Journal of Materials Chemistry B

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal | View Issue

Cite this: J. Mater. Chem. B, 2026, 14, 1764

Molecular versatility of polyproline II helices: from natural proteins to biomimetic materials

Francisco C. Seco, †[ab] Sandro Amador, †[ab] Felipe Conzuelo, [a] Ana C. Baptista, [b] Leonor Morgado *[cd] and Ana S. Pina *[a]

Received 12th August 2025,
Accepted 5th December 2025

DOI: 10.1039/d5tb01841b

rsc.li/materials-b

The polyproline II helix is more common as a secondary structure than previously thought. Its significance extends to mediating various protein–protein interactions, both structurally and functionally, in natural systems. This structure is associated with the formation of supramolecular assemblies, which have been investigated for biomimetic applications. In this review, we highlight three examples of polyproline II helix utilization: structurally in collagen and in biomolecular condensates, and as functional motifs in hyperactive antifreeze proteins. The review examines the mechanisms underlying the properties of PPII and consolidates practical design principles for engineering PPII-based assemblies, with an emphasis on sequence composition, residue propensity, and crosslinking strategies that enhance stability and functionality. Additionally, by critically comparing spectroscopic methods (CD, VCD, ROA and NMR) and AI-based prediction tools, we summarize their respective strengths and limitations, providing a practical decision framework for selecting the most suitable characterization techniques for different sample types and research objectives.

Introduction

The prevalence of polyproline II (PPII) helix structures in nature is more significant than initially thought, when compared with the most common secondary structure motifs, α-helices and β-strands. Recent estimates suggest that at least 5–10% of all residues in globular proteins adopt PPII conformations, with even higher percentages observed in intrinsically disordered proteins (IDPs). Their prevalence demonstrates the structural and functional importance of PPII helices across diverse protein families, playing crucial roles in protein–protein interactions and molecular recognition processes.[1–3]

Interestingly, PPII helices share a close structural relationship with both α-helices and β-strands. A subtle alteration in just one of the dihedral angles of the PPII helix is sufficient to transition into either of these common secondary structures, highlighting the conformational versatility of PPII motifs.[2,4]

The PPII helix exhibits a left-handed orientation, featuring three residues per turn and a 3.1 Å rise per residue, characterized by distinct backbone peptide bonds, with specific dihedral angles of approximately −75° and 145° for ϕ and ψ, respectively. This configuration results in a shape resembling a triangular prism. Unlike other secondary structures, the PPII helix lacks internal hydrogen bonds. This absence leads to an extended conformation with increased exposure to the surrounding solvent for both the backbone and the side chains. Consequently, the PPII helix demonstrates lower stability compared to other helical conformations. The reduced stability of the PPII helix explains why molecules adopting this conformation frequently engage in interactions with different molecular structures to enhance their stability, often relying on external partners for hydration of the helix. However, the first crystal structure of an oligoproline adopting the PPII conformation revealed that, even in the absence of water molecules, the helix can be intrinsically stabilized by specific n → π* interactions between adjacent amide carbonyl groups, highlighting an internal mechanism contributing to PPII helix stability.[5]

In contrast, the α-helix is right-handed with 3.6 residues per turn and a 1.5 Å rise per residue. It features internal hydrogen bonds that significantly contribute to its stability. β-sheets, on the other hand, consist of extended strands arranged in a zigzag pattern, forming hydrogen bonds between adjacent strands rather than within a single chain.[3,4,6]

One of the key challenges concerning the detection of the PPII helix stems precisely from the unique structural characteristics. The absence of internal hydrogen bonds in PPII helices

[a] Instituto de Tecnologia Química e Biológica António Xavier (ITQB), Universidade NOVA de Lisboa, Av. da República, 2780-157, Oeiras, Portugal. E-mail: ana.pina@itqb.unl.pt
[b] CENIMAT/I3N, Department of Materials Science, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[c] UCIBIO – Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal. E-mail: mlmorgado@fct.unl.pt
[d] Associate Laboratory i4HB – Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
† The authors contributed equally, being co-first authors of the review article

confers remarkable flexibility and dynamic behavior, making them difficult to capture in static structural analysis. Such inherent flexibility allows PPII helices to rapidly transition between conformations, further complicating their identification. Moreover, PPII helices are frequently found in IDPs, which lack a stable tertiary structure under physiological conditions. The combination of PPII helices' inherent flexibility and their occurrence in highly dynamic IDPs creates a scenario where traditional structural determination methods, such as X-ray crystallography, may fail to accurately capture these elusive conformations.[7,8] Due to these factors and the initial use of protein structure assignment software, which only covered α and β structures, many PPII structures may have gone unnoticed. One of the bases of this software is a known methodology, the Ramachandran map.[9] This map is based on dihedral angles between planes of the backbone, with consideration for the possible steric clashes, yielding a map with possible and impossible regions for a known structure. This map was further improved to a point where each region can be considered specific for a certain secondary structure. For example, the PPII structure attribution zone of the map was found to be previously only occupied by β-strands.[3,10,11]

