Architectonic Principles of Polyproline II Bundle Protein Domains

Cristian Segura Rodríguez[1] & Douglas V. Laurents[1,*]

[1]Instituto de Química Física "Blas Cabrera" CSIC
Serrano 119 Madrid, Spain
*To whom correspondence may be addressed: dlaurents@iqfr.csic.es

Abstract

Glycine rich polyproline II helix assemblies are an emerging class of natural domains found in several proteins with different functions and diverse origins. The distinct properties of these domains relative to those composed of α-helices and β-sheets could make glycine-rich polyproline II helix assemblies a useful building block for protein design. However, the architectonic bases of these structures are not well known. Here, we compare and analyze their structures to uncover common features. These protein domains are found to be highly tolerant of distinct flanking sequences. This speaks to the robustness of this fold and strongly suggests that glycine rich polyproline II assemblies could be grafted with other protein domains to engineer new structures and functions. These domains are also well packed with few or no cavities. Moreover, a significant trend towards anti-parallel helix configuration is observed in all these domains and could provide stabilizing interactions among macrodipoles. Finally, extensive networks of Cα-H···O=C hydrogen bonds are detected in these domains. Despite their diverse evolutionary origins and activities, glycine-rich polyproline II helix assemblies share architectonic features which could help guide the design of novel proteins.

Introduction

Protein structure comparison is a useful approach to uncover principles of protein architecture. By the mid-1970's, a few dozen protein structures were known, which permitted Levitt and Chothia [1] to compare them via simplified topology/packing diagrams. Based on their results, they proposed an early structural classification scheme based on the content of protein secondary structure. They also found that when proteins are composed of just β secondary structure, the β-strands tend to be antiparallel and that in proteins of mixed α-helical & β-strand secondary structures, the α-helices tend to be positioned on the outside and β-strands tend to be parallel and on the inside. Moveover, certain preferred combinations of elements of secondary structure, such as two α-helices arranged in an antiparallel manner and two β-strands in parallel connected by an α-helix were discovered. Additionally, they uncovered that elements of protein secondary structure which are adjacent in the sequence are frequently in contact in the tertiary structure. Thanks to modern experiments in the protein design field, we now know that proximity of sequence and structures plays an important role in helping proteins fold quickly [2].

Further comparative studies identified the tendency of protein secondary structures to arrange into autonomously folding regions called domains in larger proteins and a general lack of knots[3]. Despite the relative success of protein comparative studies and new machine learning applications such as AlphaFold[4] and RoseTTAFold[5], they have largely focused on α-helices and β-sheet based protein structure and mostly ignore polyproline II helices [6].

Polyproline II helices are left-handed spirals with three residues per turn (**Figure 1A**). They are rather extended, as they advance 3.1 Å per residue compared to 1.5 Å per residue for α-helices and 3.0 – 3.8 Å per residues for β-strands. The polyproline II (PPII) helix gets its name because this is the conformation adopted by peptides composed purely of proline in aqueous solution [7]. However, when every third proline residue is substituted by a glycine residue, three PPII helices can associate and H-bond, giving rise the well-known and abundant collagen triple helix structure (**Figure 1B**). Furthermore, peptides containing the sequence $(\text{Pro-Gly-Gly})_N$ adopts polyproline II helices which assemble into double-layered sheets with proline residues on the outside and glycine residues and an extensive hydrogen bonding network on the inside [8] (**Figure 1C**). A similar structure

is adopted by $(Ala\text{-}Gly\text{-}Gly)_N$[9]. Remarkably, the substitution of all three prolines for glycine gives rise to the complete 3D network of PPII helices in which each PPII helix is surrounded by and forms hydrogen bonds to six others [10] (**Figure 1D**). A closely related structure is adopted by $(Ala\text{-}Gly\text{-}Gly\text{-}Gly)_N$ [9]. Whether these chains have a preference for a parallel or anti-parallel orientation remains an open question. Except for collagen and collagen-like triple helices with one third glycine, throughout the 1970's, 1980's and 1990's the sequences and conformations described above with two thirds and even 100% glycine were not observed in natural proteins and seemed to be academic curiosities found only in synthetic peptides.

This view changed in the last two decades as various natural protein domains and proteins have been found to be composed of several glycine-rich PPII helices arranged in a bundle. They include the antifreeze proteins from primitive arthropods [11,12] a universally conserved translation factor [13, 14, 15], enzymes capable of forming new carbon – carbon bonds to incorporate CO_2 into organic compounds [16,17], the key tip domain in bacteriophage tail fibrils [18], and a kinase linked to human leukemia [19]. Their names and some characteristics are listed in **Table 1**.

In particular, the antifreeze proteins have proven valuable as systems to investigate the bases of the conformational stability of glycine-rich PPII helical bundles and have revealed disulfide bonds, nonconventional Cα-H···O=C hydrogen bonds, and a lower than anticipated conformational entropy in the denatured state as factors favoring the folded structure [20,21,22]. Nevertheless, some important questions remain regarding the roles of : 1) flanking sequences, which are known to be important for amyloid structures [23], 2) packing density, which is expected to impact the stabilizing contribution of van der Waals interactions, 3) the parallel versus antiparallel orientation of the helices[10], which we hypothesize could be stabilizing or destabilizing due to macrodipole interactions and 4) unconventional Cα-H···O=C hydrogen bonds.

Thanks to the diverse evolutionary origin, function and complexity of the PPII helical protein domains reported recently, a comparative study of their structural features is now possible. Following the approach pioneered by Levitt & Chothia [1], and by characterizing helix orientation and the sequence composition of flanking segments, we aim to address the questions posed above to uncover architectural principles for PPII helical bundle domains. This new knowledge may help guide the design of novel proteins based on PPII helical assemblies.

Methods

X-ray crystal structures of PPII helical bundle proteins and domains were examined and represented as 2D topological diagrams following the approach of Levitt & Chothia (1976) [1]. We tabulated the residues in the turns connecting the PPII helices to uncover characteristics of flanking residues. For longer connecting sequences, the five residues immediately preceding and the N-terminus and the five residues immediately following the C-terminus of each PPII helix were tabulated. The proteins' packing densities at buried positions in PPII helices were evaluated with the program Voronoia[24]; for comparison, the same analysis was also applied to α-helices and β-strands in some representative proteins previously studied by Liang and Dill[25].

Glycine residues have two Hα and in NMR spectra the chemical shifts of these two hydrogens become distinct when a hydrogen bond is formed. In *Aa*AFP, there are 14 Gly residues whose two Hα show chemical shifts differing by 0.50 ppm or more. This number agrees well with the 15 Cα-H participating in H-bonds identified by Gates *et al.* (2017)[20]. These Gly residues have one Hα positioned to accept a hydrogen bond from a nearby carbonyl group. Considering that for these H-bonds, the Cα-H··O=C distance is 3.54 +/- 0.2 (1σ) Å and the Cα-H···O=C bond angle is 157 +/- 22º (1σ), as criteria to detect putative Cα-H···O=C H-bonds in other glycine-rich PPII helical bundle proteins, we used a distance limit of < 3.75 Å (3.54 Å + 1σ) and a bond angle > 135º (157º – 1σ). The search was carried out using the program PYMOL and each detected putative Cα-H··O=C H-bond was subsequently checked and curated manually.

Results and Discussion

Topological diagrams for all reported polyproline II helical bundle proteins are shown in **Figure 2-7** in order of increasing complexity alongside their sequences and 3D structures. The snow flea antifreeze proteins have the simplest structures, which consist of two layers of PPII helices connected by short loops (**Fig. 2 & 3**). The two disulfide bonds present in *Hh*APF contribute decisively for the protein's conformational stability. The PPII helices alternative having their N- or C-termini up or down and this antiparallel arrangement would promote favorable electrostatic interactions between the helix macrodipoles. The N- and C-terminal PPII helices have only two other PPII helices as neighbors and are relatively poor in Gly residues.

