptide.⁷¹ The results indicated that direct water-peptide interactions could not explain the observed preference for PPII, which was determined rather by minimizing steric conflicts within the peptide. Stereoelectronic effects can serve as another important factor of PPII stabilization.⁷²

In their review of existing experimental and computer modeling data on PPII, Creamer and Campbell concluded that the major factor determining the PPII conformational propensities of amino acid residues is backbone solvation, followed by steric interactions and, to a lesser extent, side-chain-backbone hydrogen bonds, with conformational entropy not playing a major role.⁷³

Results identifying PPII as a predominant conformation of Ala-based peptides, a system that models protein backbone, strongly suggest that PPII represents one of the basic structures of polypeptide chain. Spectroscopic studies of short peptides provide a wealth of data on the PPII helix, including propensity to form PPII depending on the length of polypeptide chain and amino acid composition, impact of end and near-neighbor effects, conformational transitions involving PPII, and non-cooperative character of this structure. These studies also produced results that allow the assessment of the nature of forces stabilizing PPII. Host-guest peptide studies provide data on the amino acid propensities

to form the PPII structure, which will be discussed later in this review. Results of the spectroscopic studies of model peptides can be utilized in the analysis of structure of the unfolded polypeptides and proteins.

Denatured and natively unfolded proteins

Based on the work of the Creamer group on residue propensities to adopt α -helical and PPII structures, Creamer and Campbell suggested that residues in unfolded proteins are distributed almost exclusively between the α -helical and PPII conformational regions.⁷³ After analyzing spectroscopic data, Shi *et al.* interpreted them as pointing to the ability of the PPII helix to maintain local order in largely unfolded proteins and peptides, with the fraction of the PPII depending on the denaturing agents and temperature and specifically stabilized by urea at low temperatures.⁴¹ In particular, Creamer's group used CD to demonstrate that urea promoted formation of the PPII structure in short peptides and denatured proteins.⁷⁴ Solid-state NMR analysis of the villin headpiece subdomain (HP35) in folded, partially denatured, and fully denatured (by guanidine hydrochloride) states showed that in the process of unfolding of the polypeptide chain, conformation shifts from the α -helical to PPII and in the fully denatured state adopts a combination of partially extended and PPII conformations.⁷⁵

Just as in conflicting interpretations of the structure of the XAO peptide, the results showing the presence of local structure in denatured proteins seemed to be at variance with the work showing that the radii of gyration of the chemically denatured proteins correspond to the random-coil state.⁷⁶ However, analysis of the ensembles generated using Monte Carlo computations from known protein structures, by artificially simulating random conformations of only about 8% of residues with the rest of the chain remaining in ordered conformation, showed that such systems also conformed to the expected random-coil end-to-end distances and radii of gyration.⁷⁷ Thus, unfolded polypeptides and proteins, while conforming to the random-coil scaling, can retain a high degree of local structure.^{78,79}

Unstructured, natively unfolded, or intrinsically disordered proteins (IDPs) and domains are increasingly recognized as widely spread, functional, and structurally conserved.^{19,80,81} They display varying degrees of disorder, from folded proteins with local disorder to unstructured ones. The formal trinity classification of "folded–molten globule–unstructured" proteins can also be supplemented with the "pre-molten globule".⁸¹ Though the content of secondary structure in unfolded regions and proteins varies, there are different opinions as to whether, in a highly disordered state, they can possess some local secondary structure¹⁹ or its presence is minimal.⁸¹

Evidence has been accumulating that natively unfolded proteins possess various degrees of local secondary structure and that the presence of the PPII in such proteins as a flexible structure can be significant.^{47,82}

Optical spectroscopy^{47,82} and NMR⁸³ are the prevailing techniques used to characterize unfolded structures. Observations made on unstructured proteins using spectroscopy provide evidence of high PPII structural content and underpin the conclusion that it forms a substantial part of such molecules. A multivariate mapping analysis of a large data set of ROA spectra of natively unfolded and denatured, as well as folded, proteins accompanied by *ab initio* simulations of the Raman and ROA spectra of alanine oligopeptides showed high content of PPII in natively unfolded structures, confirming that the PPII conformation plays a significant structural role in natively unfolded proteins and disordered peptides.⁴⁷ In comparison with the natively unfolded proteins, denatured proteins were found by this analysis to contain less PPII and more α -helical or β -structures. Thus, spectroscopic data suggest that PPII not only is a non-native structure found in denatured proteins and peptides but also constitutes a local structure of the natively unfolded proteins.

PPII is a structure adopted in a wide variety of unfolded proteins divergent both structurally and functionally. For example, ROA spectra analysis showed that PPII is an important conformational element in the natively unfolded milk casein proteins and the Bowman–Birk inhibitor, described as supporting the “static” type of disorder.^{28,84} In another work, natively unfolded wheat protein ω -gliadin and the T-A-1 peptide from the Dx5 subunit of high-molecular-weight glutenin were found by ROA to contain high proportions of PPII, while more structured A-gliadin contained less PPII.⁸⁵ One of the tooth enamel amelogenins, the recombinant porcine rP172, exists as an unfolded protein but has been shown by CD and NMR to possess local PPII, extended β -strand, α -helix, and turn/loop structures.⁸⁶ β -Strand, turn/loop, and PPII were found by CD spectroscopy in the intrinsically disordered chaperone-like protein *LjDP1* from the model legume *Lotus japonicus*.⁸⁷ CD spectroscopic analysis of the natively unfolded protein human securin also showed a degree of PPII structure.⁸⁸ Spectroscopic analysis of the spider dragline silk proteins before spinning revealed unordered proteins with high content of PPII structure and several α -helical segments.⁸⁹

A separate category includes proline-rich unstructured proteins, since high proline content shifts structural bias towards the PPII helix. Comparison of the bias of amino acid composition of IDPs^{90–93} and PPII amino acid preferences shows that they partially overlap and that the same

residues can favor both “disorder” and PPII. FTIR spectroscopy of the acidic salivary protein PRP1 (proline-rich protein 1) characterized by unique amino acid composition with high content of proline (27.5%), glutamine (22.8%), and glycine (20.8%) in its mostly unstructured state showed large regions of PPII helix alternating with unstructured random coil as well as a small fraction of β -sheet.⁹⁴ The proline-rich calcium-binding protein RVCaB found in radish vacuoles differs by amino acid sequence from other calcium-binding proteins. It has a high content of Glu, Val, Lys, and Pro and no aromatic amino acid residues. CD spectroscopy showed that RVCaB has about 25% of PPII helix with the rest of the protein remaining unstructured, and differential scanning calorimetry showed no highly cooperative folding transitions.⁹⁵ Unstructured human salivary PRP IB-5 has been shown by CD and NMR to contain PPII helices, with the PPII elements probably involved in protein–tannin binding.⁹⁶

The presence of PPII as a significant or even dominant local structure in the unfolded peptides and proteins has been identified over the last 20 years by spectroscopic and other experimental techniques as well as theoretical computations. This remarkable fact has implications for our understanding of unfolded proteins and protein folding, as well as for the random-coil model.^{77,78,97}

PPII Helices in Folded Proteins

Coil libraries and structure surveys

Spectroscopic methods provided important information on the PPII helices as a conformation of polypeptide chain in unfolded systems. Identification of the PPII by CD spectroscopy has been extended to folded proteins^{26,82,98} (Fig. 2). However, only analyses of high-resolution experimental crystallographic structures allowed to assign PPII helices in proteins and conclusively identify PPII helix as a distinct class of protein regular structure. Results of protein structure surveys unquestionably established PPII among the other major regular structures— α -helix and β -sheet.

Early report of the presence of isolated PPII helices in folded proteins can be traced to the work of Soman and Ramakrishnan.⁹⁹ Initially, Adzhubei *et al.* carried out an analysis of 68 PDB structures in order to identify amino acid residue conformations that did not fall into the accepted major classes and assess their frequencies. They found that PPII was the only such conformation occurring on a scale comparable with the α -helix and β -strand. A restricted library approach where β -sheets were excluded from the B (β) region of the Ramachandran plot was

used in this work. The analysis showed that PPII was present as a conformation of single residues and as a regular structure—the PPII helix, with the PPII conformation content comparable to the β -strand conformation^{6,60} (Fig. 3). Insufficient number of high-resolution structures in the PDB release used in this work resulted in the data set that included structures with resolution below 2.0 Å, affecting the reliability of the results. However, these results were subsequently corroborated in several analyses of coil libraries^{3,4,100} and structure surveys^{2,7} utilizing data sets composed of high-resolution structures (Figs. 1 and 3). Results published by other groups also showed the presence of PPII conformation, though it was not explicitly described.^{101,102}

Jha *et al.* reported PPII as the predominant conformation in a large coil library composed of 2020 PDB chains with resolution better than 2.0 Å and with length over 20 residues³ (Fig. 3). The PPII preference expressed in modified Chou–Fasman parameters (representing normalized percentage fractions) varied from 42.3 (regular α - and β -structures and turns removed) to 33.2 (regular α - and β -structures removed) in the library subsets compiled according to different criteria, including subsets compiled using stringent selection rules. Though the question of how to compile a better balanced coil library is not yet settled, in the three different libraries analyzed by Jha *et al.*, PPII was either prevalent (two libraries) or present at the level comparable with α -helical and β -strand conforma-

tions. Preference for PPII was not directly linked to solvation, as high levels of PPII were observed even for buried residues, but was affected by chemical and structural properties of the nearest neighbors. The former observation might be explained by the fact that coil libraries do not adequately represent the solvation patterns of unstructured proteins where side chains and backbone are more exposed than in folded proteins. This bias might result in different patterns of forces that stabilize PPII in folded and unfolded proteins.

Coil libraries' conformational preferences differ from the model peptides in that there is a relatively strong presence of helical conformations, which has been attributed to the context effects.⁵⁶ Though estimates of the level of PPII in coil libraries vary,¹⁰³ it has been reported to correlate with PPII content in unfolded peptides.³

Adzhubei and Sternberg used a data set of 80 PDB structures in a survey of PPII helices in folded proteins and established that they constitute a distinct class of regular, extended, left-handed helices populating a compact cluster on the Ramachandran plot with the ϕ , ψ torsional angle distribution maximum at -75° , 145° , and have the three-residues-per-turn, 3-fold symmetry ($n = -3$)¹ (Fig. 1). PPII helices were observed as predominantly short stretches of three to eight residues. At least one PPII helix was found in nearly all of the analyzed proteins, resulting in residue frequencies in the PPII helices of 2% to 4% of the data set. PPII

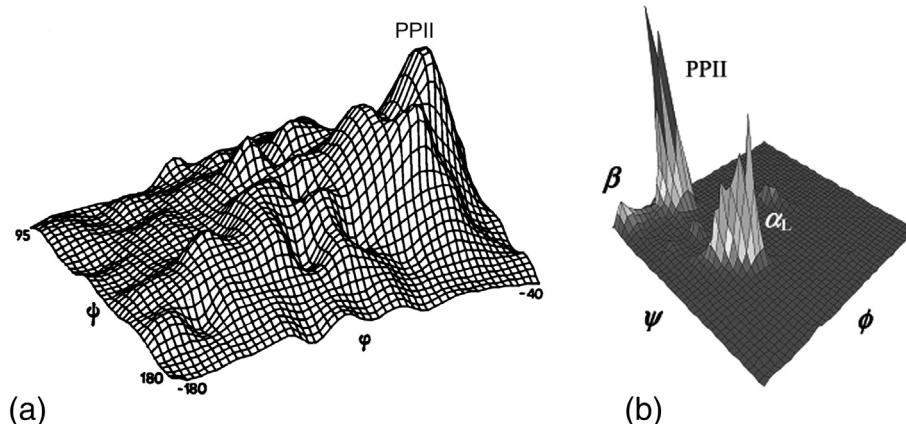


Fig. 3. PPII structure in restricted libraries. (a) PPII cluster identified in an early survey in which a restricted library was constructed by removing β -sheet residues from the β region of the Ramachandran plot. This survey included 68 protein chains from the PDB. The cluster is centered at ϕ , ψ -65° , 140° . Adapted from Adzhubei *et al.*⁶⁰ [Copyright © 1987 Taylor & Francis. Adzhubei, A.A., *et al.* (1987). *J. Biomol. Struct. Dyn.* **5**, 689–704. Reproduced with permission of Taylor & Francis.] (b) Ramachandran plot of the restricted library with helices and sheets removed. The PPII cluster is centered at ϕ , ψ -75° , 150° . The survey incorporated 2020 protein chains with sequence identity below 25% and resolution better than 2.0 Å from the PDB. Adapted from Jha *et al.*³ [Copyright © 2005 American Chemical Society. Jha, A.K., *et al.* (2005). *Biochemistry* **44**, 9691–9702. Reproduced with permission of American Chemical Society.] The distribution maxima in both surveys lie within $\pm 10^\circ$. The PPII cluster in (a) contained around 19% of the residues; that in (b) contained around 33%, the differences arising possibly due to discrepancies in the criteria defining the PPII bins and the PPII structure itself, as well as the size of the data sets.

helices are located mostly on protein surface forming hydrogen bonds with water and retain a high degree of conservation in protein structures.^{1,18}

A detailed survey of a large data set consisting of 3268 proteins from a database of structurally aligned

families revealed similar PPII occurrence.⁷ The ϕ , ψ of PPII helices and β -strands are positioned close in the conformational space. A subset of structures with a resolution threshold of 1.8 Å was used in the survey to reliably delineate the PPII helix and β -strand clusters,

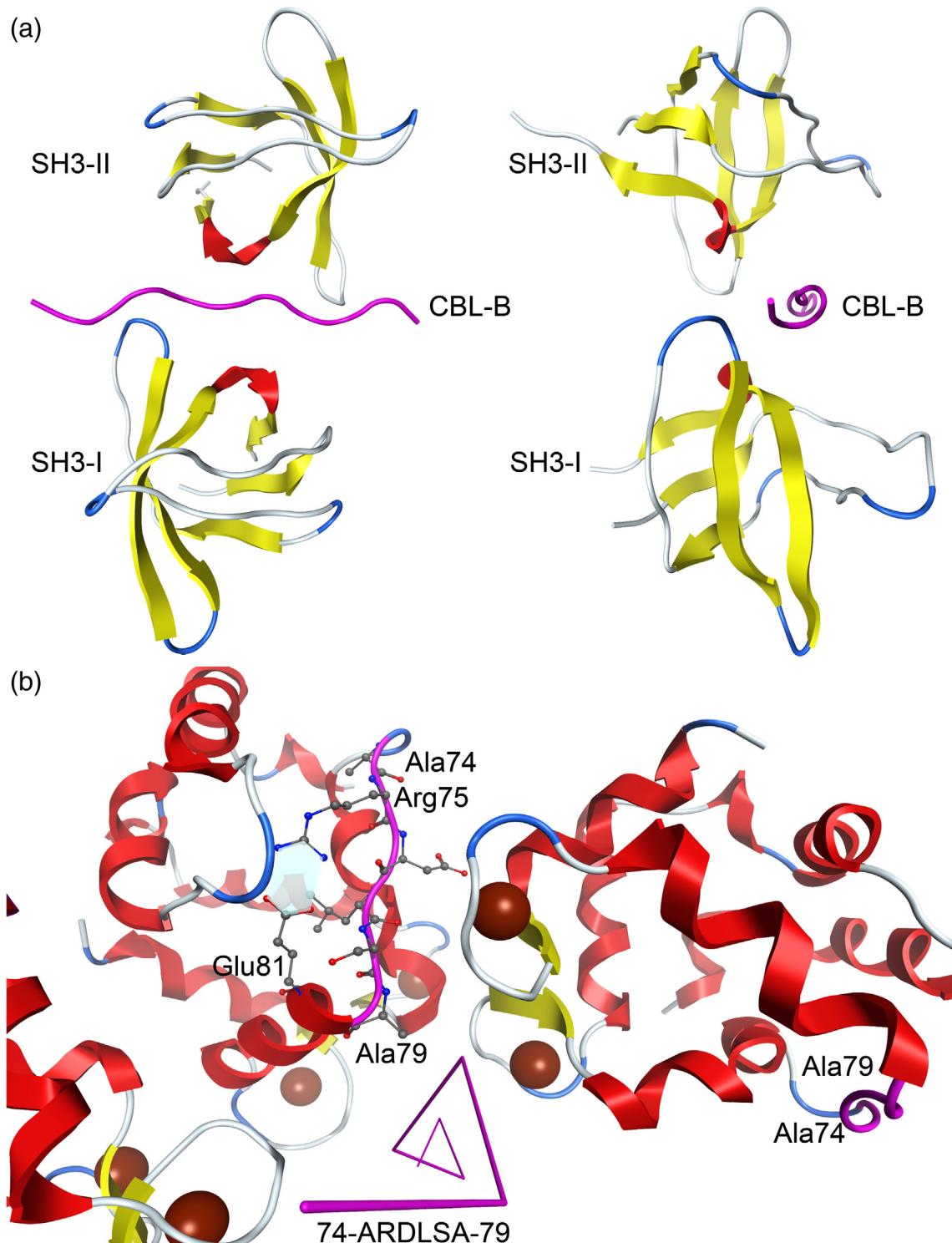


Fig. 4 (legend on page 2113)

showing that they represent distinct structural classes, with the isolated β -strand conformations partially overlapping the PPII cluster. These findings were confirmed both for the PPII conformation of single residues and for PPII helices on various data sets and using different PPII definition criteria.^{2,5,14,26}