Amino acid propensity can also be a strategy to estimate the conformation of proteins. For PPII, it has been evaluated and even cross-checked in terms of the peptide length. Longer peptides have a higher PPII propensity when composed of residues like Ala, Met, Thr, Lys, and Leu, and for shorter ones, Asp, Ile, and Glu appear as PPII-promoting residues.[10,12] Regions rich in specific amino acids, primarily proline and glycine, are also associated with a greater propensity to form these structures.[13] The name of these structures comes precisely from the PPII helices that were initially associated with the proline-rich regions of polypeptides. This association stemmed from proline's unique side chain structure, which naturally satisfies the dihedral angle constraints required for PPII formation. However, subsequent research has revealed that the presence of proline is not a prerequisite for PPII structure formation. While proline does exhibit a higher propensity for PPII conformation compared to other amino acids, PPII helices can form in sequences with diverse amino acid compositions. This finding has broadened our understanding of PPII prevalence and its potential functional roles in proteins beyond proline-rich regions.[14]

Recent advancements in structural prediction (such as AlphaFold[15] and PEP-FOLD,[16] calculation and analysis software, along with the development of specialized tools like PPIIPRED[6] and PolyprOnline[11]) have significantly improved our ability to identify and characterize PPII structures. When combined with progress in spectroscopic techniques, these computational tools now enable researchers to detect and analyze PPII conformations that might have previously gone unnoticed. However, the detection and characterization of these structures remain complex processes, and work is still required in this area.[17]

Intriguingly, the challenges in detecting and characterizing PPII structures are closely linked to their functional significance in proteins. The conformational flexibility and environmental sensitivity of PPII helices, which complicate their identification, are the same features that contribute to their diverse roles in biological processes. Due to the structural characteristics of the PPII helix, especially its reliance on PPII–solvent or PPII–protein interactions for increased stability, there are many examples in nature of proteins with PPII-promoting motifs in their binding domains for improved function.

One major case of proteins with PPII-promoting motifs is the PPII-adopting antifreeze glycoproteins. The known PPII conformation of these proteins is essential for binding to ice, as the helix allows for a functional geometry that promotes the binding within the ice crystal lattice grooves, preventing recrystallization.[18]

Higher-order PPII-based supramolecular counterparts to PPII-adopting antifreeze glycoproteins are also common in nature, in the form of PPII-adopting antifreeze proteins. These proteins are known to have a higher degree of activity due to their ice-binding site (IBS), composed of PPII helices bundled together, which improves IBS flatness and increases surface area for ice-binding interactions.[19,20]

PPII-based supramolecular assemblies have also been linked to structural roles. The collagen triple helix, composed of three left-handed PPII helices intertwined to form a right-handed supercoiled triple helix, is the best natural example of the use of PPII helices for these roles, as collagen is the most abundant structural protein in the extracellular matrix.[21,22]

It is also theorized that the presence of PPII helices in proline-rich motifs and glycine-rich motifs may contribute to liquid–liquid phase separation through their ability to participate in multivalent and weak interactions. The extended conformation of PPII helices exposes a large surface area for potential interactions, allowing them to participate in both homotypic (PPII–PPII) and heterotypic (PPII with other motifs) associations.[23] A prime example is the Fused in Sarcoma (FUS) protein, which contains glycine-rich motifs that facilitate RNA binding, promote the formation of biomolecular condensates, and play crucial roles in transcriptional regulation.[24,25]

This review addresses and explores all these examples in which PPII structures, specifically in the formation of higher-order structures, are or could be involved in biological processes (Fig. 1). For each case, examples of artificial mimetic systems are also examined, which aim to replicate the natural properties of these systems to utilize them as tools in different scientific fields. Additionally, this review includes a chapter focusing on the primary methods of identifying PPII structures. This section is designed to provide the reader with an effective understanding of the most efficient techniques and analysis for studying this type of structure.

Collagen

In nature, PPII–PPII interactions can direct the assembly of higher-order structures, and the major example is collagen. There are about 28 types of collagen in vertebrate extracellular matrices,[26] organized by their structural characteristics and functional roles in the body, and the most prevalent are the fibrillar collagens (Fig. 2a and 3a), about 90%, which comprise type I, II, III, V, XI, XXIV, and XXVII collagen.