Like *Hh*AFP, *Bs*Obg is also composed of six PPII helices (**Fig. 4**) arrange in two layers. However, these PPII helices are connected by segments of varying length and complexity; some even include α-helices and β-strands. The PPII helical domain of *Bs*Obg is not connected by disulfide bonds and are also arranged to promote favorable PPII macrodipole interactions. Like the snow flea AFPs, the central PPII helices, which have four bordering PPII helices, tend to contain more Gly residues.

Acetophenone carboxylase and acetone carboxylase are enzymes which incorporate CO_2 in the form of bicarbonate into organic compounds through the formation of new carbon-carbon bonds:

$$R\text{-}CO\text{-}CH_3 + 2ATP + HCO_3^- + H_2O + H^+ \rightarrow R\text{-}COH\text{-}CH_2\text{-}CO_2^- + 2\ ADP + 2\ P_i,$$

Where "R" is a methyl or phenyl group in the case of acetone carboxylase or acetophenone carboxylase, respectively. For acetone carboxylase, one ATP is converted into one AMP and two P_i. These enzymes contain a PPII domain consisting of eight rather short PPII helices [16,17]. Remarkably, one PPII helix is surrounded by six others and has a high glycine content and the whole PPII helical bundle domain is buried and surrounded by other enzyme domains[16,17]. These structures show that PPII helical bundle domains need not be long or solvent exposed, which would afford water molecules as H-bonding partners, to be stable.

The bacteriophage S16, which infects Salmonella bacteria, has a long tail fiber composed of several domains. The last domain (gp38), right at the fiber's tip, is composed of a complex glycine-rich PPII helical bundle [18]. Composed of ten PPII helices, this domain included two PPII helices which are completely surrounded by other PPII helices and are composed completely of glycine (**Figure 6**). In fact, for a given PPII helix, its glycine content in all these proteins increases with the number neighboring PPII helices (**Figure 8A**). The phage gp38 domain's amino acid sequence is well conserved across T-even bacteriophages (except T4). It serves as a stable platform for five surface loops which are highly variable in length and sequence. These loops recognize outer membrane proteins on the surface of prey bacteria which is the key initial event in infection. Residue substitutions in these loops have been shown to alter prey specificity. Amyloid stability and polymorphism is frequently impacted by the sequence of neighboring segments. By contrast, the PPII helical bundle of the phage gp38 domain remains stable despite the variation of the surface loops' sequence and this speaks to its conformational stability and the robustness of the fold.

Human Anaplastic Leukemia Kinase is a receptor tyrosine kinase that plays key physiological roles in neuronal development and memory[26] although its aberrant signaling is linked to obesity and cancer [27]. The extracellular glycine-rich domain is key for binding peptide activators [28] and is composed of a remarkably large bundle of fourteen PPII helices [29]. Three of these helices are surrounded by other PPII helices and are composed almost exclusively of glycine residues[19] (**Figure 7**). Some of the segments connecting the PPII helices are very short and consist of just a few residues. Other connectors are long contain loops, β-hairpins or α-helices (**Figure 7**).

Impact of flanking segments. Residues in flanking segments have well known to impact the formation and stability of β-sheet assemblies in amyloids [23]. For example, Ala2 is on the edge of the Aβ amyloid structure and undergoes transitory unfolding events [30]. Nevertheless, substitution of Ala2 by Val is impactful, being protective against Alzheimer's disease when heterozygous and disease-linked when homozygous [31] while A2T is protective [32,33]. In the case of the short *Hh*AFP, all flanking regions and loops connecting the PPII helices are short (2 – 4 residues only) and rich in small polar residues; namely Gly, Cys, Asn and Thr, as well as Pro, which is a good turn forming residue (**Figure 2, Figure 8B**). This same trend is seen in a homologous six PPII helix AFP from *Megaphorura arctica*[34] as well as in the longer antifreeze protein *Gr*AFP[12], although it also contains three Val and one Leu residue in the connecting turns. Substitution of these highly exposed hydrophobic residues by polar residues to remove a reserve hydrophobic effect [35] is expected to be stabilizing and could be tested experimentally. In the case of *Aa*Apc, the content of aromatic flanking residues is notable and is consistent with this PPII helical assembly being mostly buried inside the protein. Interestingly, one of the aromatic flanking residues (W440) contacts a buried potassium ion forming a cation-π interaction. Some of the PPII helices; namely helix 7 of GsAFP, helix 4 of AaAPC, helices 3 and 4 of S16LTF, and helix 1 of *Hs*ALK have a cationic residue right at the C-terminus or in the first two residues following it. These residues are expected to contribute favorable charge-PPII helix macrodipole interactions[36]. With regard to the PPII domain of Obg, which is a translation regulator, its flanking residues are notably rich in cationic residues and play key roles in recognizing the ribosomal peptidyl transferase center [37].

Packing Density: The formation of hydrogen bonds by buried peptide groups was reported to increase their packing density[38]. This strengthens van der Waals interactions and contributes favorably to protein conformational stability[38]. The short NH··OC hydrogen bond distances in *Hh*AFP's polyproline II helices compared to those of α-helices and β-sheets reported previously[20] led us to consider whether its packing density is higher. Interestingly, the *Hh*AFP does have a high packing density relative to other glycine rich PPII helical bundle proteins (**Table 2**). Nevertheless, a comparison of packing densities suggests that in general, glycine rich PPII helical bundles have a packing density similar to proteins composed of α-helices and β-sheet secondary structures (**Table 2**). This would imply that energetic contributions from van der Waals interactions would be similar.

Parallel versus Antiparallel Configuration: Polyproline II helices have a macrodipole with an excess of positive charge near the N-terminus and negative charge near the C-terminus [39]. A similar macrodipole is present in α-helices and a strong tendency towards antiparallel arrangements has been observed in proteins composed of four long α-helices arranged in a bundle[40]. This led us to assess the PPII helix orientation in glycine-rich PPII helical bundle domains. We found that there is a clear trend towards antiparallel orientation as 96% of successive PPII helices have their N-terminus placed on the same side as the C-terminus of the preceding PPII helix (**Figure 2-7**). Indeed, no PPII helix in this set of proteins has all parallel neighbors, all have at least one antiparallel neighbor, and all PPII helices have at least half of their neighbors in an antiparallel orientation (**Sup. Table 1**). Antiparallel orientations place helix macrodipoles in energetically favorable configurations and could thereby contribute to the conformational stability of these proteins.

Putative Networks of Cα-H···O=C H-bonds: Using the geometry criteria (Cα-H··O=C distance and bond angle cutoffs of 3.75 Å and 135º) described in the Methods Section, we identified 44 potential Cα-H···O=C H-bonds in *Gs*AFP, 15 in *Bs*Obg, 10 in *Ac*APC, 26 in S16LTF and 37 in *Hs*ALK. A complete list of the donors, acceptors, distances and angles of these H-bonds is listed in **Sup. Table 2**. The somewhat lower number of Cα-H···O=C H-bonds are present in *Aa*APC (**Sup. Table 2**) may be due to the irregular geometry of helices 1 and 6. Moreover, the glycine-rich PPII helical bundle domain of AaAPC is mostly buried. Thus, it may be stabilized by contacts with the

surrounding protein elements and therefore require fewer hydrogen bonds. Whereas the existence of these H-bonds should be corroborated by other methods such as NMR spectroscopy in the future, and even though these H-bonds are weaker than canonical N-H··O=C H-bonds, due to their large numbers these Cα-H···O=C H-bonds are likely to provide an important source of conformational stability.

Conclusions: In the last twenty years, the 3D structures of several proteins or protein domains formed by glycine-rich polyproline II helical bundles have been determined. By comparing and analyzing these structures, we have found that they share some common architectonic principles; in particular, they: 1) are tolerant of different loop lengths and flanking residues, 2) are as tightly packed as proteins composed of α-helices and β-sheets, 3) strongly prefer anti-parallel configurations and 4) generally have extensive networks of Cα-H···O=C hydrogen bonds. Due to the three dimensional and inter-helical nature of their NH··O=C and Cα-H···O=C hydrogen bonds, glycine-rich polyproline II domains have a internal stiffness and resistance to twisting or bending that sets them apart from α-helices and β-sheets and could make them a novel structural element for protein design. Protein designers have already started to incorporate polyproline II helices as a building block [41] and a binding target [42], but have not grafted a glycine-rich polyproline II helical bundle domain yet. . Future work in this direction could well benefit from a consideration of the architectonic principles identified here.