An important outcome of the protein structure surveys is that, in spite of the marked differences in the analyzed data sets and PPII identification criteria, they result in relatively small fluctuations of characteristics of the PPII structures. This, in turn, confirms that PPII is a secondary-structure class with stable

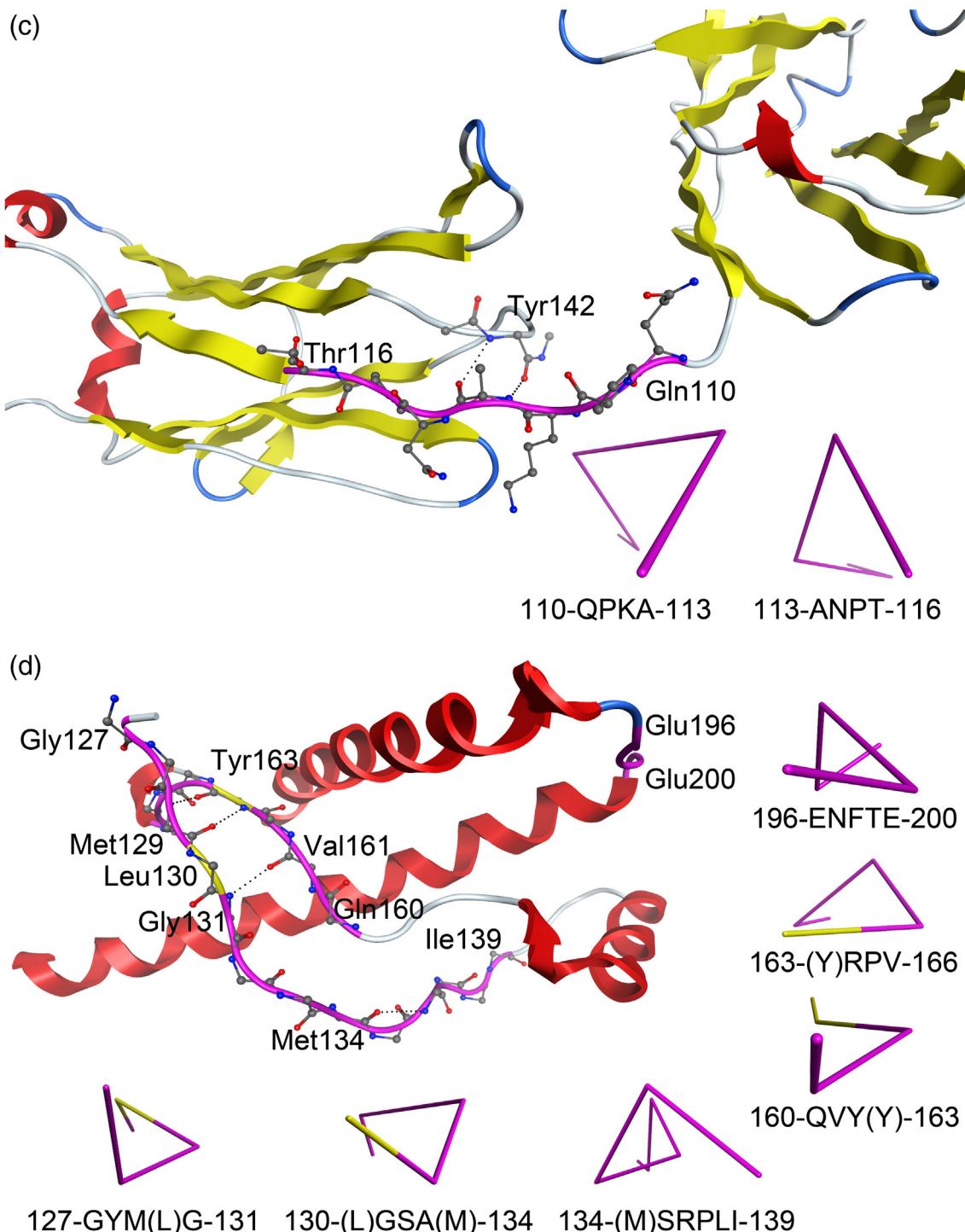


Fig. 4 (legend on page 2113)

conformational characteristics. The PPII conformation, including but not limited to residues within the PPII helices, populates a distinct cluster with the distribution maximum at $\phi, \psi = -65^\circ, 140^\circ$ of the Ramachandran plot and comprises 18–19% of all residues in the analyzed data sets^{3,6} (Fig. 3). If the coil regions in folded proteins are considered, the PPII content there is much higher, at the level comparable

with the α -helical conformation. If only stretches of PPII helices are included in the data set, the cluster they occupy is very similar with the ϕ, ψ maximum at $-75^\circ, 145^\circ$ ¹ (Fig. 1a). Depending on the length criterion, PPII helices encompass 2% ($l \geq 4$ residues) to 4% ($l \geq 3$ residues) of all amino acid residues.

In general, they are relatively short with the shortest formed by three residues and the majority

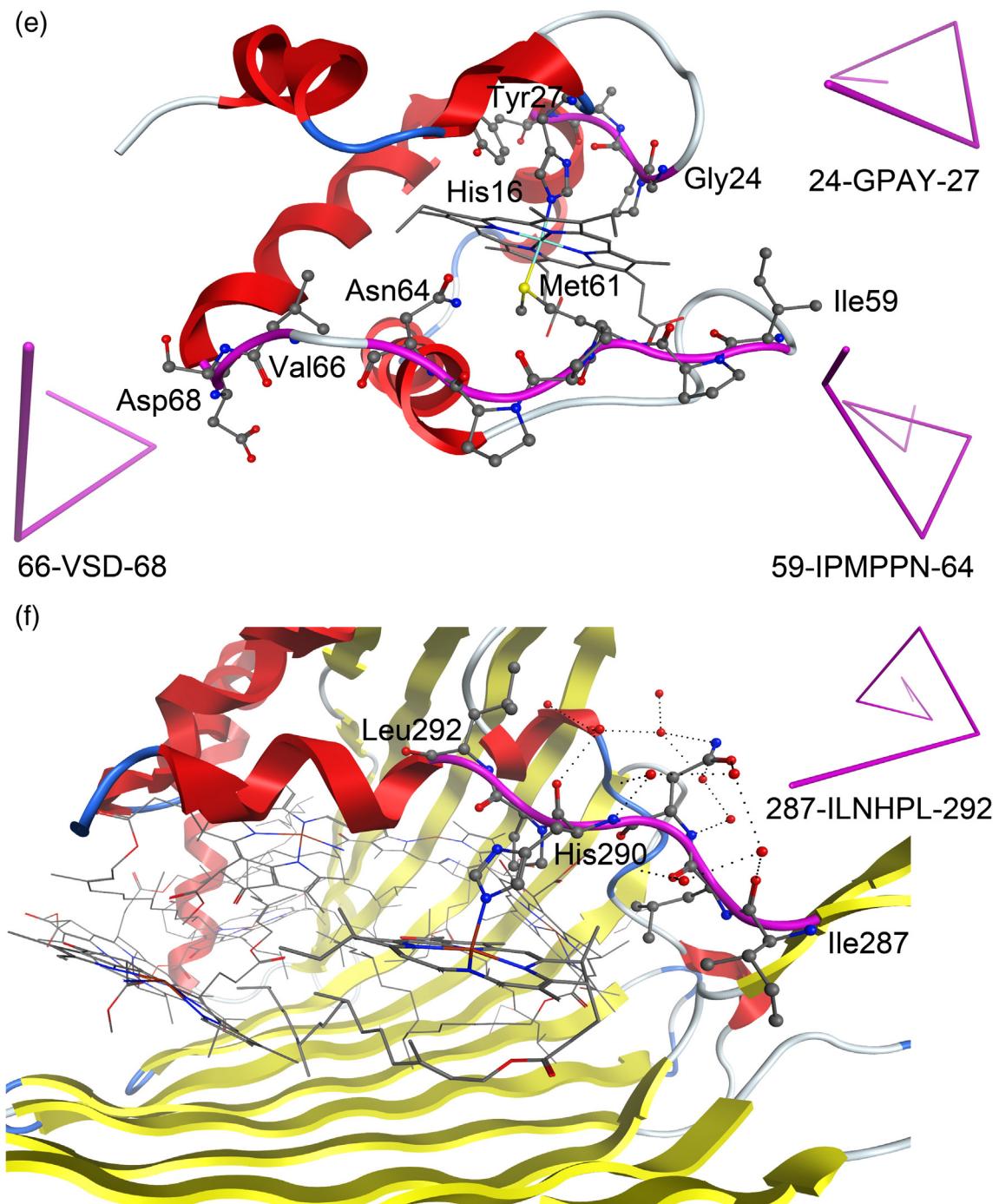


Fig. 4 (legend on page 2113)

formed by four to six residues, although longer PPII helices are also present in proteins.^{1,5,7} However, the PPII helix is an extended structure with the spatial length per residue twice the length of an α -helix, and a three-residue PPII helix has the same spatial length as a six-residue α -helix (Fig. 1).

It has been shown that longer PPII helices occur preferably in the exposed areas of folded proteins.^{1,5,7,18}

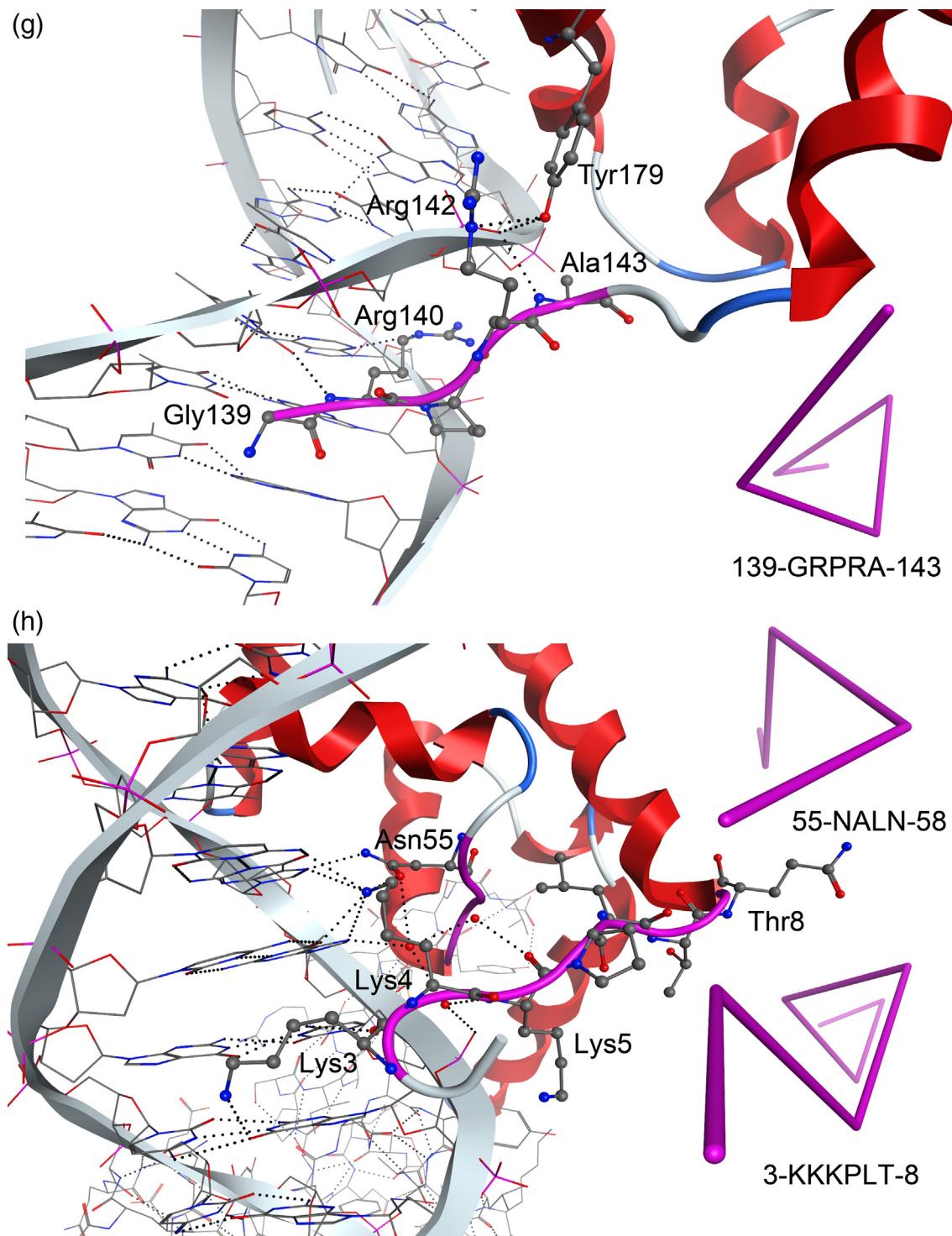


Fig. 4 (legend on next page)

Calculations of the solvent-accessible area for Ala polypeptides correctly predicted the non-proline residue preferences for the PPII over β -strand in coil libraries.⁷⁰ There is a positive correlation between the measured PPII content and simulated solvent-accessible surface.⁷³

PPII helices were found to be stabilized by predominantly non-local interactions. These interactions include mostly main-chain hydrogen bonds with water molecules, as well as side-chain–main-chain hydrogen bonds.⁷ Arginine forms most of such non-local hydrogen bonds, while the number of non-local and local hydrogen bonds is comparable for Ser and Thr, with less hydrogen bonds formed by Glu, Gln, and Asp. Gln forms side-chain–main-chain hydrogen bonds of the type $i, i + 1$ within the PPII helix.⁵ Local side-chain–main-chain interactions are relatively rare.⁷ This is generally in line with an earlier analysis of local hydrogen bonds in PPII,⁷³ although there are discrepancies due to the differences in the definitions of local hydrogen bonds and also possibly because of

different data sets and software used to assign hydrogen bonds. Subsequent study supported the results showing that conserved, buried residues distant in sequence from the PPII helices form side-chain hydrogen bonds to their main chain, resulting in “truss-like” structures that support PPII, with Arg being the most common residue, and that steric forces also play a part in stabilizing PPII in folded proteins.¹⁰⁴

PPII amino acid propensities

Protein structure surveys and experimental studies suggest that amino acid conformational propensities, including the PPII propensities, exist at the level of amino acid residues.^{14,16} Recent reviews by Schweitzer-Stenner and Elam *et al.* cover this area,^{35,105} allowing to examine here just a few relevant points.