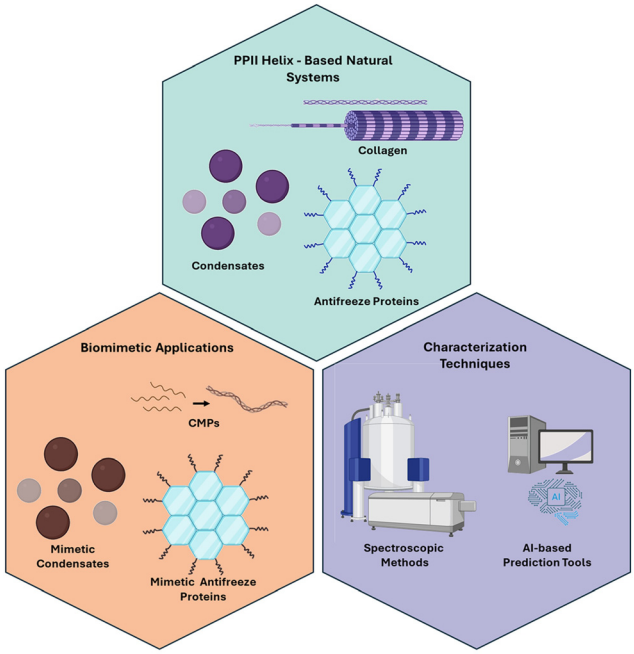

Fig. 1 Schematic overview illustrating the scope of the review: (top) natural systems featuring PPII helix-based assemblies; (left) biomimetic applications; (right) characterization toolbox integrating spectroscopic methods and AI-based computational tools for PPII systems.

The collagen molecule has a triple helix configuration, composed of three intertwined PPII helices, named PPII strands, which contain conserved motifs that promote PPII–PPII interaction-mediated assembly.[27,28] Each collagen type is defined by the organization and amino acid sequence of its PPII strands. Type I and type III collagen are used as models in many studies, as the former is formed by a heterotrimeric triple helix and the latter, a homotrimeric triple helix (Fig. 2a). A homotrimeric triple helix is formed by three PPII strands that share the same sequence, while a heterotrimeric triple helix can be formed by a combination of different PPII strands. If we name different strands in the helix as A, B, or C,[29] type I collagen would contain an AAB heterotrimeric triple helix (α1, α1, α2) while, for example, type XI collagen would have an ABC heterotrimeric triple helix (α1, α2, α3) (Fig. 2b). Key factors, such as the chain register and residue stagger of the triple helix, are crucial for the organization of the supramolecular assembly, mediated by the PPII–PPII interactions. Chain register refers to the orientation of each of the three PPII strands. A heterotrimer like collagen type I, for example, can be oriented in an AAB, BAA, or ABA register (Fig. 2b), one of which is promoted depending on the function of the specific collagen type and the stabilizing effect it has on the triple helix. The one residue stagger refers to PPII–PPII interactions at the sequence level, which requires each strand in the triple helix to have exactly one space between the terminal flank residue of one strand and the residue of the adjacent strand (Fig. 3b and c).[30]

Each PPII strand is usually composed of 10 repeats of Gly-Xaa-Yaa triplets, per 3 turns of the helix. While the Xaa and Yaa amino acid positions can be occupied by several amino acids, in native collagen, these positions are, in most cases, occupied by proline and hydroxyproline, the latter obtained from a prolyl 4-hydroxylation reaction upon a pre-existing proline residue in the Yaa position. There is also a record of an extremely low presence of aromatic amino acids within the PPII strand, and almost exclusively phenylalanine (Phe) (about 1%) is found.[31] The high propensity of proline–hydroxyproline in the triplets is key for the organization of the strands into PPII secondary structures,[32] and subsequently promotes the triple helix supramolecular configuration. Glycine residues also play a crucial role in the assembly of PPII strands in the triple helix conformation. A glycine residue from one PPII strand (PPII-A) forms hydrogen bonds with the backbone of an adjacent strand (PPII-B), typically involving the Xaa position. This arrangement allows the side chains of Xaa and Yaa residues to face outwards, exposed to the solvent. When proline and hydroxyproline