References

1. Levitt M, Chothia C (1976) Structural patterns in globular proteins. Nature 261:552–558.
2. Baker D (2019) What has de novo protein design taught us about protein folding and biophysics? Protein Sci 28:678–683.
3. Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34:167–339.
4. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.
5. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science 373:871–876.
6. Laurents DV (2022) AlphaFold 2 and NMR Spectroscopy: Partners to Understand Protein Structure, Dynamics and Function. Front Mol Biosci 9:906437.

7. COWAN PM, MCGAVIN S, NORTH AC (1955) The polypeptide chain configuration of collagen. Nature 176:1062–1064.
8. Traub W (1969) Polymers of tripeptides as collagen models. V. An x-ray study of poly(L-prolyl-glycyl-glycine). J Mol Biol 43:479–485.
9. Lotz B, Keith HD (1971) The crystal structures of poly(LAla-Gly-Gly-Gly)II and Poly(LAla-Gly-Gly)II. J Mol Biol 61:195–200.
10. Crick FH, Rich A (1955) Structure of polyglycine II. Nature 176:780–781.
11. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, Kent SBH (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. J Am Chem Soc 130:9695–9701.
12. Scholl CL, Tsuda S, Graham LA, Davies PL (2021) Crystal waters on the nine polyproline type II helical bundle springtail antifreeze protein from Granisotoma rainieri match the ice lattice. FEBS J 288:4332–4347.
13. Buglino J, Shen V, Hakimian P, Lima CD (2002) Structural and biochemical analysis of the Obg GTP binding protein. Structure 10:1581–1592.
14. Kukimoto-Niino M, Murayama K, Inoue M, Terada T, Tame JRH, Kuramitsu S, Shirouzu M, Yokoyama S (2004) Crystal structure of the GTP-binding protein Obg from Thermus thermophilus HB8. J Mol Biol 337:761–770.
15. Gkekas S, Singh RK, Shkumatov AV, Messens J, Fauvart M, Verstraeten N, Michiels J, Versées W (2017) Structural and biochemical analysis of Escherichia coli ObgE, a central regulator of bacterial persistence. J Biol Chem 292:5871–5883.
16. Mus F, Eilers BJ, Alleman AB, Kabasakal BV, Wells JN, Murray JW, Nocek BP, DuBois JL, Peters JW (2017) Structural Basis for the Mechanism of ATP-Dependent Acetone Carboxylation. Sci Rep 7:7234.
17. Warkentin E, Weidenweber S, Schühle K, Demmer U, Heider J, Ermler U (2017) A rare polyglycine type II-like helix motif in naturally occurring proteins. Proteins 85:2017–2023.
18. Dunne M, Denyes JM, Arndt H, Loessner MJ, Leiman PG, Klumpp J (2018) Salmonella Phage S16 Tail Fiber Adhesin Features a Rare Polyglycine Rich Domain for Host Recognition. Structure 26:1573-1582.e4.
19. Reshetnyak AV, Rossi P, Myasnikov AG, Sowaileh M, Mohanty J, Nourse A, Miller DJ, Lax I, Schlessinger J, Kalodimos CG (2021) Mechanism for the activation of the anaplastic lymphoma kinase receptor. Nature 600:153–157.
20. Gates ZP, Baxa MC, Yu W, Riback JA, Li H, Roux B, Kent SBH, Sosnick TR (2017) Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. Proc Natl Acad Sci U S A 114:2241–2246.
21. Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M, Laurents DV (2018) The Singular NMR Fingerprint of a Polyproline II Helical Bundle. J Am Chem Soc 140:16988–17000.
22. Treviño MÁ, López-Sánchez R, Moya MR, Pantoja-Uceda D, Mompeán M, Laurents DV (2022) Insight into polyproline II helical bundle stability in an antifreeze protein denatured state. Biophysical Journal [Internet] 121:4560–4568. Available from: https://doi.org/10.1016/j.bpj.2022.10.034
23. Ulamec SM, Brockwell DJ, Radford SE (2020) Looking Beyond the Core: The Role of Flanking Regions in the Aggregation of Amyloidogenic Peptides and Proteins. Front Neurosci 14:611285.
24. Rother K, Hildebrand PW, Goede A, Gruening B, Preissner R (2009) Voronoia: analyzing packing in protein structures. Nucleic Acids Res 37:D393-395.

25. Liang J, Dill KA (2001) Are proteins well-packed? Biophys J 81:751−766.
26. Weiss JB, Xue C, Benice T, Xue L, Morris SW, Raber J (2012) Anaplastic lymphoma kinase and leukocyte tyrosine kinase: functions and genetic interactions in learning, memory and adult neurogenesis. Pharmacol Biochem Behav 100:566−574.
27. Hallberg B, Palmer RH (2016) The role of the ALK receptor in cancer biology. Ann Oncol 27 Suppl 3:iii4−iii15.
28. Reshetnyak AV, Murray PB, Shi X, Mo ES, Mohanty J, Tome F, Bai H, Gunel M, Lax I, Schlessinger J (2015) Augmentor α and β (FAM150) are ligands of the receptor tyrosine kinases ALK and LTK: Hierarchy and specificity of ligand-receptor interactions. Proc Natl Acad Sci U S A 112:15862−15867.
29. Li T, Stayrook SE, Tsutsui Y, Zhang J, Wang Y, Li H, Proffitt A, Krimmer SG, Ahmed M, Belliveau O, et al. (2021) Structural basis for ligand reception by anaplastic lymphoma kinase. Nature 600:148−152.
30. Lu J-X, Qiang W, Yau W-M, Schwieters CD, Meredith SC, Tycko R (2013) Molecular structure of β-amyloid fibrils in Alzheimer′s disease brain tissue. Cell 154:1257−1268.
31. Di Fede G, Catania M, Morbin M, Rossi G, Suardi S, Mazzoleni G, Merlin M, Giovagnoli AR, Prioni S, Erbetta A, et al. (2009) A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis. Science 323:1473−1477.
32. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, et al. (2012) A mutation in APP protects against Alzheimer′s disease and age-related cognitive decline. Nature 488:96–99.
33. Benilova I, Gallardo R, Ungureanu A-A, Castillo Cano V, Snellinx A, Ramakers M, Bartic C, Rousseau F, Schymkowitz J, De Strooper B (2014) The Alzheimer disease protective mutation A2T modulates kinetic and thermodynamic properties of amyloid-β (Aβ) aggregation. J Biol Chem 289:30977−30989.
34. Graham LA, Boddington ME, Holmstrup M, Davies PL (2020) Antifreeze protein complements cryoprotective dehydration in the freeze-avoiding springtail Megaphorura arctica. Sci Rep 10:3047.
35. Pakula AA, Sauer RT (1990) Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. Nature 344:363−364.
36. Huang K-Y, Horng J-C (2019) Impacts of the Terminal Charged Residues on Polyproline Conformation. J Phys Chem B 123:138−147.
37. Feng B, Mandava CS, Guo Q, Wang J, Cao W, Li N, Zhang Y, Zhang Y, Wang Z, Wu J, et al. (2014) Structural and functional insights into the mode of action of a universally conserved Obg GTPase. PLoS Biol 12:e1001866.
38. Schell D, Tsai J, Scholtz JM, Pace CN (2006) Hydrogen bonding increases packing density in the protein interior. Proteins 63:278−282.
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. Sheridan RP, Levy RM, Salemme FR (1982) alpha-Helix dipole model and electrostatic stabilization of 4-alpha-helical proteins. Proc Natl Acad Sci U S A 79:4545−4549.
41. Baker EG, Williams C, Hudson KL, Bartlett GJ, Heal JW, Porter Goff KL, Sessions RB, Crump MP, Woolfson DN (2017) Engineering protein stability with atomic precision in a monomeric miniprotein. Nat Chem Biol 13:764−770.
42. Wu K, Bai H, Chang Y-T, Redler R, McNally KE, Sheffler W, Brunette TJ, Hicks DR, Morgan TE, Stevens TJ, et al. (2023) De novo design of modular peptide-binding

proteins by superhelical matching. Nature 616:581–589.