PPII structure propensity scales were developed using the data of surveys of crystallographic protein structures deposited in the PDB including propensities derived from coil libraries^{1,3,5,7,14} and on the

Fig. 4. PPII helices in proteins. (a) SH3 domains of the Cbl-interacting protein CIN85 bound to the signal transduction protein ubiquitin ligase Cbl-b PPII helical peptide, formed by the proline–arginine-rich motif PXXXPR.¹¹⁸ The work reports peptide-induced dimerization of SH3 domains, where the Cbl-b peptide binds two molecules of the SH3 in opposite directions (class I and class II orientations, SH3-I and SH3-II in the figure), forming a heterotrimeric complex. This is reported to be the only example of such complex where SH3 domains do not interact with each other (in contrast with the SuperSH3 domain) and are linked solely by the PPII peptide. It is thought that CIN85 can serve as a binding platform for several Cbls, and it has been demonstrated in the same work that such Cbl-induced binding of the two CIN85 SH3 domains occurs *in vivo*. Image created from PDB 2BZ8. (b) PPII helix in calcium-binding EF-hand protein α -parvalbumin forms the linker connecting the CD and EF calcium-binding domains. The PPII helical linker 74-ARDLSA-79 (formed without proline) supports the salt bridge Arg75–Glu81. (Secondary structure and PPII helices were assigned in this study using DSSP secondary-structure assignment¹¹⁴ and the Adzhubei and Sternberg¹ criteria.) In the PPII helix, Ala74 and Arg75 form conserved hydrogen bonds with water molecules.¹¹⁹ A PPII– α -helix switch is formed by Ala79. Image created from PDB 1RWY. (c) PPII helix forming the linker between two structural domains in the light chain of immunoglobulin IgG1 Kol. The seven-residue-long PPII helix contains two prolines and is stabilized by the interchain Ala113–Tyr142 HN-CO and CO-NH hydrogen bonds creating a bend in the helix. The PPII– β -strand switch is formed by Thr116. Image created from PDB 2FB4. (d) PPII helices in prion PrPC protein. Most of the N-terminal part of the folded domain, up to the α -helix H1, adopts a conserved PPII structure, including the B1 strand. The PPII helix 127-GYM(L)GSA(M)SRPLI-139 (PPII-1) has two distortions forming bends, at Leu130 deviating into the β -region of the Ramachandran plot, and Met134. The B2 strand is flanked by the PPII 160-QVY(Y)RPV-166 (PPII-2), in which Tyr163 deviates to the β -region, and the first three residues have PPII conformations close to the boundary of the β -strand region¹ of the Ramachandran plot. Residues that deviate into the β -region of the Ramachandran plot are marked yellow. There is also a PPII helical linker 196-ENFTE-200 (PPII-3) between α -helices H2 and H3. Image created from the rabbit PrP PDB 3O79. (e) Cytochrome C551 of the c-type from *Pseudomonas aeruginosa* is a small electron transport protein structurally similar to the mammalian cytochrome c. The protein contains one of the longest observed PPII helices, 59-IPMPNN(A)VSD-68 with a distortion at Ala65. The helix includes one of the two heme-binding iron ligands, Met61,¹²⁰ and participates in covalent binding of the heme. PPII– α -helix structure switches are formed by Tyr27 and Asp68. Image created from PDB 451C. (f) The PPII helix in the bacteriochlorophyll A protein from the photosynthetic bacterium *Prosthecochloris aestuarii* is functionally similar to the helix in the cytochrome C551. The PPII helix 287-ILNHPL-292 incorporates the protein ligand His290 that binds one of the seven bacteriochlorophyll A molecules via the magnesium atom.¹²¹ The structure has a β -strand–PPII– α -helix supersecondary motif, with structure switches at Ile287 and Leu292 at the termini of the PPII helix. Image created from PDB 3EOJ. (g) Prokaryotic Hin recombinase complex with DNA recombination site. PPII helix 139-GRPRA-143 forms the N-terminal arm, which binds with the minor groove. Specific binding is achieved by combined minor and major groove binding, the latter by the helix–turn–helix motif.¹²² Residues Gly139, Arg140, and Pro141 are conserved in DNA invertases and are required for specific binding. Gly139 forms a van der Waals contact, and Arg140 and Arg142 form hydrogen bonds with the DNA molecule.¹²² Image created from PDB 1HCR. (h) Helix–turn–helix prokaryotic lambda repressor–operator complex. PPII helix 4-KKPLT-8 forms the N-terminal arm that binds the DNA major groove together with the helix–turn–helix motif. The N-terminal arm is essential for recognition specificity and affinity.¹²³ Residues at positions 3, 4, and 5 play a key role in the repressor function. While Lys3 at the edge of the PPII helix interacts with DNA, Lys4 side chain forms three hydrogen bonds with the DNA and protein, and Lys5 forms water-mediated hydrogen bond with DNA.¹²⁴ PPII– α -helix switch is formed by Thr8. Image created from PDB 1LMB.

basis of experimental studies of host–guest peptides. In the Gly-based host–guest peptides,^{55,56} Gly can be considered as a neutral host. Host–guest studies utilizing the Pro-based peptides reflect a bias towards PPII as exists in PRRs.^{16,62,106} However, Elam *et al.* developed a scale utilizing calorimetric data for the binding of SH3–Sos PPII helical peptide, where the peptide corresponding to Sos was used as a Pro-based host, and reported that the host context had no significant effect on the ranking of the propensities.¹⁰⁷

All 20 amino acid residues occur in the PPII structure but their propensities vary for the PPII conformation (residues occurring in the PPII area of the Ramachandran plot) and PPII helices of length three and more residues. Comparisons of propensities for the PPII helices and single residues in the PPII ϕ , ψ region of the Ramachandran plot showed statistically significant difference between the two.¹⁴ In other words, as stated by Stapley and Creamer, the PPII helices are not merely consecutively positioned PPII residues.⁵

There is evidence showing dissimilarity between the experimental Gly-based propensity scales and the propensities obtained from restricted coil libraries.³⁵ A significant divergence has been observed between the PPII helix propensities derived from the protein structure survey data and the PPII host–guest propensities.^{14,73}

Apart from hydration, tertiary interactions and neighbor effects play an important role in stabilizing PPII helices in the folded proteins.^{3,7,104} Therefore, it can be assumed that propensities for single-residue PPII conformation, PPII helices in folded proteins, and unfolded peptides in solution will differ. The PPII host–guest propensities would reflect mostly the PPII single-residue conformation rather than the regular PPII helix. Indeed, when measured as propensities of residue conformations restricted to the relevant ϕ , ψ region, experimental and statistically derived propensities of regular secondary structures were reported to agree.¹⁰⁸ In a similar approach, when stringent criteria were applied restricting ϕ , ψ angles to the narrow region of the corresponding regular structure, the coil libraries' propensities were reported to correlate with both the regular secondary-structure and experimentally derived propensities.³ Such stringent definition, however, inevitably reflects a restricted subset of the conformational class present in a coil library. This highlights the problem of different definitions of conformational classes in coil libraries.

PPII helices in folded proteins are affected by a combination of factors such as context and neighbor effects, side-chain–main-chain hydrogen bonds within the PPII helix, and tertiary interactions and solvation, which vary to a large extent. Averaging out these factors results in the observed weak PPII residue propensities shown in all surveys (with the

exception of Pro that has very strong PPII propensity). One important factor that decreases the PPII propensities of amino acid residues is that side-chain rotamer preferences do not play a major role in this structure.⁵ As a result, relatively weak PPII amino acid propensities within folded proteins in their turn contribute to the observed lack of correlation between the surveys and the data of host–guest experimental scales.^{5,14,73}

PPII structure prediction and assignment in experimentally determined structures

Weaker PPII propensities compared to the α -helical and β -structure secondary-structure classes render PPII prediction difficult. As a rule, predictions of the distribution of Ramachandran dihedral angles ϕ , ψ do not explicitly include PPII as a separate structure class. One such approach to predict coil residue distribution using neural network¹⁰⁹ however implicitly included PPII by assignment of structure bins corresponding to this conformation. There have been several attempts to expressly predict PPII helices. Neural network¹¹⁰ and support vector machine¹¹¹ predictions reached a relatively high 70–75% accuracy. Vlasov *et al.* developed a tetrapeptide-based method and reported accuracy of about 60%.¹¹² These results, however, are lower than those achieved for other secondary-structure classes.¹¹³ As a rule, PPII helix is not included as a separate structure class in prediction methods because of a limited success of the PPII structure prediction efforts. Though identified as important structural elements in homology modeling,¹⁸ PPII helices are only occasionally used in structure modeling.

Many PPII helices remain unassigned even though they are present in experimentally solved or modeled structures. This limitation hinders our understanding of protein structure and function mechanisms. The current assignment rate of PPII helices in the PDB and databases of modeled structures is very low because PPII helix is not considered as a standard secondary-structure class. Probably the most widely used secondary-structure assignment program, DSSP,¹¹⁴ employed for the PDB secondary-structure assignment, does not define PPII helices; neither are they included in the protein structure quality assessment programs. Several secondary-structure assignment methods that do include the PPII, such as XTLSSTR¹¹⁵ and SEGNO,¹¹⁶ produce somewhat divergent results. A step towards the introduction of a consensus PPII assignment has been taken in the approach that defines PPII helices based on the DSSP “coil” class. It is reported to provide the best possible overlap with the other secondary-structure assignment algorithms that include PPII.¹¹⁷ Ideally, the PPII helix should be incorporated in the standard secondary-structure classes assignment, with the

Table 1. PPII helix major function

Function associated with PPII	PPII structure	Protein and description	Structure identification method	Reference to PPII or PRR	PDB structure
<i>Structural function</i>					
PPII maintains three-dimensional structure in natively unfolded proteins and peptides.	PPII helix, PPII conformation of individual residues	<i>Natively unfolded proteins.</i> PPII is one of the predominant conformational states. Provides local order, flexibility, facilitates chain hydration.	CD VCD ROA FTIR NMR	Shi <i>et al.</i> ²⁰ Shi <i>et al.</i> ⁴¹ Zhu <i>et al.</i> ⁴⁷ Bochicchio and Tamburro ⁸²	
PPII maintains three-dimensional structure in folded proteins.	PPII helix, PPII conformation of individual residues	<i>Folded (globular) proteins.</i> PPII is one of the major regular structures. Provides extended flexible structural blocks, conserved, mostly exposed. Forms interdomain linker regions, N- and C-terminal regions, interactions sites. Forms linkers between other secondary-structure segments based on single-residue structure switches, changing the chain chirality.	X-ray NMR	Adzhubei and Sternberg ¹ Adzhubei and Sternberg ¹⁸ Siliardi and Drake ¹⁵ Stapley and Creamer ⁵ Bochicchio and Tamburro ⁸² Cubellis <i>et al.</i> ⁷ Berisio <i>et al.</i> ¹⁴	
PPII in interdomain linkers, structural domains.	PPII helix ^a (Pro+) 2FB4:110-QPKANPT-116 1FC1: 338-KAKGQPREGQ-347 2FL5: 109-PKAAPS-114 3FZU: 108L-RTVAAPS-114L 3FZU: 118H-SASTKGP-124H	<i>IgG1 FAB, light chain</i> (human). <i>IgG1 FC fragment,</i> <i>heavy chain (human).</i> <i>IgG-GAR</i> (yellow antibody), light chain. <i>IgG1 FAB,</i> <i>light chain, heavy chain.</i> PPII helix forms whole or part of interdomain linker.	X-ray	Adzhubei and Sternberg ¹⁸ This study ^a	PDB: 2FB4 PDB: 1FC1 PDB: 2FL5 PDB: 3FZU
PPII in interdomain linkers, functional domains.	PPII helix ^a (Pro-, conserved) 1RWY: 74-ARDLSA-79	<i>α-Parvalbumin</i> (rat). PPII forms linker between the CD and EF Ca^{2+} binding functional domains, contains Arg75 forming a single invariant salt bridge with Glu81. <i>β-parvalbumin</i> (avian thymic hormone, ATH) (chicken). PPII forms linker between functional domains.	X-ray	Adzhubei and Sternberg ¹ This study	PDB: 3F45 PDB: 1RWY PDB: 3FS7

(continued on next page)

Table 1 (continued)

Function associated with PPII	PPII structure	Protein and description	Structure identification method	Reference to PPII or PRR	PDB structure
PPII forms structural motifs.	PPII helix	<i>LRR proteins</i> . Proteins with short LLRs have PPII helices as structural elements on their convex side, PPII also forms cysteine-capping motif. <i>Amelogenin</i> . Tooth enamel protein, PRP, IDP. <i>Lampirin</i> . Extracellular matrix protein. <i>Elastin, titin, abductin</i> . Elastomeric proteins. PPII participates in and ensures protein self-assembly; PPII is a key structure in the elasticity mechanism of elastomeric proteins.		Matsushima <i>et al.</i> ¹²⁵ Park <i>et al.</i> ¹²⁶	
PPII is involved in protein self-assembly, elasticity.	PPII helix, PPII conformation	<i>Prion PrPC proteins</i> , <i>α-synuclein</i> , <i>tau protein</i> , <i>Alzheimer extracellular amyloid β peptide fragment 1–28</i> . Proteins involved in conformational disorders. High content of the PPII is identified, PPII possibly participates in pathogenic conformational changes.	NMR CD	Delak <i>et al.</i> ⁸⁶ Lakshminarayanan <i>et al.</i> ¹²⁷ Bochicchio and Tamburro ⁸² Bochicchio and Pepe ¹²⁸	
PPII participates in maintaining three-dimensional structure in the proteins associated with conformational diseases.	PPII helix	<i>Prion PrPC proteins</i> , <i>α-synuclein</i> , <i>tau protein</i> , <i>Alzheimer extracellular amyloid β peptide fragment 1–28</i> . Proteins involved in conformational disorders. High content of the PPII is identified, PPII possibly participates in pathogenic conformational changes.	CD ROA X-ray NMR	Syme <i>et al.</i> ⁸⁴ Blanch <i>et al.</i> ¹²⁹ Blanch <i>et al.</i> ⁵² This study	PDB: 1QLX PDB: 1QM0 PDB: 1QM2 PDB: 1AG2 PDB: 1B10 PDB: 1TQB PDB: 1TQC PDB: 1CU4 PDB: 1UW3 PDB: 1I4M PDB: 2W9E PDB: 3O79 PDB: 3HAK
<i>Protein–protein interactions</i> PRDs binding to peptide ligands that adopt PPII helical structure.	PPII helix peptide ligands (Pro+)	<i>Src tyrosine kinases and other SH3-containing proteins</i> . SH3 domain binding PPII helical protein ligands. <i>WW domain</i> . WW domain binding PPII helical protein ligands. <i>GYF domain</i> . GYF domain binding PPII helical protein ligands. <i>EVH1 domain</i> . EVH1 domain binding PPII helical protein ligands. <i>UEV domain</i> . UEV domain binding PPII helical protein ligands. Binding to PPII helical ligands is essential for PRD domain function.	X-ray NMR	Yu <i>et al.</i> ¹³⁰ Musacchio <i>et al.</i> ¹³¹ Nguyen <i>et al.</i> ¹³² Lim <i>et al.</i> ¹³³ Freund <i>et al.</i> ¹³⁴ Carl <i>et al.</i> ¹³⁵ Pornillos <i>et al.</i> ¹³⁶	