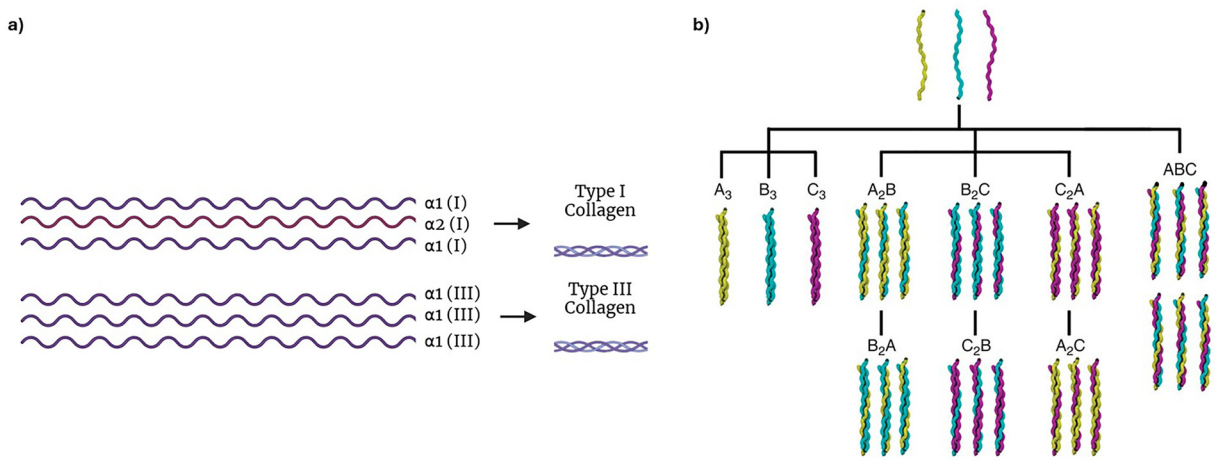

Fig. 2 (a) PPII strands present in heterotrimeric collagen type I, two α1 chains and one α2 chain, and homotrimeric type III with three α1 chains. Made with biorender.com. (b) Modes of structural organization of collagen triple helices for both homotrimers and heterotrimers. Reproduced from ref. 29, with permission from Springer Nature, Copyright 2012.