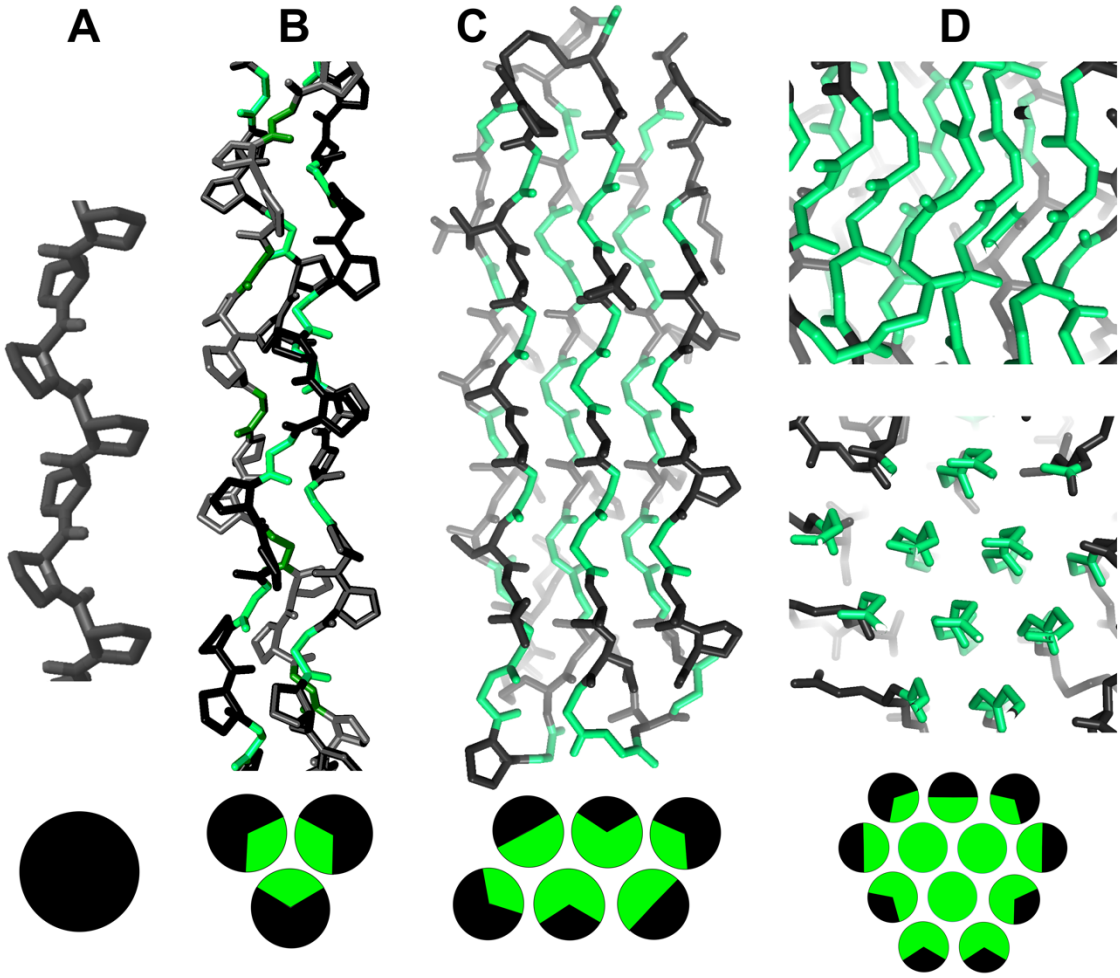
Table 1

Protein	PDB code	Independent Protein (IP) or Protein Domain (PD)	Characteristics	Reference
Hypogustrura harveyi antifreeze protein (*Hh*AFP). **Fig 2.**	3BOG	IP	Six PPII helices arranged in two layers	Pentelute et al. 2008
Granisotoma rainieri antifreeze protein (*Gr*AFP). **Fig 3.**	7JJV	IP	Nine PPII helices arranged in two layers.	Scholl et al. 2021
Bacillus subtilis Obg GTP-binding protein (*Bs*Obg) **Fig 4.**	1LNZ	PD	Six PPII helices arranged in two layers.	Buglino *et al.* 2002
T. thermophilius Obg GTP-binding protein. (*Tt*Obg)	1UDX	PD	Six PPII helices arranged in two layers.	Kukimoto-Niino *et al.* 2004
E. coli Obg GTP-binding protein (*Ec*Obg)	5M0Y	PD	Six PPII helices arranged in two layers.	Gkekas *et al.* 2017
Aromatoleum aromaticum Acetophenone carboxylase (*Aa*Apc) **Fig 5.**	5L9W	PD	Seven PPII helices arranged in three layers	Warkentin *et al.* 2017
Xanthobacter autotrophicus Acetone carboxylase (*Xa*Atc)	5SVC	PD	Seven PPII helices arranged in three layers	Mus *et al.* 2017
Salmonella φS16 long tail fiber (S16LTF). **Fig 6.**	6F45	PD	Ten PPII helices arranged in three layers	Dunne *et al.* 2018
Homo sapiens Anaplastic Leukemia Kinase. (*Hs*ALK) **Fig 7.**	7N00	PD	14 PPII helices arranged in four layers	Reshetnyak *et al.* 2021

Table 2: Protein Packing Densities

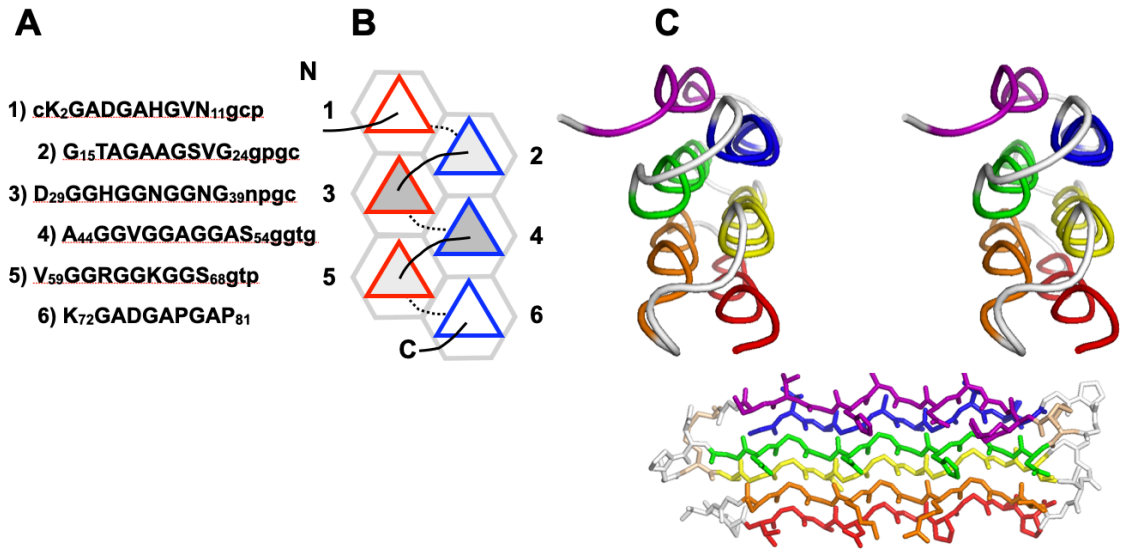
Protein	**PDB File**	**Secondary Structure**	**Packing Density** (Liang & Dill, 2001)	**Packing Density** (Measured here with Voronoia)
Myoglobin	1MBO	α-helices	0.716	0.729
Lysozyme	1LZL	α-helices (mostly) + some β-strands	0.725	0.719
Ribonuclease A	7RSA	Mostly β-strands, three α-helices	0.725	0.791
SH3 domain	1PGX	Mostly β-strands, one α-helix	0.762	0.740
*Aa*AFP	3BOB	PPII helical bundle		0.751
*Gr*AFP	7JJV	PPII helical bundle		0.743
*Bs*Obg	1LNZ	PPII helical bundle		0.742
*Tt*Obg	1UDX	PPII helical bundle		0.738
*Ec*Obg	5M0Y	PPII helical bundle		0.728
*Aa*Apc	5L9W	PPII helical bundle		0.705
*Xa*Apc	5SVC	PPII helical bundle		0.707
S16LTF	6F45	PPII helical bundle		0.732
*Hs*ALK	7N00	PPII helical bundle		0.720

Figure 1. Proline and Glycine Rich Sequences Adopt Polyproline II Helices



A. Polyproline peptides adopt isolated polyproline II helices in aqueous solution.
B. Substituting every third proline residue (black) for glycine (green) allows three chains to associate to form triple polyproline II helical structure of collagen.
C. Swapping of additional prolines by glycine permits proteins with two layers of polyproline II helices to form, such as snow flea antifreeze proteins.
D. Substitution of all proline residues by glycine gives rise to a multilayered "honeycomb" of polyproline II helices, such as that seen in peptides composed of just glycine [10] and Human Anaplastic Leukemia Kinase[19].