PRD profilin binding to PPII helix peptide ligands.	PPII helix peptide ligands (Pro+)	<i>Profilin</i> . Actin-binding regulatory protein involved in F-actin structure regulation. Binding to PPII helical ligand is essential for profilin function including profilin localization. <i>Synapsin-1</i> (human). Vesicle-associated phosphoprotein. PRR in the C-terminal region domain D. Binds actin, SH3 domain. <i>Dynamin-2</i> (human). Microtubule-associated GTPase.PRR in the C-terminal region. Binds SH3 domain. <i>cAMP-dependent protein kinase</i> (mouse, pig). Binding of the inhibitor peptide derived from naturally occurring inhibitor protein, to the catalytic loop.	X-ray	Mahoney <i>et al.</i> ¹³⁷	
PPII helix peptide ligands binding SH3.	PPII helix peptide ligands in PRR	<i>Synapsin-1</i> (human). Vesicle-associated phosphoprotein. PRR in the C-terminal region domain D. Binds actin, SH3 domain. <i>Dynamin-2</i> (human). Microtubule-associated GTPase.PRR in the C-terminal region. Binds SH3 domain. <i>cAMP-dependent protein kinase</i> (mouse, pig). Binding of the inhibitor peptide derived from naturally occurring inhibitor protein, to the catalytic loop.	X-ray	McPherson <i>et al.</i> ¹³⁸ Onofri <i>et al.</i> ¹³⁹ Gout <i>et al.</i> ¹⁴⁰	
PPII helix peptide ligands binding to enzyme.	PPII helix forms part (6 to 7 residues) of the peptide ligand (Pro-)	<i>Tyrosine-protein phosphatase non-receptor type 11</i> (Protein-tyrosine phosphatase SYP) (mouse). Binding of a PPII helical peptide ligand to the amino-terminal SH2 domain of Syp. <i>Serine proteinase inhibitor α-1-antichymotrypsin</i> (human, mouse). PPII helices form part of the reactive center loop (RCL).	X-ray	Siligardi and Drake ¹⁵ This study	PDB: 2CPK PDB: 1CDK PDB: 1FMO PDB: 3FJQ
PPII helix peptide ligands binding phosphatase SH2.	PPII helix peptide ligand (Pro+)	<i>Tyrosine-protein phosphatase non-receptor type 11</i> (Protein-tyrosine phosphatase SYP) (mouse). Binding of a PPII helical peptide ligand to the amino-terminal SH2 domain of Syp. <i>Serine proteinase inhibitor α-1-antichymotrypsin</i> (human, mouse). PPII helices form part of the reactive center loop (RCL).	X-ray	Siligardi and Drake ¹⁵ This study	PDB: 1AYA PDB: 1AYC
PPII structure in serpin reactive center loop.	PPII helix (Pro+)	<i>Serine proteinase inhibitor α-1-antichymotrypsin</i> (human, mouse). PPII helices form part of the reactive center loop (RCL).	X-ray	This study	PDB: 1QLP PDB: 1YXA
PPII structure in peptide epitope; antigen binding to MHC class II.	PPII helix antigen peptide (Pro-)	<i>Major histocompatibility complex (MHC) class II</i> . HLA-DR1. Peptides in the PPII conformation bind to MHC class II, presenting the peptides to T cells on the cell surface.	X-ray	Jardetzky <i>et al.</i> ¹⁴¹	PDB: 1SEB
PPII structure in peptide epitope; antigen binding to MHC class I.	PPII helix antigen peptide (Pro+)	<i>Major histocompatibility complex (MHC) class I</i> . HLA-B*1402 and HLA-B*2709 in complex with pCatA peptide. Peptide antigens bound to MHC I have different structure compared to MHC II. However, the Pro+ cathepsin-A signal sequence peptide binds in the PPII conformation.	X-ray	This study	PDB: 3BXN PDB: 3BP7

(continued on next page)

Table 1 (continued)

Function associated with PPII	PPII structure	Protein and description	Structure identification method	Reference to PPII or PRR	PDB structure
PPII structure in peptide epitope.	PPII helix ^a antigen peptide (Pro-) 1CU4: P109-MKHMA-P113	<i>Anti-prion FAB 3F4 in complex with its peptide epitope.</i> PPII helix in the peptide epitope. Epitope sequence corresponding to residues 104–113 of the Syrian hamster prion protein.	X-ray	This study	PDB: 1CU4
PPII helix binding to RRM motif.	PPII helix	<i>XlePABP2.</i> RNA-binding protein, poly(A) binding. <i>U2AF.</i> RNA-binding protein, splicing factor. RRM proteins. PPII helix binds RRM, which acts as a protein binding motif.	X-ray	Song <i>et al.</i> ¹⁴² Kielkopf <i>et al.</i> ¹⁴³	
<i>Protein–non-protein group binding</i>					
PPII helix supports protein ligand–heme covalent binding.	PPII helix ^a (Pro+) 451C: 59-IPMPPN-64 (Pro+, distorted) 2ZON: 59-KgaMPP-64	<i>Cytochrome c551</i> (<i>Pseudomonas aeruginosa</i> , <i>Achromobacter xylosoxidans</i>). PPII helix contains protein ligand Met, binding heme.	X-ray	Detlefsen <i>et al.</i> ¹²⁰ Adzhubei and Sternberg ¹ This study	PDB: 451C PDB: 351C PDB: 2ZON
PPII helix supports protein ligand–Bchl binding.	PPII helix ^a (Pro+) 3EOJ: 287-ILNHPL-292	<i>Bacteriochlorophyll A protein</i> (Bchl protein) (<i>Prosthecochloris aestuarii</i>). PPII helix contains protein ligand His binding bacteriochlorophyll.	X-ray	Adzhubei and Sternberg ¹ This study	PDB: 3EOJ
PPII helical regions are binding polyphenolic compounds tannins.	PPII helix PPII conformation	<i>Proline-rich salivary human protein PRP1, basic salivary proline-rich protein 4 (IB-5) and others.</i> IDPs. PRRs extend over most of the sequence, PPII is the probable structure of plant tannins binding sites.	FTIR NMR	Williamson ¹⁷ Elangovan <i>et al.</i> ⁹⁴ Pascal <i>et al.</i> ⁹⁶	
<i>Protein–nucleic acid interactions</i>					
PPII helix forms nucleic acid binding structural motifs.	PPII helix ^a (Pro+) 1HCR: 139-GRPRA-143 1LMB: 4-KKPLT-8 (Pro-) 9ANT: 6-QTYTR-10	<i>Hin recombinase, Antennapedia homeodomain.</i> DNA-binding proteins. PPII helical motifs bind DNA minor groove. <i>Lambda repressor protein.</i> DNA-binding protein. PPII helical motif binds DNA major groove.	X-ray	Hicks and Hsu ²³ Vlasov <i>et al.</i> ¹⁴⁴ This study	PDB: 1HCR PDB: 1LMB PDB: 9ANT

Pro+, PPII helix contains proline; Pro-, PPII helix without proline.

^a Secondary structure and PPII assigned in this study using DSSP¹¹⁴ and the Adzhubei and Sternberg criteria.¹

next step being the introduction of the PPII secondary structure in PDB entries.

Functional role of PPII

The role of PPII helices is twofold: they participate in maintaining the three-dimensional structure of proteins and support specific biological molecular functions. In folded proteins, PPII helices form flexible structural blocks immediately preceding α -helices,¹ in loops and interdomain linker regions, and in the N- and C-terminal regions^{17–19} (Fig. 4). PPII flexibility has been suggested to be its major feature underlying and ensuring its role in protein structure and its function in protein interactions.⁸² A representative list of major functions carried by the PPII structure is given in Table 1.

Depending on the available data, function assignment to a PPII helical region can be categorized as either (1) direct (e.g., based on the localization of a specific function to the PPII helix identified in the experimentally solved structure), (2) indirect (e.g., based on spectroscopic data showing high content of PPII in the structural entity responsible for the function), or (3) inferred (e.g., when the function is linked with a PRP or PRR where the PPII structure is highly probable).

A unique feature of the PPII is that a switch from the PPII helix to other regular structures can occur at a single residue and requires a shift of only one dihedral angle, forming structural motifs PPII- α -helix, β -strand-PPII, and PPII- β -strand observed in the structures of folded proteins.¹ Switching between the left-handed PPII and right-handed α - and 3(10)-helices occurs by shifting of dihedral angle ψ , resulting in the change of the chain chirality and transition from the extended PPII to α -helical conformation within one amino acid residue. The PPII- β -strand switch requires a relatively small shift of angle ϕ (Fig. 1). This fascinating feature of structural plasticity of the PPII helices is reflected in the way they are defined in several structure surveys—the first and last dihedral angles of a PPII helix are allowed to deviate from the PPII conformation.^{1,5,7} The PPII boundary residues acting as structure switches are partially incorporated in the hydrogen-bonding pattern of the adjacent structure.

The PPII helix thus acts as a flexible structured linker between helices and β -strands, or following after rigid β -strands, and entering the N-terminus of α -helices and 3(10)-helices. However, the α -helix-PPII single-residue structure switches were not found in proteins, and the α -helix-PPII structural motif always includes a short intervening linker. This is due to the existing steric restrictions revealed by the hard sphere repulsion modeling and the analysis of X-ray structures.¹⁴⁵ These restrictions preclude residues in β and PPII conformations from immedi-

ately following a stretch of α -helix formed by three or more residues. PPII helical interdomain linkers, structure switches, and motifs observed in protein structures are shown in Fig. 4.

Despite their generally flexible character, PPII helices as admittedly a major local structure in the natively unfolded proteins and peptides^{47,107} provide a degree of local structural rigidity. Assessment of the IDP structure, using the calorimetric host-guest PPII propensity scale, showed high PPII content in these proteins.¹⁰⁷ This analysis indicated a wide range of functions for the unfolded proteins with high PPII content including exceptionally high propensity for phosphorylation and showed that PPII regions were linked with phosphorylation sites. PRRs mostly also lack a high degree of order; large PRRs display various degrees of unfolding.¹⁷ Williamson analyzed PRRs and gave a detailed assessment of their function and relevant structural mechanisms.¹⁷ He concluded that PRRs are often involved in protein binding and sometimes play a structural role, but their function remains unclear in many cases.

An important PPII structural function is its involvement in elasticity and self-assembly processes. It was shown by isothermal titration calorimetry and variable-temperature CD that PPII participates in the self-assembly of the natively unfolded extracellular enamel matrix protein amelogenin.^{127,146} Earlier results from the Tamburro group showed that PPII helices are highly represented in elastomeric proteins elastin, titin, and abductin, as well as in the extracellular matrix protein lampirin, and that they are responsible for the self-assembly of elastin, lampirin, and possibly abductin.⁸² PPII helices interspersed by coil were identified by CD and NMR in the titin PEVK module.¹⁴⁷ It has been suggested that PPII can be responsible for intermolecular interactions resulting in fibrillar supramolecular structures.⁸² In elastomeric proteins titin and abductin formed by tandemly repeated sequences, as well as in elastin, PPII structures are partly responsible for entropic elasticity.⁸²

Tropoelastin is a large multidomain protein with the Gly-rich and Pro-rich hydrophobic domains and the Lys-rich/Ala-rich and Lys-rich/Pro-rich cross-linking domains. Covalently cross-linked tropoelastin molecules form elastin. Tamburro *et al.* have suggested that tropoelastin in water solution has prevalent PPII structure in equilibrium with unorderd conformation, which shifts to the β -turn and unorderd conformation when conditions change, for instance, in the less polar TFE.¹⁴⁸ The PPII structure was reported for both the Gly-rich and Pro-rich hydrophobic domains and the cross-linking Lys-rich/Pro-rich domains.^{128,148} A model of elastin elasticity has been proposed based on the conformational dynamic equilibrium between the extended PPII and β -strand conformations and the folded β -turn conformations, facilitated by the high PPII content in elastin.¹²⁸

Protein interactions

Perhaps the major and best known functional role of PPII structure is to mediate protein interactions via the PPII helical linear peptide ligands and their recognition domains. PPII was identified as the conformation of a wide range of protein ligands.^{15,82} Proline-containing motifs are most abundant in several proteomes.¹⁴⁹ Intrinsically disordered regions are often found involved in molecular recognition,¹⁵⁰ and PRRs as a part of such regions accommodate many of the PPII helix ligands. Some of them, however, are formed by shorter proline-rich sequences (PRSSs) or do not contain proline. The majority, although not all, of specific molecular PPII-mediated functions are based on the interactions involving PPII helices, as shown in Table 1.

The role of the PPII helix as a basic protein binding structural motif is the result of its extended structure and preferred location on the surface in folded proteins. The absence of fixed patterns of intra- or interchain hydrogen bonds makes PPII helix more flexible than α -helix or β -sheet, allowing conformational adjustments. The non-regular character of the existing hydrogen bonds stabilizing PPII allows it to form intermolecular ligand–receptor hydrogen bonds¹⁵ and favors it as a conformation responsible for protein–protein interactions.⁷ Another feature of the PPII important for its role as a structural motif of protein interactions is its tendency to form amphipathic helices.⁵

Though proline is not strictly necessary for the formation of PPII, its presence can provide additional important advantages. It is well established that protein–protein binding motifs are often enriched in proline,^{22,150} and PRSSs have high preference to adopt the PPII conformation.¹⁷ An important distinct feature of the PPII is its ability to bind rapidly and reversibly.^{17,151} It has been suggested that the PPII binding is favorable in terms of conformational entropy⁸² and supports low-affinity binding²² and that the low-affinity binding combined with high sequence specificity due to the presence of proline¹³² makes proline-based PPII favorable in dynamic complexes. It becomes essential in intracellular signaling where such complexes are assembled and disassembled as part of the signal transduction process.^{152,153} High-specificity and low-affinity binding is generally a feature of the intrinsically disordered regions,¹⁵⁰ which also provide a large available surface or platform for multiple interactions. PPII represents a major conformation in such regions; for example, SH3 binding motifs are conserved in intrinsically disordered regions.¹⁵⁴

PPII and proline-recognition domains

The PPII helix was found in protein ligands participating in major classes of interactions, such

as the Src tyrosine kinases SH3 domain-ligand,^{130,131,133,155} the MHCII–peptide epitopes,¹⁵⁶ WW domains,¹³² and profilin¹³⁷–protein ligands. Siligardi and Drake¹⁵ analyzed ligands modeled by corresponding peptides. They have concluded that for peptide ligands bound to their targets, there is a preference for extended structures, including the PPII. Their results indicated that the conformation of peptides to a large extent depends on the environment, specifically that peptide ligands adopt the PPII structure in the bound state. Dunker's group in their publications also suggested that, upon binding, some peptide ligands undergo transition from intrinsically unordered to ordered structure, which includes the PPII conformation.^{81,150} However, there is evidence from calorimetric and modeling results that SH3 ligands may have PPII conformation in both unbound and bound state.¹⁵⁷ The role of PPII in various protein interactions is outlined in Table 1.