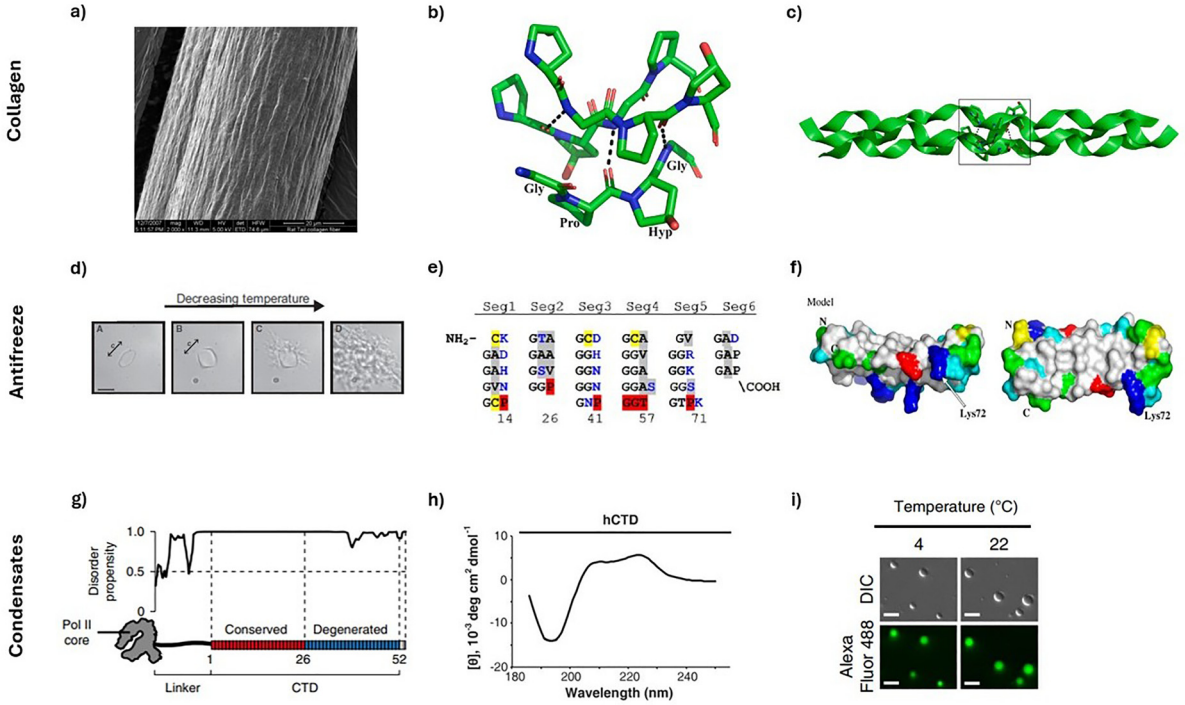

Fig. 3 (a) SEM image of collagen fiber. Reproduced from ref. 26, with permission from Elsevier, Copyright 2012. (b) Section of collagen triple helix crystal structure. Note the interstrand hydrogen bonds and the ring pucker of Pro and Hyp. Reproduced from ref. 33, with permission from Elsevier, Copyright 2022. (c) Collagen α-chains intertwined in the tripe helix structure. Reproduced from ref. 33, with permission from Elsevier, Copyright 2022. (d) Sequence of snow flea AFP (HhsfAFP) aligned in its structural segments. In yellow the cysteine residues, in gray small residues usually hydrophobic, in blue the charged or hydrophilic residues and in red the terminal prolines and other residues that could disrupt the PPII helix. Reproduced from ref. 75, with permission from John Wiley & Sons, Copyright 2021. (e) Effect of GrAFP in ice shaping with decreasing temperature. Reproduced from ref. 20, with permission from Elsevier Copyright 2007. (f) Side view of HhsfAFP (left) and putative IBS (right). Cysteines are colored yellow, prolines in green, charged residues in red and blue and uncharged polar residues in cyan. Hydrophobic residues in white. Note the flatness and propensity for hydrophobic residues of the IBS. Reproduced from ref. 20, with permission from Elsevier Copyright 2007. (g) Disorder analysis (top) and schematic view of the low complexity hCTD sequence of Pol II; (h) circular dichroism difference spectra of MBP-hCTD and MBP; (i) LLPS of MBP-hCTD in 16% dextran during incubation of the sample at different temperatures for 1 h. Fig. 3g–i reproduced from the ref. 114, with permission from Springer Nature, Copyright 2018.

occupy these Xaa and Yaa positions, respectively, their outward face configuration allows for an interaction between their rings to occur, the so-called ring pucker (Fig. 3b),[33] promoting the PPII structure of the strand. Proline can appear in *cis* or *trans* configuration, the latter being crucial for overall triple helix stability. The ring pucker of proline and derivatives is the main contributor to stabilizing stereoelectronic interactions. Proline promotes the occurrence of a C^{γ}-endo pucker, while hydroxyproline promotes the C^{γ}-exo pucker. The occurrence of the ring puckers is essential for the formation of *trans* peptide bonds, present in collagen, through $n \rightarrow \pi^*$ interactions (lone electron pair of an oxygen with an adjacent carbonyl group), which are stronger when in the presence of the C^{γ}-exo pucker conformation.[34,35] The assembly of the supramolecular structure occurs through an association event between the three PPII strands. This follows a ''zipper'' model where a trimer begins forming through the PPII–PPII interactions, also named triplet–triplet interactions, at the C-terminal, comprising hydrogen bonding and stereoelectronic interactions.[27,36]

The different factors mentioned thus far in triplet–triplet interactions, such as hydrogen bonding and stereochemical effects of the ring pucker, among others, which help in collagen type-specific register and stagger of the involved PPII strands,[37] have been studied extensively as part of an emerging field dedicated to the development of collagen mimics for several biotechnological applications, known as collagen mimetic peptides.[38]

Collagen mimetic peptides

Collagen mimetic peptides (CMPs) are capable of self-assembly in a similar fashion to innate collagen. They organize into an initial PPII strand, which can trimerize with other PPII strands, yielding a triple helix. This supramolecular assembly is then capable of fibrillating with other triple helices to form nanofibers. Some CMPs have been used for the self-assembly of hydrogels, since water molecules can be trapped and embedded into the nanofiber matrix.[39] The research into CMPs has focused on improving triple helix stability, mainly through the mimicry of the triplet region of collagen, narrowed down to assemblies of Gly-Pro-Hyp triplets.[40] Assemblies of mimetic peptides with variable amounts of these triplets still organize into a PPII structure, since the ring pucker conformation, especially the C^{γ}-exo pucker of Hyp,

reorganizes the peptide backbone chain to the respective φ and ψ torsion angles that promote the PPII helix.[13]

CMP improvement is based on obtaining sequences that can still adopt a PPII helix conformation, as there is a train of thought that improving propensity of the PPII helix will lead to better interactions among PPII monomers to form stable triple helices.[34,35] CMPs can form homotrimeric or heterotrimeric triple helices, needing to follow the correct chain register and one residue stagger between each PPII chain, to obtain the highest stability from the supramolecular assembly they originate. As such, several strategies have been developed to tackle this issue (Table 1), some even focused on changing the Gly-Pro-Hyp sequence by other types of repetitive Gly-Xaa-Yaa motifs.

There are 7 main strategies used so far when it comes to obtaining higher stability CMPs: PPII–PPII crosslink strategies, here exemplified by the on-resin chain branching method (aided by solid-phase peptide synthesis or SPPS) and the use of disulfide bridging, and amino acid propensity strategies which focus on the use of different amino acids to evaluate PPII and triple helix stability, including the use of charged, aromatic or non-canonical amino acids for improved PPII and triple helix stability.