Figure 2. *Hypogustrura harveyi* antifreeze protein (*Hh*AFP).

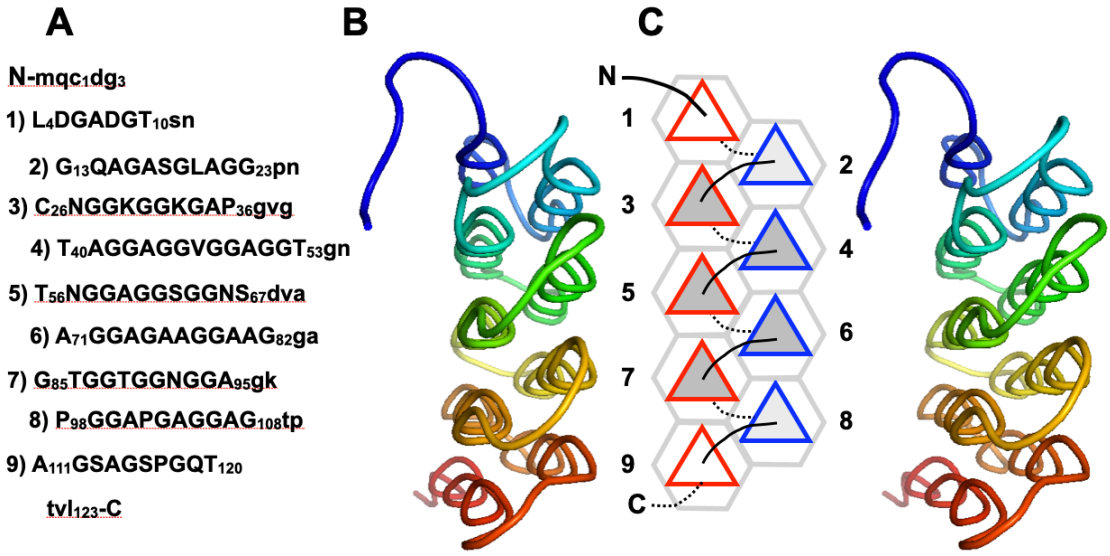


A. Primary structure of *Hf*AFP. Residues in PPII helices (which are numbered) are shown in capital letters, other residues in the connecting loops are written in lowercase.

B. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. PPII helices with two, three and four neighboring helices are shaded white, light gray and medium gray, respectively.

C. (*top*) Cross-eyed stereo ribbon diagram of the *Hh*AFP tertiary structure (PDB 3BOG) colored blue to red from the N- to the C-terminus, with connecting turns in gray, showing the six PPII helices end on (*top*) and along the long axis (*bottom*).

Figure 3. *Granisotoma rainieri* antifreeze protein

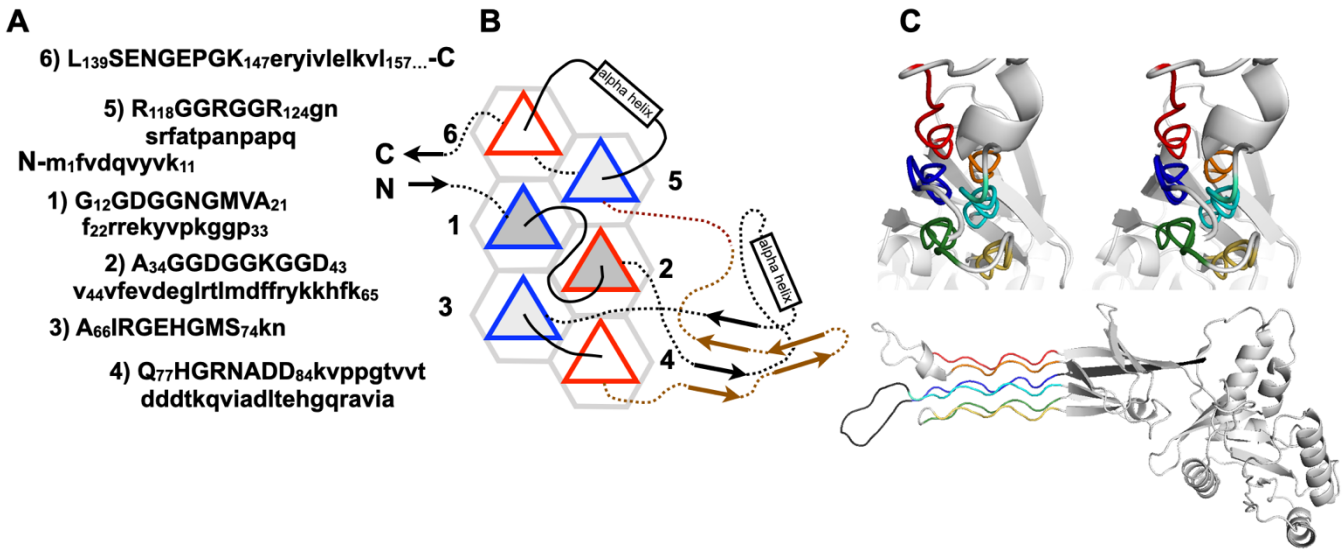


A. Primary structure of *Gr*AFP. Residues in PPII helices (which are numbered) are shown in capital letters, others are written in lowercase.

B. Cross-eyed stereo ribbon diagram of the *Gr*AFP tertiary structure (PDB 7JJV) colored blue to red from the N- to the C-terminus, showing the nine PPII helices end on.

C. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. PPII helices which two, three and four neighboring helices are shaded white, light gray and medium gray, respectively.

Figure 4. *Bacillus subtilis* Obg GTP-binding protein (*Bs*Obg)

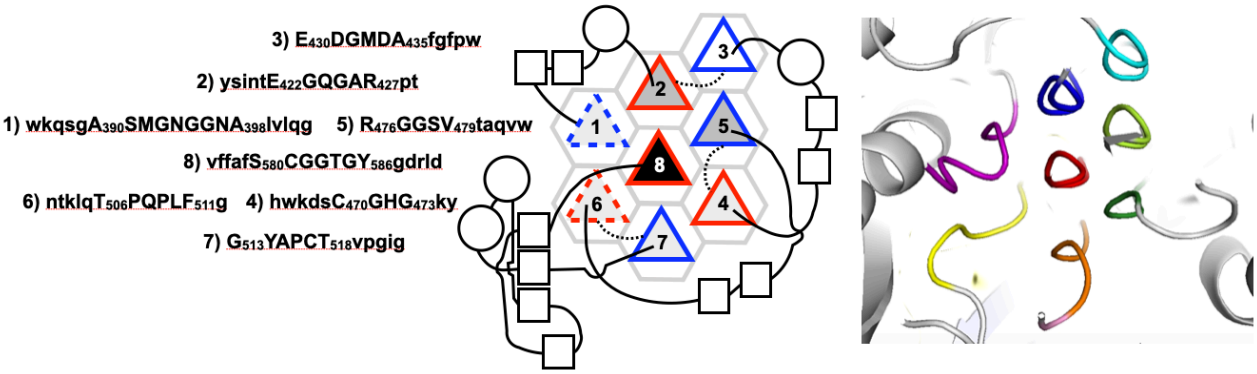


A. Primary structure of *Bs*Obg. Residues in PPII helices (which are numbered) are shown in capital letters, others are written in lowercase.

B. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. For clarity, the connections between PPII helices 4 and 5 are colored brown. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. PPII helices which two, three and four neighboring helices are shaded white, light gray and medium gray, respectively.

C. Cross-eyed stereo ribbon diagram of the *Bs*Obg tertiary structure (PDB 1LNZ) colored blue to red from the N- to the C-terminus, showing the six PPII helices end on (top). The six PPII helices together with the rest of the protein (light gray) are shown in the lower panel.

Figure 5. *Aromatoleum aromaticum* Acetophenone Carboxylase (*Aa*APC)

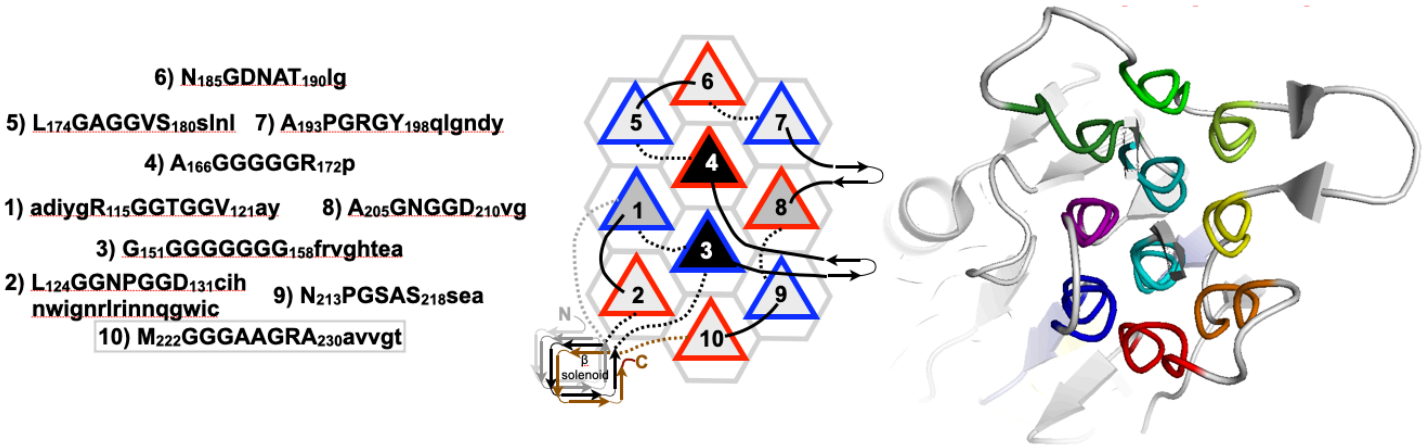


A. The primary structure of *Aa*APC PPII helices is shown, with the PPII helices numbered. Residues forming the PPII helices are written in capital letters; those of flanking sequence are shown in lower case.

B. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. The dashed outline of PPII helices 1 and 6 refers to their more deviant geometry. PPII helices with two, three, four and six neighboring helices are shaded white, light gray, medium gray and black, respectively. α-helices and β-strands of the connecting segments are represented as circles and squares, respectively. Squares which are aligned and close together represent β-strands hydrogen bonded in β-sheets.

C. The *Aa*APC tertiary structure (PDB 5L9W) with PPII helices colored purple, blue, cyan, dark green, lime green, yellow, orange and red from the N- to the C-terminus, showing the eight PPII helices end on.

Figure 6. *Salmonella* φS16 Long Tail Fiber gp38 Domain (S16LTF)



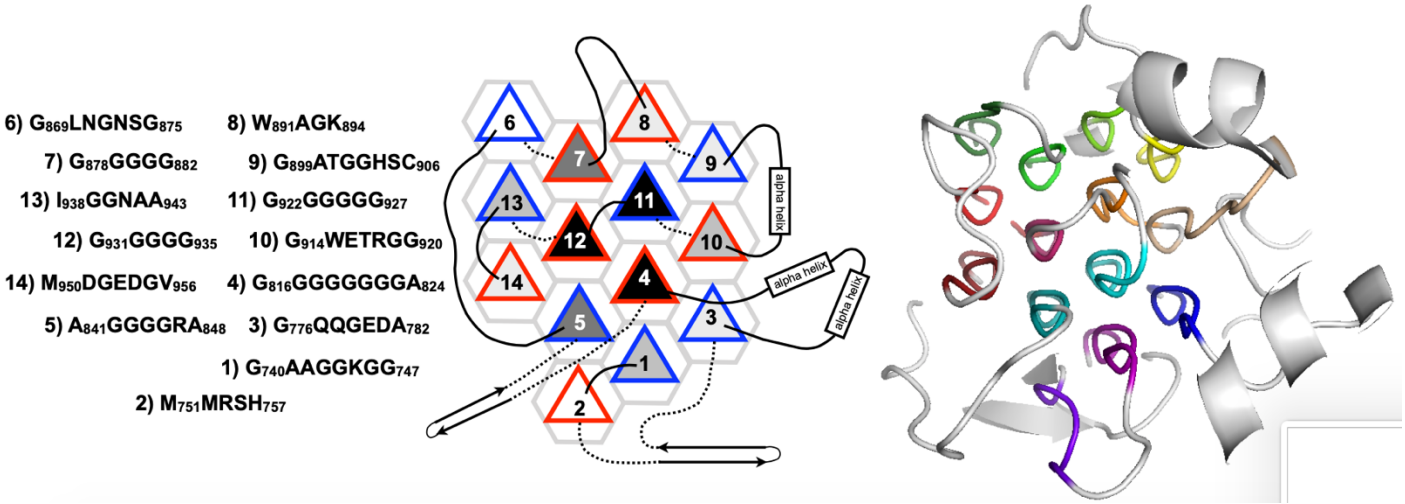
.

A. The primary structure of S16LFT gp38 domain PPII helices and flanking segments is shown, which PPII helices labeled. Residues forming the PPII helices are written in capital letters, those forming the connecting segments are written in lowercase.

B. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. PPII helices with three, four and six neighboring helices are shaded light gray, medium gray and black respectively. PPII helix 1 is preceded by two turns of a β-solenoid (gray), another layer of this β-solenoid is contributed by the segment linking PPII helices 2 and 3 (black) and a final layer is formed by the segment following PPII helix 10 (brown).

C. The S16LFT tertiary structure (PDB 6F45) colored purple to red from the N- to the C-terminus, showing the ten PPII helices end on.

Figure 7. *Homo sapiens* Anaplastic Leukemia Kinase (*Hs*ALK)