Proline-rich ligands, including ligands with the PPII helical structure, interact with proline-recognition domains (PRDs). The role of PPII as the conformation responsible for interactions with PRDs is well known and widely studied to a large extent because of the important role of PRDs in signaling.¹⁵⁸ SH3 is one of the superfamily of such PRDs. It is one of the most abundant protein recognition modules in eukaryotes involved in a vast number of cellular functions ranging from proliferation and differentiation to cell death.^{159–162} PRDs that bind peptide ligands in the PPII helical conformation include SH3, the WW domain (two most abundant modules), EVH1,^{135,163} GYF,^{164,165} UEV,¹³⁶ and profilin. The CAP-Gly domain has topological similarities with SH3 and binds proline-rich ligands, though their exact conformation is not known.¹⁶⁶ Consensus recognition motifs differ for PRD families, but one of their common features is the presence of one or more prolines; PRDs also share the principal characteristics of the three-dimensional interactions with ligands.^{158,161}

SH3 mediates protein–protein interactions by binding short proline-rich ligand sequences, commonly the PXXP (XPXXX) PPII helical motif and other proline-based^{22,167} as well as non-proline motifs.¹⁶² (Reviewed by Kay *et al.*,²² Musacchio,¹⁶⁸ and Rath *et al.*²¹) In Src and c-Abl tyrosine kinases, intramolecular binding between the SH3 domain and the PPII helix located in the SH2–kinase linker region and the SH2 domain interaction with the C-terminal phosphotyrosine cause inactivation of the enzyme.^{169–171} The release of these interactions and binding to external ligands increase the kinase activity.¹⁷² SH3 interactions with the proteins containing PPII ligands together with the SH2 interactions with specific proteins can serve as a mechanism of activation, inactivation, and targeting of the Src and other kinases to substrates and subcellular compartments.^{173,174} The structural model for the

common SH3–XPXXP recognition¹³³ and the mechanism of WW-ligand interaction proposed by Lim *et al.*¹³² specifically state that the ligand should have the PPII structure. In an interesting example (Fig. 4a), SH3 domains of the Cbl-interacting protein CIN85 bind the signal transduction protein ubiquitin ligase Cbl-b PPII helical peptide, formed by the proline–arginine-rich motif PXXXPR.¹¹⁸

WW domains are found in many signaling and structural proteins including RNA-binding proteins. They also act by mediating protein–protein interactions. While the PPII helical binding motifs interacting with WW domains are different from the PPII motifs interacting with SH3 and employ a different mechanism of ligand recognition,¹³² the SH3 and WW binding sites have structural similarity¹⁷⁵ and display elements of functional similarity.^{132,176}

Profilins are small proteins interacting with three types of ligands: the proline-rich PPII helical peptide ligands,¹³⁷ actin monomers and actin-related proteins, and phosphatidylinositol lipids in the cell membrane.^{177–179} PPII helix binding sites are located in the N- and C-termini of profilins, and PPII binding is essential for the functioning of the majority of profilins,^{179,180} allowing them to co-localize with other partners. They are versatile proteins conserved at the structural level and implicated in actin assembly dynamics and formation of the actin cytoskeleton.¹⁸⁰ Profilins interact with a number of mammalian proteins and are engaged in the regulation of actin-based processes, membrane trafficking, and nuclear transport, but their exact role has not been established.¹⁸¹

Proline-rich regions

Generally, PRRs can be classified into three categories: non-repetitive PRRs that possess a wide range of functions including SH3 binding, short repetitive PRRs with structural and protein binding functions, and longer tandemly repeated PRRs, such as in salivary proteins, which are also involved in binding.^{5,17} Protein ligands with PRSSs involved in protein–protein interactions and forming PPII helices can be classified according to their counterpart PRDs, type of sequence motifs, and structural mechanism of interaction.^{22,153,158}

Some PRRs, usually shorter proline-rich PPII motifs, are structurally well characterized, and the mechanisms of their interaction with PRDs are known.¹⁵⁸ Larger PRRs are often intrinsically disordered¹⁵⁰; for some, there is NMR and other spectroscopic evidence of their forming PPII structure. In the absence of direct structural data, the presence of PPII and its functional role can only be conjectured for PRSSs on the basis of their amino acid composition. As an example of such conjecture, a PPII helical structure can be suggested for the novel glycosylation motif responsible for a recently discov-

ered O-linked glycosylation system in bacteria.¹⁸² Mammalian O-linked glycosylation sites were shown to be enriched in Pro, Thr, Ser, and Ala.¹⁸³ It has been shown that O-linked glycosylation in bacteria is associated with an alanine-, serine-, and proline-rich, low-complexity region near the protein N-termini and specifically with a sequence motif SAPA.¹⁸² Such sequence should have a high PPII propensity.

Though PRRs are likely to contain PPII helices, this cannot be arbitrarily extended to all functionally important PRSSs. Precise structure information on PRRs is difficult to obtain as they are often missing from the crystallographic data, and suggestion of the PPII structure for specific functional sites based solely on their sequence would be inevitably tentative. In one such example where crystallographic structure is available, tandemly repeated PRSSs form the functionally important motifs of the C-terminal domain of RNA polymerase II. Though the PPII helical structure is probable for such sequence, the motif SPTS in the phosphorylated C-terminal domain peptide, responsible for binding the protein Pcf11 CID domain, was shown by crystallography to adopt the β-turn structure.¹⁸⁴

An important functional role of proline-rich motifs is demonstrated by structural data for a number of proteins. An unusually (eight-residue) long, non-repetitive, proline-rich PPII helix has been identified in the crystallographic structure of exonuclease TREX1, a protein involved in the processing of DNA ends. It has been suggested that the PPII helix located on the protein surface acts as a binding motif responsible for the interaction of TREX1 with SET protein.¹⁸⁵

Myelin basic protein is an unusual multifunctional protein associated with myelin and involved in maintaining myelin sheath and an array of complex functions in the mammalian central nervous system. It is mostly intrinsically disordered and has been shown by spectroscopic methods to contain an SH3-binding motif that adopts the PPII helical structure.^{186,187} By binding proteins via their SH3 domains, myelin basic protein can connect these proteins to the membrane and thus participate in transmembrane signal transduction.¹⁸⁸ Another IDP, the human salivary PRP IB-5, has been shown by CD and NMR to contain PPII, with the PPII helices likely to be involved in binding tannins—a protection mechanism against food tannins.⁹⁶

It has been demonstrated that some proline-rich PPII helical motifs can be associated with cellular processes leading to cancer and involved in infectious diseases.¹⁵³ For instance, Cas is an adaptor protein involved in the generation of certain cellular responses, including cell migration and integrin signaling, and plays a role in oncogenic processes.¹⁸⁹ It has been extensively characterized as a substrate of Src family kinases. Cas has an Src binding sequence near its C-terminus that contains a ligand for the SH3

domain (RPLPSPP) followed by a motif that, when phosphorylated, can interact with the SH2 domain. The PPII helical segment of Cas is essential for its function and its disruption leads to the loss of Src binding, in turn, affecting Cas phosphorylation.¹⁹⁰ The importance of the PPII segment for the function of Cas has been further confirmed by mutation studies.¹⁹¹

PPII helices have been shown to mediate protein–protein host–pathogen interactions.¹⁹² The PPII structure is present in the HIV protein Nef, one of the key proteins responsible for HIV infectivity, which binds SH3 via a consensus binding PPII motif (PXXP).¹⁹³ This motif is highly conserved in the HIV and SIV Nef.^{194,195} Nef–SH3 binding is essential in the life cycle of the virus,¹⁹⁶ and the PPII motif is a key determinant of the Nef ability to interact with its partners.¹⁹⁵ HIV Gag and Ebola Vp40 matrix proteins contain the PTAP peptide, binding the human protein Tsg101 UEV domain as part of the process necessary to facilitate virus budding. The PTAP peptide structure is a PPII helix flanked by extended conformation.¹⁹⁶ Binding to proline-rich ligands was suggested to be a factor of allergenicity of a major allergen, the birch pollen profilin, as the IgE-reactive epitopes overlap its binding sites interacting with proline-rich ligands.¹⁹⁷

In *Listeria monocytogenes*, a major pathogen responsible for food infections, ActA protein interacts with the EVH1 domain of vasodilator-stimulated phosphoprotein via its tandemly repeated PRS motif, which facilitates cytoskeleton actin motility and is essential for bacteria virulence.¹⁹⁸ It has also been shown that a polyproline motif plays a key role in the pathogenesis of the enteropathogenic *Escherichia coli* (EPEC), which causes intestinal inflammation and severe diarrhea.¹⁹⁹ The EPEC virulence factor Tir has a proline-rich SH3 binding motif that interacts with the SH3 domains of the host Tec and Abl tyrosine kinases; in addition, it has a phosphorylation site that interacts with kinase SH2 domains. These interactions lead to formation of the membrane protrusions, which are essential for the progression of the EPEC infection.¹⁹⁹

PPII helices with and without proline

PPII helices that include prolines can be located in proteins and regions with amino acid composition not enriched in proline, where the term proline-containing PPII helices seems more appropriate. A distinction should be made between proline-containing PPII helices and PPII within PRRs.

For some peptide ligands adopting the PPII structure, for instance, T-cell peptide epitopes of the MHC class II, proline is not mandatory. Accordingly, proline-containing PPII can be designated as Pro+, and PPII without proline, as Pro-. Antigenic

peptides bound to MHC class II invariably form PPII helix.^{141,156,200,201} The PPII helix in the cleft forms hydrogen bonds between the peptide backbone and protein side chains.¹⁴¹ In MHC class I molecules, the peptide binding mechanism is different from the MHC II, and the peptides generally bind in a different conformation. Still, a Pro+ peptide ligand bound to MHC I has the PPII helical structure (see Table 1).

In cytochrome C551, a Pro+ PPII helix contains the methionine residue that binds the heme^{1,120} (Fig. 4e). Similarly, the Pro+ PPII helix in bacteriochlorophyll A protein contains one of the chlorophyll-binding histidines (Fig. 4f). In fact, the PPII structure is abundant in both bacteriochlorophyll A protein and cytochrome C551, and the latter contains one of the longest known PPII helices.

In some proteins, PPII helices form interdomain linkers.¹⁸ For instance, Pro–PPII helix forms a linker between functional calcium-binding domains in α -parvalbumin (Fig. 4b). The conserved salt bridge important for the structure stability²⁰² is supported by this PPII helix, which includes residue Arg forming the salt bridge. A Pro+ PPII helix also connects structural domains in the light chain of immunoglobulin IgG1 Kol (Fig. 4c).

PPII helices form nucleic acid-binding structural motifs. Some DNA-binding motifs reported earlier as extended structures by Siligardi and Drake¹⁵ have been assigned in this study as PPII helices using DSSP¹¹⁴ secondary-structure assignment and the PPII helix criteria of Adzhubei and Sternberg.¹ These PPII helices are located in the N-terminal arms of the Hin recombinase (Pro+) and antennapedia homeodomain (Pro–), both binding in the minor groove of DNA,^{122,203} and the N-terminal arm of the lambda repressor protein (Pro+), binding in the major groove^{123,124} (Table 1 and Fig. 4). A survey of structures of nucleic acid-binding proteins showed that PPII helices are often involved in protein interactions with DNA and RNA, being commonly located at the interaction sites and forming specific and non-specific hydrogen bonds.²³ These findings are supported by the results of analysis of the structures of protein linker regions interacting with nucleic acids.¹⁴⁴

Functional domains of the poly(A)-binding protein XlePABP2 form heterodimers, in which the N-terminal PPII helix interacts with the RNA recognition motif (RRM) of the protein and occludes it. Upon binding of RNA, this interaction is abolished and the PPII helix is released into the solvent.¹⁴² Similarly, the RNA-binding splicing factor U2AF forms heterodimers, and such a dimerization occurs through the interaction of an atypical RRM in the smaller domain U2AF35 and PPII helix of the UA2F65 domain via reciprocal tryptophan recognition resembling the SH3-ligand Trp interaction.¹⁴³ In both proteins, the RRM acts as a protein-binding motif interacting with the PPII helix.

Finally, PPII helices have an important structural function in leucine-rich repeat (LRR) proteins. Proteins with short LRRs have PPII helices as structural elements on their convex side¹²⁵ replacing the α -helix of longer LRRs, or a combination of PPII and other structures such as 3(10)-helix and β -turn.²⁰⁴ PPII helices can form interlocking structures, filling the openings on the LRR solenoid convex side and thus protecting the hydrophobic core of the concave side.²⁰⁴ PPII helices were also identified in the cysteine-capping motif, protecting the hydrophobic core in small LRR proteins and proteoglycans.¹²⁶

PPII in proteins involved in conformational disorders

Transmissible spongiform encephalopathies or prion diseases are triggered, according to the protein-only hypothesis by structural changes in the prion protein PrP.^{205,206} The soluble cell surface glycoprotein form of PrP, the PrPC, undergoes dramatic structural changes, converting it into the protease-resistant infectious protein PrPSc. PrPC is a predominantly α -helical protein comprising α -helices H1, H2, and H3, and two short β -strands B1 and B2 forming a distorted β -sheet. Besides, PrPC has an extended, partially unstructured N-terminal domain, with four octapeptide repeats PHGGGWGQ. The N-terminal domain contains PPII helices interspersed with β -turns as identified by spectroscopic methods.^{129,207,208} Recent studies suggest that PrPSc is mostly formed by β -structure, which incorporates the N-terminus, with the α -helices H2 and H3 linked by a disulfide bridge. The structure adopts a parallel left-handed helical β -fold. Crystallographic studies of PrPSc are hampered by its insolubility, and the information about its structure is based on the combination of various data and methods, including electron and atomic force microscopy, low-resolution diffraction of various amyloid fibrils, analysis of known protein structures and structure modeling, and IR spectroscopy.^{207,209} Recently, a model of the prion HET-s protein amyloid fibrils from fungus *Podospora anserina* has been suggested on the basis of solid-state NMR data to be a solenoid structure, in which the fibril core is formed by in-register parallel β -sheets.^{210,211}

Although the N-terminal domain of prion proteins is not explicitly required for the formation of infectious amyloid form, its importance in the PrP function has been gradually recognized. The findings of Gill *et al.* suggest that the PPII structure is present in the partially unstructured N-terminal domain of the prion protein PrP from the Chinese hamster ovary cells and brain cells of mice *in vivo*. It serves as a signal for proline 4-hydroxylation, which is important for regulating PrP activity and may occur in other prion proteins.²⁰⁷ In a number of works by the Barron

group, the PPII helix was identified as an intermediate structure in a critical conformational transition in human lysozyme, a ubiquitous protein involved in systemic amyloidoses. This transition from an α -helix to a PPII helix precedes the formation of β -sheet and leads to aggregation into β -fibrils.⁵² The authors argue that the PPII structure forms the "static type of disorder",²⁸ which can play a crucial role in aggregation of human lysozyme into amyloid fibrils. Such an aggregation can have a wider implication for proteins undergoing similar amyloidogenic processes, such as prions, where PPII potentially can induce fibrinogenesis.⁵²

Natively unfolded proteins have a prevailing PPII structure, and it has been suggested that in natively unfolded proteins, the amino acid composition favoring PPII can also ensure compensation mechanisms protecting such proteins from fibrinogenesis.⁴⁷ Denatured reduced proteins such as the truncated (94–233) sheep PrP, according to spectroscopic data, have a lower content of PPII but a high propensity to form β -strand.⁴⁷ It can be assumed that the presence of PPII in proteins related to conformational diseases will indicate their healthy state, and evidence of conformational transition of PPII to β -structure would point to the possibility of fibrinogenesis. ROA spectra indicate a high content of the PPII in the natively unfolded brain protein α -synuclein that in the misfolded state forms fibrils associated with Parkinson's disease. Similar data were obtained for the tau protein located primarily in neurons in the central nervous system, which in its pathological hyperphosphorylated state forms neurofibrillary tangles characteristic for Alzheimer's disease.⁸⁴

Examination of the crystallographic and NMR structures of the folded domain of several prion PrPC proteins (using DSSP¹¹⁴ secondary-structure assignment and the PPII helix criteria of Adzhubei and Sternberg¹) performed in this study revealed conserved PPII helices forming most of the N-terminal part of the structures, up to the α -helix H1. The short two-residue strand B-1 located in the N-terminal PPII region adopts PPII conformation and is flanked on both sides by the PPII helices. The two-residue strand B-2 is also followed by the PPII helix (Fig. 4d). The occurrence of PPII residues in β -sheets is not unusual,³ and in each of these strands, one residue deviates towards the β -region of the Ramachandran plot supporting the interchain β -structure hydrogen bonds, and the other residue has the PPII conformation.