Focusing on the PPII–PPII crosslink strategies, these strategies are based on the presence of a branch domain, aside from the Gly-Pro-Hyp or Gly-Xaa-Yaa motif domain that increases propensity for PPII and triple helix conformations. This branch domain contains motifs such as terminal Lys or Cys residues that can induce a covalent tether and improve CMP stability. The use of dilysine crosslinks consists of taking advantage of the Lys residue in the backbone, with the Lys amines in the resin during SPPS. Interestingly, these dilysine crosslinks also occur naturally between adjacent triple helices, during fibrillation.[51,52] This method was first used in homotrimeric triple helices,[41] and later for heterotrimeric triple helices.[42] Cys residues can be used to form disulfide bridges between themselves. Their use at the C-terminal of peptides with PPII/triple helix motifs (Gly-Pro-Hyp) has been reported. Ottl *et al.* showed how a heterotrimeric triple helix forming CMP (heterotrimer D), where each chain is tethered to another by disulfide bridges at the C-terminal, had a higher T_m (41 °C) than the heterotrimeric collagen type I (38 °C).[43] Later, Kotch and Raines took advantage of this disulfide bridging between differing length chains of (Gly-Pro-Yaa)$_n$, where Y could be Pro or Hyp, to form triple helices, evidenced by the positive peak near 225 nm and negative peak at 200–210 nm in their CD spectra. When the triple helices contained Hyp instead of Pro, the T_m was much higher (47 °C *vs.* 26 °C), which can be explained by the promotion of higher stereochemical stability interactions. These triple helices, even tethered by the disulfide bridge, could interact to form fibrils almost 400 nm in length.[53]

More recently, a novel disulfide bridge between homocysteine (Hcy) and Cys residues has been shown to significantly enhance the stability of CMPs compared to traditional Cys–Cys bridges (35 °C compared to 28 °C). These CMPs were based on a (PPG)$_{10}$ sequence trimer, where specific proline residues were substituted with either Cys or Hcy to form interstrand disulfide bonds. This Hcy–Cys bond better fits the natural triple helix geometry and reduces structural strain, providing a stronger covalent link for CMP stabilization and self-assembly.[54]

Amino acid propensity strategies for PPII and triple helix configurations work on changes to select areas of the Gly-Pro-Hyp triplets to improve stability. An initial study on the propensity of canonical amino acids for PPII configuration in a repetitive Pro guest peptide Ac-(Pro-X-Pro)$_n$-NH$_2$ was conducted in 2004. It was presented through computational methods that replacing X by Gly, Gln, or Ala in comparison with Pro was less desirable.[13]

Replacing the X or Y position of triple helix host peptides (Ac-(Gly-Pro-Hyp)$_3$-Gly-X/Pro-Hyp/Y-(Gly-Pro-Hyp)$_4$-Gly-Gly-CONH$_2$) can also influence the stability of the triple helix and provide information on the best motifs with PPII propensity, as per the work of Persikov and Brodsky. Here, Pro and Hyp show higher stability when in X and Y, respectively, followed by Glu and Ala for X, and Arg and Met for Y. The PPII propensity is led by Pro, but Gln and Arg quickly follow. The respective T_m for their triple helix-forming host peptides were all on par with proline, being above the 40 °C mark, with differences being attributed to possible distinct interchain and solvent interactions.[45] These propensities were later investigated, drawing a correlation between residue motifs interactions and increased triple helix thermal stability, with charged residues in Gly-Xaa-Yaa positions leading to a 17 °C increase in T_m and hydrophobic residues like Leu, only increasing the stability by 7 °C.[46]

Pairwise interactions, be they hydrophobic or electrostatic interactions, can alter the stability of the PPII helix and triple helix. In comparison with a (Gly-Pro-Hyp)$_2$ triple helix forming hexapeptide (T_m = 47.3 °C), both Gly-Pro-Lys-Gly-Glu-Hyp and Gly-Pro-Lys-Gly-Asp-Hyp hexapeptides exhibit similar or even greater stability (T_m = 47.8 °C and T_m = 47.1 °C, respectively) while maintaining their homotrimeric triple helix conformation in the respective CD spectra (maximum near 225 nm and minimum near 198 nm).[49] We argue that this pairwise interaction and the presence of Hyp in these peptides culminate in stabilizing electrostatic and stereochemical interactions, in contrast to the simple stereochemical and hydrogen bonding of the Gly-Pro-Hyp hexapeptide.