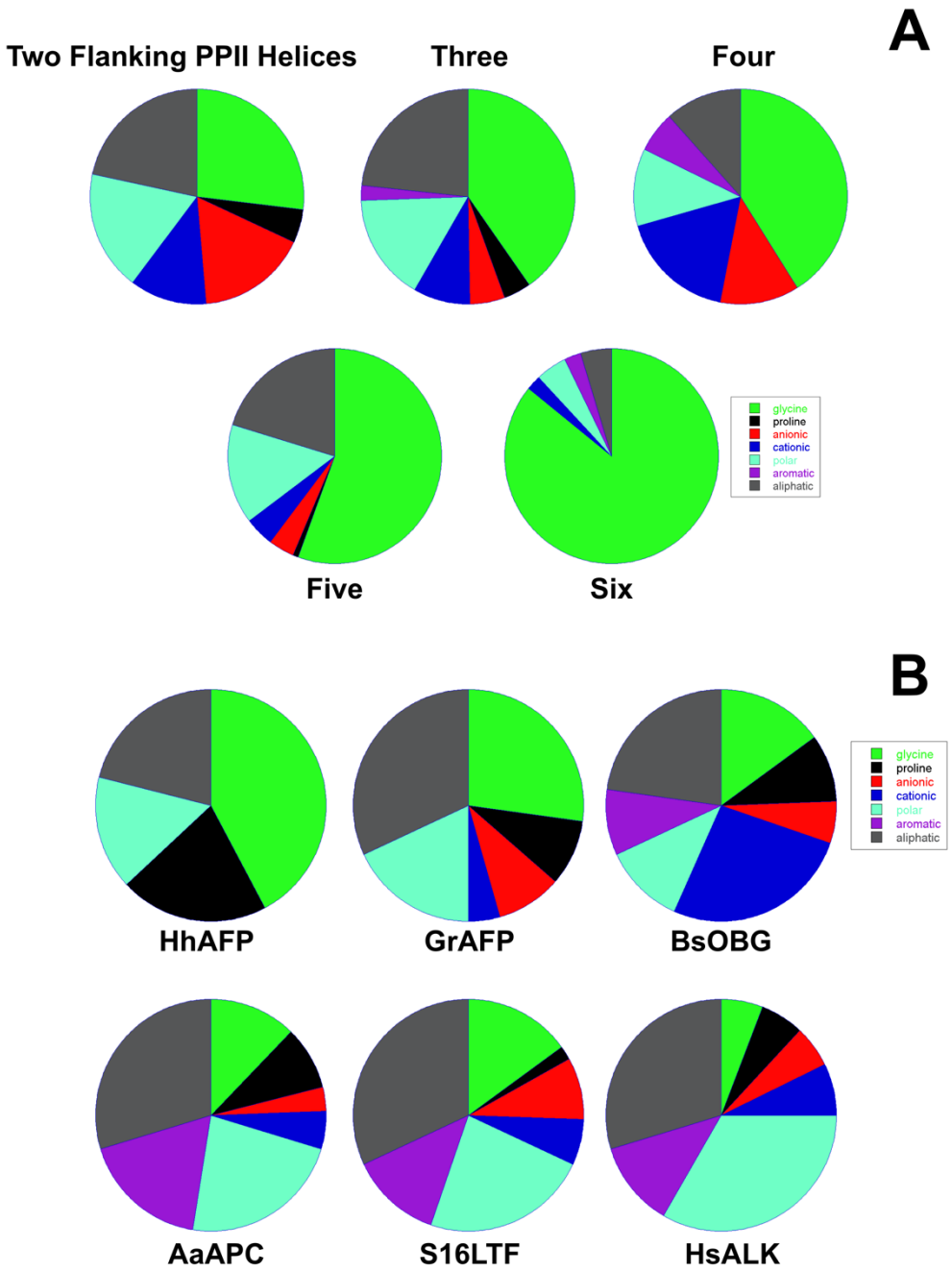


A. Primary structure of the fourteen PPII helices in *Hs*ALK. Helices are numbered according to the order along the sequence.

B. Schematic diagram representing the PPII helices as triangles. Connecting loops above the plane are shown as solid lines; those below the plane are shown as dotted lines. For clarity, the sequence of the connecting segments are not shown. PPII helical N-termini pointing up are outlined in red; N-termini pointing down are outlined in blue. PPII helices which two, three and four, five and six neighboring helices are shaded white, light gray, medium gray, dark gray and black respectively. PPII helices with 5 or 6 neighbors are labeled in white.

C. The tertiary structure of the PPII helical bundle domain of HsALK (PDB: 7N00) is shown end-on with helices colored in a rainbow spectrum from purple (helix 1) to dark red (helix 14).

Figure 8. Glycine Content Increases with the Number of Neighboring PPII helices and Glycine-Rich PPII Helical Assemblies are Tolerant of Diverse Flanking Sequences



A. The Glycine content increases when a PPII helix is surrounded by more PPII helices. Residue color code: Gly=light green, Pro=black, anionic (Asp, Glu)=red, cationic (Arg, Lys, His)=blue, polar (Thr, Ser, Asn, Gln, Cys-SH)=cyan, aromatic (Phe, Tyr, Trp)=purple, aliphatic (Ala, Met, Val, Leu, Ile, Cys S-S)=gray.

B. Glycine-rich PPII helical assemblies are robust enough to remain folded in the context of distinct flanking sequences, which can be rich in glycine and proline residues in the case of *Hh*AFP, cationic or aromatic residues for *Bs*OBG or *Aa*APC, respectively. The color code is the same as that used in panel **A**.

Supporting Table 1: Packing Density of Buried PPII Helices

*Hh*AFP· PDB: 3BOG

PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.769	2	KGADGAHGV	9
2	0.755	3	GTAGAAGSVG	10
3	0.741	4	DGGHGGNGGNG	11
4	0.737	4	AGGVGGAGGAS	11
5	0.766	3	VGGRGGKGGS	10
6	0.761	2	KGADGAPGAP	10
Average	0.751	3.00		10.167

*for buried positions.

*Gs*AFP PDB: 7JJV

PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.783	2	LDGADGT	7
2	0.756	3	GQAGASGLAGG	10
3	0.742	4	CNGGKGGKGAP	11
4	0.740	4	TAGGAGGVGGAGGT	14
5	0.731	4	TNGGAGGSGGNS	12
6	0.742	4	AGGAGAAGGAAG	12
7	0.740	4	GTGGTGGNGGA	11
8	0.721	3	PGGAPGAGGAG	11
9	0.773	2	AGSAGSPGQT	10
Average:	0.743	3.33		10.889

*Bs*OBG PDB: 1LNZ

PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.759	4	GGDGGNGMVA	10
2	0.760	4	AGGDGGKGGD	10
3	0.771	3	AIRGEHGMS	9
4	0.784	2	QHGRNADD	8
5	0.737	3	RGGRGGR	7
6	0.729	2	LSENGEPGK	9
Average:	0.742	3.00		8.83

Supporting Table 1: Packing Density of Buried PPII Helices *(Continued)*

*Aa*APC PDB: 5L9W

PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.708	*3*	ASMGNGGNA	9
2	0.708	4	EGQGAR	6
3	0.740	2	EDGMDAFG	8
4	0.686	3	CGHG	4
5	0.735	4	RGGV	4
6	0.762	*3*	TPQPL	5
7	0.703	3	GYAPCT	6
8	0.677	6	SCGGTGYG	8
average:	0.705	3.50		6.25

* Helices 1 and 6 are irregular.

S16LTF PDB: 6F45

PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.746	4	RGGTGGV	7
2	0.750	3	LGGNPGGD	8
3	0.722	6	GGGGGGGG	8
4	0.741	6	AGGGGGR	7
5	0.739	3	LGAGGVS	7
6	0.759	3	NGDNAT	6
7	0.773	3	APGRGY	6
8	0.765	4	AGNGGD	6
9	0.774	3	NPGSAS	6
10	0.738	3	MGGGAAGRA	9
Average:	0.732	3.80		7

Supporting Table 1: Packing Density of Buried PPII Helices (*Continued*)

*Hs*ALK	PDB: 7N00			
PPII helix	Packing Density*	Number of bordering PPII helices	Sequence	Length
1	0.682	4	GAAGGKGG	8
2	0.664	2	MMRSH	5
3	0.741	3	GQQGEDA	7
4	0.699	6	GGGGGGGGA	9
5	0.699	5	AAGGGGRA	8
6	0.721	2	GLNGNSG	7
7	0.721	5	GGGGG	5
8	0.739	3	WAGK	4
9	0.722	3	GATGGHSC	8
10	0.707	4	GWETRGG	7
11	0.700	6	GGGGGG	6
12	0.703	6	GGGGGG	6
13	0.740	4	YIGGNAA	7
14	0.742	3	MDGEDGV	7
Average:	0.720	4.00		6.71

Supporting Table 2: Cα-H···O=C Hydrogen Bonds

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

*Hh*AFP PDB: 3BOG

PPII helices	Donor	Acceptor	Distance	Angle
1 <=> 2	G6	A19	3.5	167.4
	G9	T16	3.6	157.7
1 <=> 3	G31	A4	3.4	165
	G34	A7	3.6	169.5
	G37	V10	3.4	159.4
2 <=> 3	G21	G31	3.5	164.7
	G18	G34	3.4	165
	G15	G37	3.4	161.3
2 <=> 4	G46	T16	3.7	169.1
	G49	A19	3.6	170.8
	G52	S22	3.5	170.9
	G55	G25	3.4	171.1
3 <=> 4	G30	G52	3.4	155.5
	G33	G49	3.5	173.9
	G36	G46	3.4	166.5
3 <=> 5	G61	G31	3.3	171.6
	G64	G34	3.3	168.7
	G67	G37	3.5	170.6
4 <=> 5	G45	G67	3.6	163
	G48	G64	3.6	163
	G51	G61	3.5	167.7
4 <=> 6	G73	G46	3.6	166.9
	G76	G49	3.6	172.8
	G79	G52	3.5	176.8
5 <=> 6	G60	G79	3.6	162
	G63	G76	3.4	179.6
	G66	G73	3.6	169.5