This suggests that PPII can have an important functional role in the conformational transition to the PrPSc structure. The N-terminal PPII helix is located immediately before and partially overlaps with the putative aggregation initiation site suggested for prions on the basis of analysis of aggregation properties of peptides, corresponding

to the sequence region 106–157 of the human prion protein.²¹² The second PPII helix lies in the region that potentially can play the key role in conformational transition from PrPC to PrPSc.^{213–215} This PPII helix borders the YYR motif identified as a PrPSc-selective epitope.²¹⁶ PPII helix is close to the β -strand in the conformational space, and the conformational transition PPII– β -strand requires a shift of only one torsional angle, ϕ (Fig. 1). Therefore, comparatively minor conformational changes in the PPII structure can trigger the formation of β -sheet.⁸⁴ It is possible that the normal PrPC structure is maintained by the presence of the PPII helical structure. The relative positions of the PPII helices in PrPC (Fig. 4d) show that if their conformation shifts to β -strands, a structural rearrangement can result in the formation of a long β -sheet, possibly triggering conformational changes leading to the PrPSc structure.

Conclusions

The left-handed PPII helix is a distinct structural class that constitutes one of the three ubiquitous regular secondary-structure classes in folded proteins—right-handed helices, including α -helix and 3(10)-helix, β -sheet, and the left-handed PPII helix.^{1–3,7,105} The designation “PPII helix” is somewhat misleading, since proline is not mandatory for this structure (though it has the highest occurrence in these structures among all amino acids), as up to 46% of PPII helices in folded proteins contain no proline.⁷

PPII helices occur on a much lesser scale than the α -helices and β -structure: depending on the length criterion, they encompass 2% (four or more amino acids in length) to 4% (three or more amino acids) of all residues. This makes PPII helices the fourth most abundant regular secondary structure in folded proteins, following the α -helix, β -structure, and 3(10)-helix. The occurrence rate of shorter PPII helices (three or more residues) is comparable with that of the 3(10)-helices, which also mostly form short one-turn helices. Other regular secondary structures—the left-handed α -helix and π -helix—are rare.^{217,218} In unfolded proteins and polypeptides, PPII is identified by spectroscopic methods as a predominant conformation supporting local order in polypeptide chain, though α -helical and β -conformations are also present.^{20,21,41}

A striking feature of the PPII is that it is a well-defined regular structure with repetitive torsional ϕ , ψ angles, and yet devoid of the regular patterns of main-chain intrachain or interchain hydrogen bonds that stabilize other regular structures such as α -helices and β -structure. The main stabilizing factor and hydrogen-bonding partner for the PPII in many cases is water. The absence of rigid hydrogen-bonding patterns ensures another unique and

functionally important quality of the PPII among regular structures, its flexibility. Its extended conformation and the absence of regular hydrogen bonds make PPII an ideal structure for a wide range of molecular interactions.

A characteristic feature of the major secondary-structure classes is that they occur with high frequencies both as distinctive clusters of conformations of individual residues and as regular or periodic structures. PPII differs from α -helix and β -sheet in the way that its distribution frequency in folded proteins is shifted towards the prevalence of PPII conformation. PPII helices and PPII conformation dominate coil regions of folded proteins,³ which by their conformational composition resemble unfolded proteins.

PPII helices are directly involved or mediate a wide range of molecular functions including but not limited to signaling, transcription, cell motility, immune response, elasticity of elastomeric proteins, and possibly conformational transitions in amyloid proteins.

PPII has a major structural role in both folded and unfolded proteins. An important structural function of PPII helices is to form flexible blocks in the folded protein structure.^{18,82} PPII supports numerous protein–protein and protein–nucleic acid interactions where PPII helices act as a peptide ligand or form structural binding motifs.^{15,22,23} PPII structures represent possibly the most widely spread binding motif in proteins. Some of the PPII functions are as yet undiscovered, since there is no conclusive information on the structure of many functionally important PRRs where PPII is likely to be the predominant conformation, judging by the amino acid composition of PRRs.

Unlike other secondary structures, most of the PPII helices identified in proteins are linked to a specific function. For some functions, participation of the PPII structure has been inferred from location or amino acid composition of the functional sites. Therefore, while α -helices and β -sheets represent “building blocks” of protein structure, PPII helices can be designated as “functional blocks” or a “functional secondary structure”. Functionally important PPII helices described in this review have not been meant to represent an exhaustive list but rather to show examples of their versatility. A detailed survey of the functional role of PPII helices, covering all available structural data in folded and unfolded proteins, would be of great benefit for the understanding of the protein structure–function relationship.

Prediction of the PPII structure is a difficult task because of the elusive nature of amino acid propensities to adopt this structure. PPII amino acid preferences calculated in a number of statistical surveys consistently appear to be weak compared to α -helix and β -structure,^{1,7,14} differ between surveys, and can depend on the data set and criteria applied to define PPII.¹⁴

PPII helices are conserved in protein structure and can be useful in structure modeling and fold recognition methods.^{7,18} Despite its common occurrence and functional importance, PPII helix remains a ghost structural class in terms of formal classification, at least in folded proteins, where it rarely appears in secondary-structure annotations. In a crucial omission, it remains outside of the PDB standard secondary-structure classes. The reason for this is the fact that PPII is not assigned by the DSSP secondary-structure annotation.¹¹⁴ Its inclusion in the DSSP and PDB secondary-structure classes would solve this problem. Just as useful would be the availability of structural data including PPII in the databases of unfolded proteins.^{219,220}

Acknowledgements

This work was supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences and by State Contract No. 11.519.11.2008. Computations were performed at the EIMB RAS "Genome" center supported by State Contract No. 16.552.11.7069 with the Russian Ministry of Education and Science. We thank Vladimir Chekhov for designing and preparing figures.

Received 26 November 2012;

Received in revised form 28 February 2013;

Accepted 11 March 2013

Available online 16 March 2013

Keywords:

PPII conformation; extended left-handed helix; secondary structure; protein interactions; folded and unfolded proteins

Abbreviations used:

PPII, poly-L-proline type II; VCD, vibrational circular dichroism; ROA, Raman optical activity; MD, molecular dynamics; IDP, intrinsically disordered protein; PRS, proline-rich sequence; PRR, proline-rich region; PRP, proline-rich protein; PRD, proline-recognition domain; PDB, Protein Data Bank; PGA, poly-L-glutamic acid; PL, poly-L-lysine; IR, infrared; FTIR, Fourier transform infrared; EPEC, enteropathogenic *Escherichia coli*; RRM, RNA recognition motif; LRR, leucine-rich repeat.

References

- Adzhubei, A. A. & Sternberg, M. J. (1993). Left-handed polyproline II helices commonly occur in globular proteins. *J. Mol. Biol.* **229**, 472–493.
- Hollingsworth, S. A., Berkholz, D. S. & Karplus, P. A. (2009). On the occurrence of linear groups in proteins. *Protein Sci.* **18**, 1321–1325.
- Jha, A. K., Colubri, A., Zaman, M. H., Koide, S., Sosnick, T. R. & Freed, K. F. (2005). Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry*, **44**, 9691–9702.
- Fitzkee, N. C., Fleming, P. J. & Rose, G. D. (2005). The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins: Struct. Funct. Bioinform.* **58**, 852–854.
- Stapley, B. J. & Creamer, T. P. (1999). A survey of left-handed polyproline II helices. *Protein Sci.* **8**, 587–595.
- Adzhubei, A. A., Eisenmenger, F., Tumanyan, V. G., Zinke, M., Brodzinski, S. & Esipova, N. G. (1987). Third type of secondary structure: noncooperative mobile conformation. Protein Data Bank analysis. *Biochem. Biophys. Res. Commun.* **146**, 934–938.
- Cubellis, M. V., Caillez, F., Blundell, T. L. & Lovell, S. C. (2005). Properties of polyproline II, a secondary structure element implicated in protein–protein interactions. *Proteins*, **58**, 880–892.
- Arnott, S. & Dover, S. D. (1968). The structure of poly-L-proline II. *Acta Crystallogr., Sect. B*, **24**, 599–601.
- Shoulders, M. D. & Raines, R. T. (2009). Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958.
- Makarov, A. A., Esipova, N. G., Pankov, Y. A., Lobachev, V. M. & Grishkovsky, B. A. (1975). A conformational study of beta-melanocyte-stimulation hormone. *Biochem. Biophys. Res. Commun.* **67**, 1378–1383.
- Makarov, A. A., Esipova, N. G., Lobachov, V. M., Grishkovsky, B. A. & Pankov, Y. A. (1984). Secondary structure of polypeptide hormones of the anterior lobe of the pituitary gland. *Biopolymers*, **23**, 5–22.
- Makarov, A. A., Lobachov, V. M., Adzhubei, I. A. & Esipova, N. G. (1992). Natural polypeptides in left-handed helical conformation. A circular dichroism study of the linker histones' C-terminal fragments and beta-endorphin. *FEBS Lett.* **306**, 63–65.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
- Berisio, R., Loguerio, S., De Simone, A., Zagari, A. & Vitagliano, L. (2006). Polyproline helices in protein structures: a statistical survey. *Protein Pept. Lett.* **13**, 847–854.
- Siligardi, G. & Drake, A. F. (1995). The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. *Biopolymers*, **37**, 281–292.
- Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F. & Creamer, T. P. (2001). Host–guest study of left-handed polyproline II helix formation. *Biochemistry*, **40**, 14376–14383.
- Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260.
- Adzhubei, A. A. & Sternberg, M. J. (1994). Conservation of polyproline II helices in homologous

- proteins: implications for structure prediction by model building. *Protein Sci.* **3**, 2395–2410.
19. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208.
 20. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **62**, 163–240.
 21. Rath, A., Davidson, A. R. & Deber, C. M. (2005). The structure of “unstructured” regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers*, **80**, 179–185.
 22. Kay, B. K., Williamson, M. P. & Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* **14**, 231–241.
 23. Hicks, J. M. & Hsu, V. L. (2004). The extended left-handed helix: a simple nucleic acid-binding motif. *Proteins*, **55**, 330–338.
 24. Woody, R. W. (1992). Circular dichroism and conformation of unordered polypeptides. *Adv. Biophys. Chem.* **2**, 37–79.
 25. Makarov, A. A., Adzhubei, I. A., Protasevich, I. I., Lobachov, V. M. & Espipova, N. G. (1993). Scanning microcalorimetry and circular dichroism study of melting of the natural polypeptides in the left-handed helical conformation. *J. Protein Chem.* **12**, 85–91.
 26. Sreerama, N. & Woody, R. W. (1994). Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. *Biochemistry*, **33**, 10022–10025.
 27. Keiderling, T. A. (2002). Protein and peptide secondary structure and conformational determination with vibrational circular dichroism. *Curr. Opin. Chem. Biol.* **6**, 682–688.
 28. Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vasák, M. & Barron, L. D. (2001). Solution structure of native proteins with irregular folds from Raman optical activity. *Biopolymers*, **58**, 138–151.
 29. McColl, I. H., Blanch, E. W., Hecht, L., Kallenbach, N. R. & Barron, L. D. (2004). Vibrational raman optical activity characterization of poly(L-proline) II helix in alanine oligopeptides. *J. Am. Chem. Soc.* **126**, 5076–5077.
 30. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl Acad. Sci. USA*, **99**, 9190–9195.
 31. Tiffany, M. L. & Krimm, S. (1972). Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers*, **11**, 2309–2316.
 32. Tiffany, M. L. & Krimm, S. (1968). New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers*, **6**, 1379–1382.
 33. Krimm, S. & Tiffany, M. L. (1974). The circular dichroism spectrum and structure of unordered polypeptides and proteins. *Israel J. Chem.* **12**, 189–200.
 34. Tiffany, M. Lois & Krimm, Samuel (1969). Circular dichroism of the “random” polypeptide chain. *Biopolymers*, **8**, 347–359.
 35. Schweitzer-Stenner, R. (2012). Conformational propensities and residual structures in unfolded peptides and proteins. *Mol. Biosyst.* **8**, 122–133.
 36. Krimm, Samuel, Mark, J. E. & Tiffany, M. Lois (1969). Influence of counterions on the helical conformations of charged polypeptide chains. *Biopolymers*, **8**, 695–697.
 37. Drake, A. F., Siligardi, G. & Gibbons, W. A. (1988). Reassessment of the electronic circular dichroism criteria for random coil conformations of poly(L-lysine) and the implications for protein folding and denaturation studies. *Biophys. Chem.* **31**, 143–146.
 38. Dukor, R. K. & Keiderling, T. A. (1991). Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers*, **31**, 1747–1761.
 39. Kuemin, M., Schweizer, S., Ochsenfeld, C. & Wennemers, H. (2009). Effects of terminal functional groups on the stability of the polyproline II structure: a combined experimental and theoretical study. *J. Am. Chem. Soc.* **131**, 15474–15482.
 40. He, L., Navarro, A. E., Shi, Z. & Kallenbach, N. R. (2011). End effects influence short model peptide conformation. *J. Am. Chem. Soc.* **134**, 1571–1576.
 41. Shi, Z., Chen, K., Liu, Z. & Kallenbach, N. R. (2006). Conformation of the backbone in unfolded proteins. *Chem. Rev.* **106**, 1877–1897.
 42. Chen, K., Liu, Z. & Kallenbach, N. R. (2004). The polyproline II conformation in short alanine peptides is noncooperative. *Proc. Natl Acad. Sci. USA*, **101**, 15352–15357.
 43. Schweitzer-Stenner, R., Eker, F., Griebenow, K., Cao, X. & Nafie, L. A. (2004). The conformation of tetraalanine in water determined by polarized raman, FT-IR, and VCD spectroscopy. *J. Am. Chem. Soc.* **126**, 2768–2776.
 44. Makarov, A. A., Adzhubei, I. A., Protasevich, I. I., Lobachov, V. M. & Fasman, G. D. (1994). Melting of the left-handed helical conformation of charged poly-L-lysine. *Biopolymers*, **34**, 1123–1124.
 45. Makowska, J., Rodziewicz-Motowidło, S., Bagińska, K., Vila, J. A., Liwo, A., Chmurzyński, L. & Scheraga, H. A. (2006). Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl Acad. Sci. USA*, **103**, 1744–1749.
 46. Makowska, J., Rodziewicz-Motowidło, S., Bagińska, K., Makowski, M., Vila, J. A., Liwo, A. et al. (2007). Further evidence for the absence of polyproline II stretch in the XAO peptide. *Biophys. J.* **92**, 2904–2917.
 47. Zhu, F., Kapitan, J., Tranter, G. E., Pudney, P. D. A., Isaacs, N. W., Hecht, L. & Barron, L. D. (2008). Residual structure in disordered peptides and unfolded proteins from multivariate analysis and ab initio simulation of Raman optical activity data. *Proteins*, **70**, 823–833.
 48. Schweitzer-Stenner, R. & Measey, T. J. (2007). The alanine-rich XAO peptide adopts a heterogeneous population, including turn-like and polyproline II conformations. *Proc. Natl Acad. Sci. USA*, **104**, 6649–6654.
 49. Zagrovic, B., Lipfert, J., Sorin, E. J., Millett, I. S., Van Gunsteren, W. F., Doniach, S. & Pande, V. S. (2005). Unusual compactness of a polyproline type II structure. *Proc. Natl Acad. Sci. USA*, **102**, 11698–11703.