These pairwise interactions were also evaluated in contrast to different chain registers for the overall stability of heterotrimeric CMPs.[55,56] Gauba and Hartgerink prepared (Gly-Pro-Hyp)$_{10}$, (Gly-Glu-Hyp)$_{10}$, (Gly-Asp-Hyp)$_{10}$, (Gly-Pro-Arg)$_{10}$, and (Gly-Pro-Lys)$_{10}$ peptides. By mixing several of these chain peptides, they found a remarkably simple way to obtain triple helix ABC heterotrimeric peptides that could be as stable as homotrimeric arrangements of (Gly-Pro-Hyp) CMPs. This was shown through CD, when a (Gly-Asp-Hyp)$_{10}$-(Gly-Pro-Lys)$_{10}$-(Gly-Pro-Hyp)$_{10}$ presented a T_m = 65 °C in comparison with 3 (Gly-Pro-Hyp)$_{10}$, a known CMP (T_m = 67.5 °C). This was related to the capacity for side chains of lysine and aspartate to be ideally positioned to form salt bridges in the heterotrimer.[56] The authors considered the difference in stability to also be a case of a lower propensity to form triple helices compared to Gly or Hyp, which is in accordance with the previous studies based on computational methods, accounting for solvent and residue interaction.[45,49]

Seeing as these motifs appear as possible triple helix stabilizers, they were also evaluated for their capacity to induce the

Table 1 Strategies for stability enhancement of the CMPs triple helix and respective monomer sequence and secondary structure

Strategy		Sequence	Self-assembly	Structure	Thermal stability (T_m in °C)	Implications for material properties	Ref.
PPII–PPII crosslinks	On-resin chain branching	Ac-(Gly-Pro-Hyp)$_n$ – (Gly-X-Y)$_n$ – Ahx-Lys-Lys-Tyr-Gly	Branch of each chain (A1, A2, A3) – connection at C-terminal of Ahx(A1) – Lys1(A2) and Lys2(A2)-Ahx(A3)	PPII homotrimeric triple helix (circular dichroism – CD)	THP$_2$: (α1(IV)$_{1263–1277}$ + ((GPO),GVKGDKGNPG-WPGAP) – 58,5	Stabilizes triple helix; controls fibril nucleation; enhances hydrogel strength and thermal stability	41 and 42
	Disulfide bridging	Ac/H$_2$N-(Gly-Pro-Hyp)$_5$-(Gly-X-Y)$_4$-GC(C)G	Chain A1,B,A2 – disulfide bridging at C-terminal: Cys(A1)-Cys1(B) and Cys2(B) – Cys(A2)	PPII heterotrimeric triple helix	Heterotrimer D – 41	Mimics natural cystine bonds; facilitates nucleation; enhances folding kinetics; increases mechanical and enzymatic stability	43 and 44
Amino acid propensity	Canonical amino acids (PPII propensity)	Ac-(Pro)$_3$-X-(Pro)$_3$-Gly-Tyr-NH$_2$ X – Pro, Ala, Gln. Gly or Val	—	Computational methods (PPII content): Pro-Pro-Pro – 66% Pro-Ala-Pro – 46–66% Pro-Ala-Ala-Pro – 40–50% Pro-Gln-Pro – 44–66% Pro-Gly-Pro – 44-54% Pro-Val-Pro – 33–46%	—	Favors PPII formation; pre-organize chains for triple helix assembly; enhance thermal resilience	13
	Canonical amino acids (triple helix propensity)	(Ac-(Gly-Pro-Hyp)$_3$-Gly-X/Pro-Hyp/Y-(Gly-Pro-Hyp)$_4$-Gly-Gly-CONH$_2$ X or Y comprise all canonical amino acids	—	PPII Triple Helix (CD) Computational methods (Best 5 residues for triple helix Xaa/Yaa/PPII structure): Pro/Hyp/Pro Glu/Arg/Gln Ala/Met/Arg Lys/Ile/Lys Arg/Gln/Thr	Host peptide X/Host peptide Y: Pro – 47.3/Hyp – 47.3 Glu – 42.9/Arg – 47.2 Ala – 41.7/Met – 42.6 Lys – 41.5/Ile – 41.5 Arg – 40.6/Gln – 41.3	Control local stability, fibril nucleation; enable design of thermally robust biomaterials	45 and 46
	Non canonical amino acids	(Gly-Pro-Y)$_{10}$ Y – Pro, Hyp or Flp (4(R)-fluoro-L-proline)	Hyperstability from using Flp – higher promotor of C$^{\gamma}$-exo ring pucker and restriction of all torsion angles (φ, ψ, and ω) to PPII threshold	PPII triple helix (CD)	(Gly-Pro-Pro)$_{10}$ – 36 (Gly-Pro-Hyp)$_{10}$ – 67 (Gly-Pro-Flp)$_{10}$ – 87	Improves stereochemical stability; increases T_m; enhances mechanical modulus and durability	47 and 48
	Electrostatic interactions (charged Residues)	(Gly-Pro-Hyp)$_8$ where n = 4 is Gly-X-Y-Gly-X′-Y′ X or Y – Ala, Leu, Lys, Arg, Glu	Stabilizing interactions in a tandem between electrostatic interactions (Lys-Asp or Lys-Glu) and through the presence of Pro and Hyp.	PPII homotrimeric triple helix	Gly-Pro-Hyp-Gly-Pro-Hyp – 47.3 Gly-Pro-Lys-Gly-Glu-Hyp – 47.8 Gly-Glu-Hyp-Gly-Pro-Lys – 38 Gly-Glu-Lys-Gly-Pro-Hyp – 35	Stabilizes through ionic bonds; facilitates/Controls fibril organization; improves hydrogel mechanics & stability	49
	Hydrophobic aromatic interactions	CR – Ac-(Gly-Pro-Hyp)$_6$-(Pro-Arg-Gly)$_6$-NH$_2$ CK – Ac-(Gly-Pro-Hyp)$_6$-(Pro-Lys-Gly)$_6$-NH$_2$ CF – Ac-(Gly-Pro-Hyp)$_6$-(Phe-Hyp-Gly)$_6$-NH$_2$	Chain A1A2B: C-terminal cation – π interactions	Monomeric PPII peptides PPII homotrimeric and heterotrimeric triple helix (CD)	(CR)$_3$ – 42.6 (CR)$_2$(CF)$_1$ – 40.9 (CR)$_1$(CF)$_2$ – 40.2 (CK)$_3$ – 27.5 (CK)$_2$(CF)$_1$ – 28.1	Strengthens fibril formation; improves hydrogel stiffness and elasticity; potentially enhances biostability	50