Supporting Table 2: Cα-H···O=C Hydrogen Bonds *(Continued)*

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

*Gs*AFP PDB 7JJV

PPII helices	Donor	Acceptor	Distance	Angle
1 <=> 2	G6	A17	3.5	161.9
	G9	Q14	3.7	170.9
1 <=> 3	G29	L4	3.4	166.6
	G32	A7	3.6	166.9
2 <=> 3	G16	G32	3.6	174
	G19	G29	3.6	167.3
2 <=> 4	G43	Q14	3.6	174
	G46	A17	3.6	167.8
	G49	L20	3.6	164.7
	G52	G23	3.4	172
3 <=> 4	G28	G49	3.3	172.4
	G31	G46	3.2	157.9
	G34	G43	3.5	167.6
3 <=> 5	G59	G29	3.5	171.9
	G62	G32	3.4	169
	G65	A35	3.5	175.8
	S67	G37	3.4	150.4
4 <=> 5	G42	G65	3.5	175.4
	G45	G62	3.5	169.1
	G48	G59	3.4	157
4 <=> 6	G73	G43	3.5	176.8
	G79	G49	3.5	174.9
	G82	G52	3.9	173.8
5 <=> 6	G58	G79	3.3	169.1
	G61	A76	3.5	164.5
	G64	G73	3.4	172.7
5 <=> 7	G85	T56	3.3	168.4
	G88	G59	3.5	164.5
	G91	G62	3.5	174.1
	G94	G65	3.5	172.8
6 <=> 7	G72	G94	3.4	176.6
	G75	G91	3.8	163.3
	G78	G88	3.6	174.4
6 <=> 8	G100	G73	3.5	174.6
	G103	A76	3.5	164.1
	G106	G79	3.4	169.7
7 <=> 8	G87	G106	3.4	163
	G90	G103	3.4	160.3
	G93	G100	3.4	173.2
7 <=> 9	G88	G112	3.5	166.2
	G91	G115	3.5	164.4

	G94	G118	3.3	169.9
8 <=> 9	G99	G118	3.6	170.7
	G105	G112	3.2	165.2

Supporting Table 2: Cα-H···O=C Hydrogen Bonds *(Continued)*

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

Protein:
*Bs*OBG PDB: 1LNZ

PPII helices	Donor	Acceptor	Distance	Angle
1 <=> 2	G38	G16	3.2	170.1
	G41	G13	3.9	167
1 <=> 3	G16	G70	3.5	166.7
1 <=> 5	G18	R124	3.6	169.6
1 <=> 6	G143	G15	3.5	168.5
	G146	G12	3.7	164.6
2 <=> 3	G69	G39	3.1	157.8
	G72	G36	3.6	163.9
2 <=> 4	G36	Q77	3.5	167.9
	G39	R80	3.4	164.1
2 <=> 5	G119	G41	3.5	158.1
	G122	G38	3.4	163
	G125	G35	3.5	151.4
3 <=> 4	G79	E70	3.4	164.8
5 <=> 6	G120	E144	3.4	153
	G123	E141	3.3	146.4

Supporting Table 2: Cα-H···O=C Hydrogen Bonds *(Continued)*

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

Protein:
*Aa*APC PDB: 5L9W

PPII helices	Donor	Acceptor	Distance	Angle
1<=>2	G425	A390	3.7	156.3
2<=>5	G423	G478	3.3	156.5
2<=>8	G583	Q424	3.7	160.6
3<=>5	G478	M433	3.1	142.8
4<=>8	G582	S469	3.3	147.2
4<=>8	G585	G472	3.3	174.4
5<=>8	G477	G583	3.7	164.2
6<=>8	Y586	L510	3.4	150.5
6<=>8	G587	F511	3.2	143.6
7<=>8	G513	G585	3.7	167.9

Supporting Table 2: Cα-H···O=C Hydrogen Bonds *(Continued)*

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

S16 LTF PDB 6F45

PPII helices	Donor	Acceptor	Distance	Angle
1 <=> 2	G114	G125	3.4	145
1 <=> 3	G152	R115	3.8	175.8
	G155	T118	3.6	174
1 <=> 4	G116	G171	3.1	168.7
	G119	G168	3.4	175.7
1 <=> 5				
2 <=> 3	G126	G155	3.3	173.9
	G129	G152	3.4	165.5
	G225	N127	2.4	174.2
	G228	G130	3.4	143.4
3 <=> 4	G168	G156	3.7	169.3
	G171	G153	3.3	169.5
3 <=> 8	G153	G209	3.4	171.5
	G156	G206	3.4	172.3
3 <=> 9	G215	G155	3.4	172.3
3 <=> 10	G151	G228	3.8	159.4
	G154	G225	3.5	170.2
	G157	M222	3.7	179.1
4 <=> 5	G170	G175	3.5	173.4
	G167	G178	3.5	177.1
4 <=> 6	G186	G168	3.4	173.8
4 <=> 7	G169	G195	3.3	160.9
4 <=> 8	G208	G170	3.3	169.7
5 <=> 6	G177	G186	3.5	
	L174	A189	3.5	163.2
6 <=> 7	G195	D187	3.5	158.1
7 <=> 8	G197	A205	3.4	155.3
8 <=> 9	G206	S216	3.4	162.2
	G209	N213	4.1	
9 <=> 10	G224	G215	3.3	160.1

Supporting Table 2: Cα-H···O=C Hydrogen Bonds *(Continued)*

Note: non-glycine residues are shaded violet. marginal H-bonds are shaded blue.

*Hs*ALK 7N00

PPII helices	Donor	Acceptor	Distance	Angle
1 <=> 2	G756	G740	3.4	158.1
1 <=> 3	G744	E780	3.4	155.8
1 <=> 4	G819	G744	3.9	172.2
	G822	A741	3.5	154.5
1 <=> 5	G740	A842	3.4	171.3
	G743	G845	3.5	174.1
	G746	A848	4.1	144.3
2 <=> 5				
3 <=> 4	G776	G823	3.7	152.1
	G779	G820	3.5	158.7
3 <=> 10	G920	Q777	3.2	152.6
4 <=> 5	G843	G822	3.6	154
	G846	G819	3.3	162
4 <=> 10	G820	R924	3.8	175.7
	G823	F921	3.4	165.9
4 <=> 11	G924	G821	3.2	130.3
	G927	G818	3.7	147.9
4 <=> 12	G821	G933	3.5	176.7
5 <=> 12	G931	G846	3.3	154.7
	G934	G843	3.5	157.3
5 <=> 14	G845	G952	3.5	166.2
6 <=> 7	G879	G872	3.9	165.2
6 <=> 13	G872	G940	3.5	170.6
	G875	A943	3.2	153.8
7 <=> 8	G878	W891	3.2	165.9
	G881	K894	3.3	157.1
7 <=> 11	G925	G878	3.3	150.4
7 <=> 12	G880	G934	3.2	161.7
7 <=> 13	G939	G880	3.3	164.3
8 <=> 9	A900	K894	3.4	175.3
	G903	W891	3.7	157.1
8 <=> 11	G893	G923	3.5	166.3
9 <=> 10	G902	G919	3.3	166.5
9 <=> 11	G926	H904	3.1	167
10 <=> 11	G919	G924	3.3	161.5
11 <=> 12	G932	G925	3.6	157
	G935	G922	3.8	174.3
12 <=> 13	G933	G940	3.7	154.5
12 <=> 14	G952	G932	3.3	171

	G955	G935	3.3	154.4
13 <=> 14	G940	E953	3.3	160.6