50. Graf, J., Nguyen, P. H., Stock, G. & Schwalbe, H. (2007). Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **129**, 1179–1189.
51. Schweitzer-Stenner, R. (2009). Distribution of conformations sampled by the central amino acid residue in tripeptides inferred from amide I band profiles and NMR scalar coupling constants. *J. Phys. Chem. B*, **113**, 2922–2932.
52. Blanch, E. W., Morozova-Roche, L. A., Cochran, D. A., Doig, A. J., Hecht, L. & Barron, L. D. (2000). Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme1. *J. Mol. Biol.* **301**, 553–563.
53. Asher, S. A., Mikhonin, A. V. & Bykov, S. (2004). UV raman demonstrates that α -helical polyalanine peptides melt to polyproline II conformations. *J. Am. Chem. Soc.* **126**, 8433–8440.
54. Oh, K.-I., Lee, K.-K., Park, E.-K., Jung, Y., Hwang, G.-S. & Cho, M. (2012). A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. *Proteins: Struct. Funct. Bioinform.* **80**, 977–990.
55. Shi, Z., Chen, K., Liu, Z., Ng, A., Bracken, W. C. & Kallenbach, N. R. (2005). Polyproline II propensities from GGXGG peptides reveal an anticorrelation with beta-sheet scales. *Proc. Natl Acad. Sci. USA*, **102**, 17964–17968.
56. Hagarman, A., Measey, T. J., Mathieu, D., Schwalbe, H. & Schweitzer-Stenner, R. (2010). Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I' band profiles and NMR scalar coupling constants. *J. Am. Chem. Soc.* **132**, 540–551.
57. Rucker, A. L. & Creamer, T. P. (2002). Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* **11**, 980–985.
58. Eker, F., Griebenow, K., Cao, X., Nafie, L. A. & Schweitzer-Stenner, R. (2003). Tripeptides with ionizable side chains adopt a perturbed polyproline II structure in water. *Biochemistry*, **43**, 613–621.
59. Schweitzer-Stenner, R., Measey, T., Kakalis, L., Jordan, F., Pizzanelli, S., Forte, C. & Griebenow, K. (2007). Conformations of alanine-based peptides in water probed by FTIR, raman, vibrational circular dichroism, electronic circular dichroism, and NMR spectroscopy. *Biochemistry*, **46**, 1587–1596.
60. Adzhubei, A. A., Eisenmenger, F., Tumanyan, V. G., Zinke, M., Brodzinski, S. & Espipova, N. G. (1987). Approaching a complete classification of protein secondary structure. *J. Biomol. Struct. Dyn.* **5**, 689–704.
61. Sreerama, N. & Woody, R. W. (1999). Molecular dynamics simulations of polypeptide conformations in water: a comparison of alpha, beta, and poly(pro)II conformations. *Proteins: Struct. Funct. Genet.* **36**, 400–406.
62. Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E. & Creamer, T. P. (2003). Host-guest scale of left-handed polyproline II helix formation. *Proteins*, **53**, 68–75.
63. Poon, C.-D., Samulski, E. T., Weise, C. F. & Weisshaar, J. C. (2000). Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution? *J. Am. Chem. Soc.* **122**, 5642–5643.
64. Han, W.-G., Jalkanen, K. J., Elstner, M. & Suhai, S. (1998). Theoretical study of aqueous *N*-acetyl-L-alanine *N'*-methylamide: structures and Raman, VCD, and ROA spectra. *J. Phys. Chem. B*, **102**, 2587–2602.
65. Smith, P. E. (1999). The alanine dipeptide free energy surface in solution. *J. Chem. Phys.* **111**, 5568–5579.
66. Mukhopadhyay, P., Zuber, G. & Beratan, D. N. (2008). Characterizing aqueous solution conformations of a peptide backbone using raman optical activity computations. *Biophys. J.* **95**, 5574–5586.
67. Law, P. B. & Daggett, V. (2010). The relationship between water bridges and the polyproline II conformation: a large-scale analysis of molecular dynamics simulations and crystal structures. *Protein Eng. Des. Sel.* **23**, 27–33.
68. Mezei, M., Fleming, P. J., Srinivasan, R. & Rose, G. D. (2004). Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins*, **55**, 502–507.
69. Pappu, R. V. & Rose, G. D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **11**, 2437–2455.
70. Fleming, P. J., Fitzkee, N. C., Mezei, M., Srinivasan, R. & Rose, G. D. (2005). A novel method reveals that solvent water favors polyproline II over β -strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). *Protein Sci.* **14**, 111–118.
71. Drozdov, A. N., Grossfield, A. & Pappu, R. V. (2004). Role of solvent in determining conformational preferences of alanine dipeptide in water. *J. Am. Chem. Soc.* **126**, 2574–2581.
72. Horng, J.-C. & Raines, R. T. (2006). Stereoelectronic effects on polyproline conformation. *Protein Sci.* **15**, 74–83.
73. Creamer, T. P. & Campbell, M. N. (2002). Determinants of the polyproline II helix from modeling studies. *Adv. Protein Chem.* **62**, 263–282.
74. Whittington, S. J., Chellgren, B. W., Hermann, V. M. & Creamer, T. P. (2005). Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry*, **44**, 6269–6275.
75. Hu, K.-N., Havlin, R. H., Yau, W.-M. & Tycko, R. (2009). Quantitative determination of site-specific conformational distributions in an unfolded protein by solid state nuclear magnetic resonance. *J. Mol. Biol.* **392**, 1055–1073.
76. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N. et al. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12491–12496.
77. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.
78. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. (2005). Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl Acad. Sci. USA*, **102**, 13099–13104.

79. Cortajarena, A. L., Lois, G., Sherman, E., O'Hern, C. S., Regan, L. & Haran, G. (2008). Non-random-coil behavior as a consequence of extensive PPII structure in the denatured state. *J. Mol. Biol.* **382**, 203–212.
80. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18**, 756–764.
81. Uversky, V. N. & Dunker, A. K. (2010). Understanding protein non-folding. *Biochim. Biophys. Acta*, **1804**, 1231–1264.
82. Bochicchio, B. & Tamburro, A. M. (2002). Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. *Chirality*, **14**, 782–792.
83. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
84. Syme, C. D., Blanch, E. W., Holt, C., Jakes, R., Goedert, M., Hecht, L. & Barron, L. D. (2002). A Raman optical activity study of rheomorphism in caseins, synucleins and tau. *Eur. J. Biochem.* **269**, 148–156.
85. Blanch, E. W., Kasarda, D. D., Hecht, L., Nielsen, K. & Barron, L. D. (2003). New insight into the solution structures of wheat gluten proteins from Raman optical activity. *Biochemistry*, **42**, 5665–5673.
86. Delak, K., Harcup, C., Lakshminarayanan, R., Sun, Z., Fan, Y., Moradian-Oldak, J. & Evans, J. S. (2009). The tooth enamel protein, porcine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. *Biochemistry*, **48**, 2272–2281.
87. Haaning, S., Radutoiu, S., Hoffmann, S. V., Dittmer, J., Giehm, L., Otzen, D. E. & Stougaard, J. (2008). An unusual intrinsically disordered protein from the model legume *Lotus japonicus* stabilizes proteins in vitro. *J. Biol. Chem.* **283**, 31142–31152.
88. Sánchez-Puig, N., Veprinsev, D. B. & Fersht, A. R. (2005). Human full-length Securin is a natively unfolded protein. *Protein Sci.* **14**, 1410–1418.
89. Lefèvre, T., Leclerc, J., Rioux-Dubé, J.-F., Buffeteau, T., Paquin, M.-C., Rousseau, M.-E. *et al.* (2007). Conformation of spider silk proteins in situ in the intact major ampullate gland and in solution. *Biomacromolecules*, **8**, 2342–2344.
90. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Struct. Funct. Bioinform.* **42**, 38–48.
91. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533.
92. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. (2003). Flavors of protein disorder. *Proteins: Struct. Funct. Bioinform.* **52**, 573–584.
93. Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N. & Dunker, A. K. (2008). TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **15**, 956–963.
94. Elangovan, S., Margolis, H. C., Oppenheim, F. G. & Beniash, E. (2007). Conformational changes in salivary proline-rich protein 1 upon adsorption to calcium phosphate crystals. *Langmuir*, **23**, 11200–11205.
95. Ishijima, J., Nagasaki, N., Maeshima, M. & Miyano, M. (2007). RVCA_B, a calcium-binding protein in radish vacuoles, is predominantly an unstructured protein with a polyproline type II helix. *J. Biochem.* **142**, 201–211.
96. Pascal, C., Paté, F., Cheynier, V. & Delsuc, M.-A. (2009). Study of the interactions between a proline-rich protein and a flavan-3-ol by NMR: residual structures in the natively unfolded protein provides anchorage points for the ligands. *Biopolymers*, **91**, 745–756.
97. Kentsis, A., Mezei, M., Gindin, T. & Osman, R. (2004). Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins*, **55**, 493–501.
98. Sreerama, N. & Woody, R. W. (2003). Structural composition of β I- and β II-proteins. *Protein Sci.* **12**, 384–388.
99. Soman, K. V. & Ramakrishnan, C. (1983). Occurrence of a single helix of the collagen type in globular proteins. *J. Mol. Biol.* **170**, 1045–1048.
100. Swindells, M. B., MacArthur, M. W. & Thornton, J. M. (1995). Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* **2**, 596–603.
101. Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. (1996). The concept of a random coil: residual structure in peptides and denatured proteins. *Fold. Des.* **1**, R95–R106.
102. Hovmöller, S., Zhou, T. & Ohlson, T. (2002). Conformations of amino acids in proteins. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 768–776.
103. Beck, D. A. C., Alonso, D. O. V., Inoyama, D. & Daggett, V. (2008). The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proc. Natl Acad. Sci. USA*, **105**, 12259–12264.
104. Worth, C. L. & Blundell, T. L. (2010). On the evolutionary conservation of hydrogen bonds made by buried polar amino acids: the hidden joists, braces and trusses of protein architecture. *BMC Evol. Biol.* **10**, 161.
105. Elam, W. A., Schrank, T. P. & Hilser, V. J. (2012). Experimental and computational studies of polyproline II propensity. In *Protein and Peptide Folding, Misfolding, and Non-Folding* (Schweitzer-Stenner, R., ed.), pp. 159–185, John Wiley & Sons, Inc.
106. Brown, A. M. & Zondlo, N. J. (2012). A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions. *Biochemistry*, **51**, 5041–5051.
107. Elam, W. A., Schrank, T. P., Campagnolo, A. J. & Hilser, V. J. (2013). Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites. *Protein Sci.* **22**, 405–417.
108. Muñoz, V. & Serrano, L. (1994). Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Proteins*, **20**, 301–311.
109. Helles, G. & Fonseca, R. (2009). Predicting dihedral angle probability distributions for protein coil residues from primary sequence using neural networks. *BMC Bioinformatics*, **10**, 338.

110. Siermala, M., Juhola, M. & Vihtinen, M. (2001). On preprocessing of protein sequences for neural network prediction of polyproline type II secondary structures. *Comput. Biol. Med.* **31**, 385–398.
111. Wang, M.-L., Yao, H. & Xu, W.-B. (2005). Prediction by support vector machines and analysis by Z-score of poly-L-proline type II conformation based on local sequence. *Comput. Biol. Chem.* **29**, 95–100.
112. Vlasov, P. K., Vlasova, A. V., Tumanyan, V. G. & Esipova, N. G. (2005). A tetrapeptide-based method for polyproline II-type secondary structure prediction. *Proteins*, **61**, 763–768.
113. Pirovano, W. & Heringa, J. (2010). Protein secondary structure prediction. *Methods Mol. Biol.* **609**, 327–348.
114. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Bio-polymers*, **22**, 2577–2637.
115. King, S. M. & Johnson, W. C. (1999). Assigning secondary structure from protein coordinate data. *Proteins: Struct. Funct. Bioinform.* **35**, 313–320.
116. Cubellis, M. V., Cailliez, F. & Lovell, S. C. (2005). Secondary structure assignment that accurately reflects physical and evolutionary characteristics. *BMC Bioinformatics*, **6**, S8.
117. Mansiaux, Y., Joseph, A. P., Gelly, J.-C. & De Brevern, A. G. (2011). Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PLoS One*, **6**, e18401.
118. Jozic, D., Cárdenes, N., Deribe, Y. L., Moncalián, G., Hoeller, D., Groemping, Y. et al. (2005). Cbl promotes clustering of endocytic adaptor proteins. *Nat. Struct. Mol. Biol.* **12**, 972–979.
119. Bottoms, C. A., Schuemann, J. P., Agah, S., Henzl, M. T. & Tanner, J. J. (2004). Crystal structure of rat α-parvalbumin at 1.05 Å resolution. *Protein Sci.* **13**, 1724–1734.
120. Detlefsen, D. J., Thanabal, V., Pecoraro, V. L. & Wagner, G. (1991). Solution structure of Fe(II) cytochrome c551 from *Pseudomonas aeruginosa* as determined by two-dimensional 1H NMR. *Biochemistry*, **30**, 9040–9046.
121. Tronrud, D. E., Schmid, M. F. & Matthews, B. W. (1986). Structure and X-ray amino acid sequence of a bacteriochlorophyll a protein from *Prosthecochloris aestuarii* refined at 1.9 Å resolution. *J. Mol. Biol.* **188**, 443–454.
122. Feng, J. A., Johnson, R. C. & Dickerson, R. E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*, **263**, 348–355.
123. Beamer, L. J. & Pabo, C. O. (1992). Refined 1.8 Å crystal structure of the lambda repressor–operator complex. *J. Mol. Biol.* **227**, 177–196.
124. Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C. & Pabo, C. O. (1991). The DNA binding arm of lambda repressor: critical contacts from a flexible region. *Science*, **254**, 267–270.
125. Matsushima, N., Tanaka, T., Enkhbayar, P., Mikami, T., Taga, M., Yamada, K. & Kuroki, Y. (2007). Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics*, **8**, 124.
126. Park, H., Huxley-Jones, J., Boot-Handford, R., Bishop, P., Attwood, T. & Bella, J. (2008). LRRCE: a leucine-rich repeat cysteine capping motif unique to the chordate lineage. *BMC Genomics*, **9**, 599.
127. Lakshminarayanan, R., Yoon, I., Hegde, B. G., Fan, D., Du, C. & Moradian-Oldak, J. (2009). Analysis of secondary structure and self-assembly of amelogenin by variable temperature circular dichroism and isothermal titration calorimetry. *Proteins*, **76**, 560–569.
128. Bochicchio, B. & Pepe, A. (2011). Role of polyproline II conformation in human tropoelastin structure. *Chirality*, **23**, 694–702.
129. Blanch, E. W., Gill, A. C., Rhie, A. G. O., Hope, J., Hecht, L., Nielsen, K. & Barron, L. D. (2004). Raman optical activity demonstrates poly(L-proline) II helix in the N-terminal region of the ovine prion protein: implications for function and malfunction. *J. Mol. Biol.* **343**, 467–476.
130. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W. & Schreiber, S. L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*, **76**, 933–945.
131. Musacchio, A., Saraste, M. & Wilmanns, M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct. Mol. Biol.* **1**, 546–551.
132. Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N. & Lim, W. A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science*, **282**, 2088–2092.
133. Lim, W. A., Richards, F. M. & Fox, R. O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*, **372**, 375–379.
134. Freund, C., Dötsch, V., Nishizawa, K., Reinherz, E. L. & Wagner, G. (1999). The GYF domain is a novel structural fold that is involved in lymphoid signalling through proline-rich sequences. *Nat. Struct. Biol.* **6**, 656–660.
135. Carl, U. D., Pollmann, M., Orr, E., Gertler, F. B., Chakraborty, T. & Wehland, J. (1999). Aromatic and basic residues within the EVH1 domain of VASP specify its interaction with proline-rich ligands. *Curr. Biol.* **9**, 715–718.
136. Pornillos, O., Alam, S. L., Davis, D. R. & Sundquist, W. I. (2002). Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat. Struct. Biol.* **9**, 812–817.
137. Mahoney, N. M., Janmey, P. A. & Almo, S. C. (1997). Structure of the profilin-poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. *Nat. Struct. Mol. Biol.* **4**, 953–960.
138. McPherson, P. S., Czernik, A. J., Chilcote, T. J., Onofri, F., Benfenati, F., Greengard, P. et al. (1994). Interaction of Grb2 via its Src homology 3 domains with synaptic proteins including synapsin I. *Proc. Natl. Acad. Sci. USA*, **91**, 6486–6490.
139. Onofri, F., Giovedi, S., Kao, H.-T., Valtorta, F., Borbone, L. B., Camilli, P. D. et al. (2000). Specificity of the binding of synapsin I to Src homology 3 domains. *J. Biol. Chem.* **275**, 29857–29867.

140. Gout, I., Dhand, R., Hiles, I. D., Fry, M. J., Panayotou, G., Das, P. *et al.* (1993). The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell*, **75**, 25–36.
141. Jardetzky, T. S., Brown, J. H., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc. Natl Acad. Sci. USA*, **93**, 734–738.
142. Song, J., McGivern, J. V., Nichols, K. W., Markley, J. L. & Sheets, M. D. (2008). Structural basis for RNA recognition by a type II poly(A)-binding protein. *Proc. Natl Acad. Sci. USA*, **105**, 15317–15322.
143. Kielkopf, C. L., Rodionova, N. A., Green, M. R. & Burley, S. K. (2001). A novel peptide recognition mode revealed by the x-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*, **106**, 595–605.
144. Vlasov, P., Budzko, A., Rubin, M., Tumanyan, V., Makarov, A. & Esipova, N. (2008). Left-handed helix of polyproline II type in linker regions of DNA-binding proteins. *Biophysics*, **53**, 663–664.
145. Fitzkee, N. C. & Rose, G. D. (2004). Steric restrictions in protein folding: an α -helix cannot be followed by a contiguous β -strand. *Protein Sci.* **13**, 633–639.
146. Lakshminarayanan, R., Fan, D., Du, C. & Moradian-Oldak, J. (2007). The role of secondary structure in the entropically driven amelogenin self-assembly. *Biophys. J.* **93**, 3664–3674.
147. Ma, K., Kan, L. & Wang, K. (2001). Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. *Biochemistry*, **40**, 3427–3438.
148. Tamburro, A. M., Bochicchio, B. & Pepe, A. (2005). The dissection of human tropoelastin: from the molecular structure to the self-assembly to the elasticity mechanism. *Pathol. Biol.* **53**, 383–389.
149. Ravi Chandra, B., Gowthaman, R., Raj Akhouri, R., Gupta, D. & Sharma, A. (2004). Distribution of proline-rich (PxxP) motifs in distinct proteomes: functional and therapeutic implications for malaria and tuberculosis. *Protein Eng. Des. Sel.* **17**, 175–182.
150. Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K. & Uversky, V. N. (2006). Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **362**, 1043–1059.
151. Ahmad, M., Gu, W. & Helms, V. (2008). Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem. Int. Ed.* **47**, 7626–7630.
152. Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. (2003). The structure and function of proline recognition domains. *Sci. STKE*, **2003**, RE8.
153. Freund, C., Schmalz, H.-G., Sticht, J. & Kühne, R. (2008). Proline-rich sequence recognition domains (PRD): ligands, function and inhibition. *Handb. Exp. Pharmacol.* 407–429 http://dx.doi.org/10.1007/978-3-540-72843-6_17.
154. Ren, S., Uversky, V. N., Chen, Z., Dunker, A. K. & Obradovic, Z. (2008). Short linear motifs recognized by SH2, SH3 and Ser/Thr kinase domains are conserved in disordered protein regions. *BMC Genomics*, **9**, S26.
155. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*, **266**, 1241–1247.
156. Madden, D. R. (1995). The three-dimensional structure of peptide–MHC complexes. *Annu. Rev. Immunol.* **13**, 587–622.
157. Whitten, S. T., Yang, H.-W., Fox, R. O. & Hilser, V. J. (2008). Exploring the impact of polyproline II (PPII) conformational bias on the binding of peptides to the SEM-5 SH3 domain. *Protein Sci.* **17**, 1200–1211.
158. Ball, L. J., Kühne, R., Schneider-Mergener, J. & Oschkinat, H. (2005). Recognition of proline-rich motifs by protein–protein interaction domains. *Angew. Chem. Int. Ed Engl.* **44**, 2852–2869.
159. Parsons, S. J. & Parsons, J. T. (2004). Src family kinases, key regulators of signal transduction. *Oncogene*, **23**, 7906–7909.
160. Mayer, B. J. (2001). SH3 domains: complexity in moderation. *J. Cell Sci.* **114**, 1253–1263.
161. Li, S. S.-C. (2005). Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem. J.* **390**, 641.
162. Kaneko, T., Li, L. & Li, S. S.-C. (2008). The SH3 domain—a family of versatile peptide- and protein-recognition module. *Front. Biosci.* **13**, 4938–4952.
163. Peterson, F. C. & Volkman, B. F. (2009). Diversity of polyproline recognition by EVH1 domains. *Front. Biosci.* **14**, 833–846.
164. Freund, C., Kühne, R., Yang, H., Park, S., Reinherz, E. L. & Wagner, G. (2002). Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J.* **21**, 5985–5995.
165. Kofler, M. M. & Freund, C. (2006). The GYF domain. *FEBS J.* **273**, 245–256.
166. Saito, K., Kigawa, T., Koshiba, S., Sato, K., Matsuo, Y., Sakamoto, A. *et al.* (2004). The CAP-Gly domain of CYLD associates with the proline-rich sequence in NEMO/IKKgamma. *Structure*, **12**, 1719–1728.
167. Jia, C. Y. H., Nie, J., Wu, C., Li, C. & Li, S. S.-C. (2005). Novel Src homology 3 domain-binding motifs identified from proteomic screen of a Pro-rich region. *Mol. Cell Proteomics*, **4**, 1155–1166.
168. Musacchio, A. (2002). How SH3 domains recognize proline. *Adv. Protein Chem.* **61**, 211–268.
169. Sicheri, F., Moarefi, I. & Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature*, **385**, 602–609.
170. Engen, J., Wales, T., Hochrein, J., Meyn, M., Banu Ozkan, S., Bahar, I. & Smithgall, T. (2008). Structure and dynamic regulation of Src-family kinases. *Cell. Mol. Life Sci.* **65**, 3058–3073.
171. Alvarado, J. J., Betts, L., Moroco, J. A., Smithgall, T. E. & Yeh, J. I. (2010). Crystal structure of the Src family kinase Hck SH3–SH2 linker regulatory region supports an SH3-dominant activation mechanism. *J. Biol. Chem.* **285**, 35455–35461.
172. Porter, M., Schindler, T., Kuriyan, J. & Miller, W. T. (2000). Reciprocal regulation of hck activity by phosphorylation of Tyr527 and Tyr416. *J. Biol. Chem.* **275**, 2721–2726.

173. Boggon, T. J. & Eck, M. J. (2004). Structure and regulation of Src family kinases. *Oncogene*, **23**, 7918–7927.
174. Williams, N. K., Lucet, I. S., Klinken, S. P., Ingle, E. & Rossjohn, J. (2009). Crystal structures of the lyn protein tyrosine kinase domain in its Apo- and inhibitor-bound state. *J. Biol. Chem.* **284**, 284–291.
175. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M. & Tanokura, M. (2004). Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J. Biol. Chem.* **279**, 31833–31841.
176. Macias, M. J., Wiesner, S. & Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* **513**, 30–37.
177. Lassing, I. & Lindberg, U. (1985). Specific interaction between phosphatidylinositol 4,5-bisphosphate and profilactin. *Nature*, **314**, 472–474.
178. Machesky, L. M. & Poland, T. D. (1993). Profilin as a potential mediator of membrane-cytoskeleton communication. *Trends Cell Biol.* **3**, 381–385.
179. Lu, J. & Pollard, T. D. (2001). Profilin binding to poly-L-proline and actin monomers along with ability to catalyze actin nucleotide exchange is required for viability of fission yeast. *Mol. Biol. Cell*, **12**, 1161–1175.
180. Jockusch, B. M., Murk, K. & Rothkegel, M. (2007). The profile of profilins. *Rev. Physiol. Biochem. Pharmacol.* **159**, 131–149.
181. Birbach, A. (2008). Profilin, a multi-modal regulator of neuronal plasticity. *BioEssays*, **30**, 994–1002.
182. Vik, Å., Aas, F. E., Anonsen, J. H., Bilsborough, S., Schneider, A., Egge-Jacobsen, W. & Koomey, M. (2009). Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci.* **106**, 4447–4452.
183. Julenius, K., Mølgaard, A., Gupta, R. & Brunak, S. (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*, **15**, 153–164.
184. Meinhart, A. & Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, **430**, 223–226.
185. De Silva, U., Choudhury, S., Bailey, S. L., Harvey, S., Perrino, F. W. & Hollis, T. (2007). The crystal structure of TREX1 explains the 3' nucleotide specificity and reveals a polyproline II helix for protein partnering. *J. Biol. Chem.* **282**, 10537–10543.
186. Polverini, E., Rangaraj, G., Libich, D. S., Boggs, J. M. & Harauz, G. (2008). Binding of the proline-rich segment of myelin basic protein to SH3 domains: spectroscopic, microarray, and modeling studies of ligand conformation and effects of posttranslational modifications. *Biochemistry*, **47**, 267–282.
187. Harauz, G. & Libich, D. S. (2009). The classic basic protein of myelin-conserved structural motifs and the dynamic molecular barcode involved in membrane adhesion and protein–protein interactions. *Curr. Protein Pept. Sci.* **10**, 196–215.
188. Harauz, G., Ladizhansky, V. & Boggs, J. M. (2009). Structural polymorphism and multifunctionality of myelin basic protein. *Biochemistry*, **48**, 8094–8104.
189. Bouton, A. H., Riggins, R. B. & Bruce-Staskal, P. J. (2001). Functions of the adapter protein Cas: signal convergence and the determination of cellular responses. *Oncogene*, **20**, 6448–6458.
190. Pellicena, P. & Miller, W. T. (2001). Processive phosphorylation of p130Cas by Src depends on SH3–polyproline interactions. *J. Biol. Chem.* **276**, 28190–28196.
191. Yadav, S. S. & Miller, W. T. (2007). Cooperative activation of Src family kinases by SH3 and SH2 ligands. *Cancer Lett.* **257**, 116–123.
192. Berisio, R. & Vitagliano, L. (2012). Polyproline and triple helix motifs in host–pathogen recognition. *Curr. Protein Pept. Sci.* **13**, 855–865.
193. Arold, S., Franken, P., Strub, M. P., Hoh, F., Benichou, S., Benarous, R. & Dumas, C. (1997). The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. *Structure*, **5**, 1361–1372.
194. Hiipakka, M. & Saksela, K. (2002). Capacity of simian immunodeficiency virus strain mac Nef for high-affinity Src homology 3 (SH3) binding revealed by ligand-tailored SH3 domains. *J. Gen. Virol.* **83**, 3147–3152.
195. Saksela, K. (2011). Interactions of the HIV/SIV pathogenicity factor Nef with SH3 domain-containing host cell proteins. *Curr. HIV Res.* **9**, 531–542.
196. Hanna, Z., Weng, X., Kay, D. G., Poudrier, J., Lowell, C. & Jolicoeur, P. (2001). The pathogenicity of Human Immunodeficiency Virus (HIV) type 1 Nef in CD4C/HIV transgenic mice is abolished by mutation of its SH3-binding domain, and disease development is delayed in the absence of Hck. *J. Virol.* **75**, 9378–9392.
197. Fedorov, A. A., Ball, T., Mahoney, N. M., Valenta, R. & Almo, S. C. (1997). The molecular basis for allergen cross-reactivity: crystal structure and IgE epitope mapping of birch pollen profilin. *Structure*, **5**, 33–45.
198. Niebuhr, K., Ebel, F., Frank, R., Reinhard, M., Domann, E., Carl, U. D. et al. (1997). A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J.* **16**, 5433–5444.
199. Bommarius, B., Maxwell, D., Swimm, A., Leung, S., Corbett, A., Bornmann, W. & Kalman, D. (2007). Enteropathogenic *Escherichia coli* Tir is an SH2/3 ligand that recruits and activates tyrosine kinases required for pedestal formation. *Mol. Microbiol.* **63**, 1748–1768.
200. Murthy, V. L. & Stern, L. J. (1997). The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure*, **5**, 1385–1396.
201. Zavala-Ruiz, Z., Sundberg, E. J., Stone, J. D., DeOliveira, D. B., Chan, I. C., Svendsen, J. et al. (2003). Exploration of the P6/P7 region of the peptide-binding site of the human class II major histocompatibility complex protein HLA-DR1. *J. Biol. Chem.* **278**, 44904–44912.
202. Hoh, F., Cavé, A., Strub, M. P., Banères, J. L. & Padilla, A. (2009). Removing the invariant salt bridge of parvalbumin increases flexibility in the AB-loop

- structure. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **65**, 733–743.
203. Fraenkel, E. & Pabo, C. O. (1998). Comparison of X-ray and NMR structures for the Antennapedia homeodomain–DNA complex. *Nat. Struct. Mol. Biol.* **5**, 692–697.
204. Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. (2008). The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**, 2307–2333.
205. Prusiner, S. B. (1998). Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383.
206. Aguzzi, A., Baumann, F. & Bremer, J. (2008). The prion's elusive reason for being. *Annu. Rev. Neurosci.* **31**, 439–477.
207. Gill, A. C., Ritchie, M. A., Hunt, L. G., Steane, S. E., Davies, K. G., Bocking, S. P. *et al.* (2000). Post-translational hydroxylation at the N-terminus of the prion protein reveals presence of PPII structure in vivo. *EMBO J.* **19**, 5324–5331.
208. Taubner, L. M., Bienkiewicz, E. A., Copié, V. & Caughey, B. (2010). Structure of the flexible amino-terminal domain of prion protein bound to a sulfated glycan. *J. Mol. Biol.* **395**, 475–490.
209. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. (2004). Evidence for assembly of prions with left-handed β -helices into trimers. *Proc. Natl Acad. Sci. USA* **101**, 8342–8347.
210. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R. & Meier, B. H. (2008). Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science*, **319**, 1523–1526.
211. Van Melckebeke, H., Wasmer, C., Lange, A., Ab, E., Loquet, A., Böckmann, A. & Meier, B. H. (2010). Atomic-resolution three-dimensional structure of HET-s(218–289) amyloid fibrils by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 13765–13775.
212. Ziegler, J., Viehrig, C., Geimer, S., Rösch, P. & Schwarzinger, S. (2006). Putative aggregation initiation sites in prion protein. *FEBS Lett.* **580**, 2033–2040.
213. Kozin, S. A., Bertho, G., Mazur, A. K., Rabesona, H., Girault, J.-P., Haertlé, T. *et al.* (2001). Sheep prion protein synthetic peptide spanning helix 1 and β -strand 2 (residues 142–166) shows β -hairpin structure in solution. *J. Biol. Chem.* **276**, 46364–46370.
214. Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. & Wüthrich, K. (2005). Prion protein NMR structures of elk and of mouse/elk hybrids. *Proc. Natl Acad. Sci. USA* **102**, 646–650.
215. Hafner-Bratkovič, I., Bester, R., Pristovšek, P., Gaedtke, L., Veranič, P., Gašperšič, J. *et al.* (2011). Globular domain of the prion protein needs to be unlocked by domain swapping to support prion protein conversion. *J. Biol. Chem.* **286**, 12149–12156.
216. Paramithiotis, E., Pinard, M., Lawton, T., LaBoissière, S., Leathers, V. L., Zou, W.-Q. *et al.* (2003). A prion protein epitope selective for the pathologically misfolded conformation. *Nat. Med.* **9**, 893–899.
217. Novotny, M. & Kleywegt, G. J. (2005). A survey of left-handed helices in protein structures. *J. Mol. Biol.* **347**, 231–241.
218. Fodje, M. N. & Al-Karadaghi, S. (2002). Occurrence, conformational features and amino acid propensities for the $\{\pi\}$ -helix. *Protein Eng.* **15**, 353–358.
219. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A. *et al.* (2007). DisProt: the database of disordered proteins. *Nucleic Acids Res.* **35**, D786–D793.
220. Domenico, T. D., Walsh, I., Martin, A. J. M. & Tosatto, S. C. E. (2012). MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics*, **28**, 2080–2081.