fibrillation of several CMP triple helices. O'Leary *et al.* focused on Gly-Pro-Lys, Gly-Asp-Hyp and Gly-Pro-Hyp triplets to create a CMP $(Gly\text{-}Pro\text{-}Lys)_4$-$(Gly\text{-}Pro\text{-}Hyp)_4$-$(Gly\text{-}Asp\text{-}Hyp)_4$). This CMP was evaluated by CD, exhibiting a characteristic PPII spectra (maximum at 225 nm and minimum near 200 nm) and a thermal unfolding curve typical of a triple helix configuration, and was seen to form nanofibers through TEM, SEM, and AFM. The authors propose that by flanking a triple helix promoting motif (Gly-Pro-Hyp) with triplets that exhibit pairwise electrostatic interactions, it is possible for a Lys-Asp interchain salt bridge to form in a manner that creates a stagger between adjacent PPII mimetic chains, ending in a staggered homotrimeric triple helix that can assemble into full nanofibers, through their "sticky-ends" (Fig. 4a–c).[39] The mechanism shown here was also seen for other CMPs like the ones described by Chiu and Horng, where a host CMP peptide (H-Gly-Gly(Pro-Hyp-Gly)$_7$Gly-Gly-OH), when flanked by Glu and Lys (H-Glu-Gly(Pro-Hyp-Gly)$_7$Gly-Lys-OH), had a 0.9 °C increase in T_m.[57]

Hydrophobic interactions, such as pairwise cation-π interactions, have also been studied for improved stability of CMP, through triple helix stabilization.[50,58] A study on this matter was done by Chiang *et al.*, with host peptides $(Gly\text{-}Pro\text{-}Hyp)_6$-$(Gly\text{-}Pro\text{-}Yaa)_3$ and $(Gly\text{-}Pro\text{-}Hyp)_6$-$(Gly\text{-}Phe\text{-}Hyp)_3$. All peptides exhibited CD spectra on par with the PPII conformation (positive peak around 225nm and intense negative peak below 210 nm), and evidence of a thermal cooperative unfolding curve demonstrated the capacity for triple helix formation, consistent with the obtained nuclear magnetic resonance spectroscopy (NMR) spectra for these peptides. They found that when an AAB heterotrimer (2 chains with Arg and 1 chain with Phe) formed, it had a T_m = 40.4 °C, the highest stability of all heterotrimers, even compared to an AAB heterotrimer with 2 chains with Lys (T_m = 27.3 °C). The propensity for higher stability cation–π interactions between Arg-Phe in comparison with Lys-Phe was evidenced before,[58] which was mainly addressed by the difference in hydration of the positively charged guanidinium side-chain of Arg in comparison with the amino side chain of Lys, but we could also argue that higher number of geometries of Arg to participate in cation-π in comparison to Lys, would be a valid explanation as well. Like electrostatic pairwise interactions,

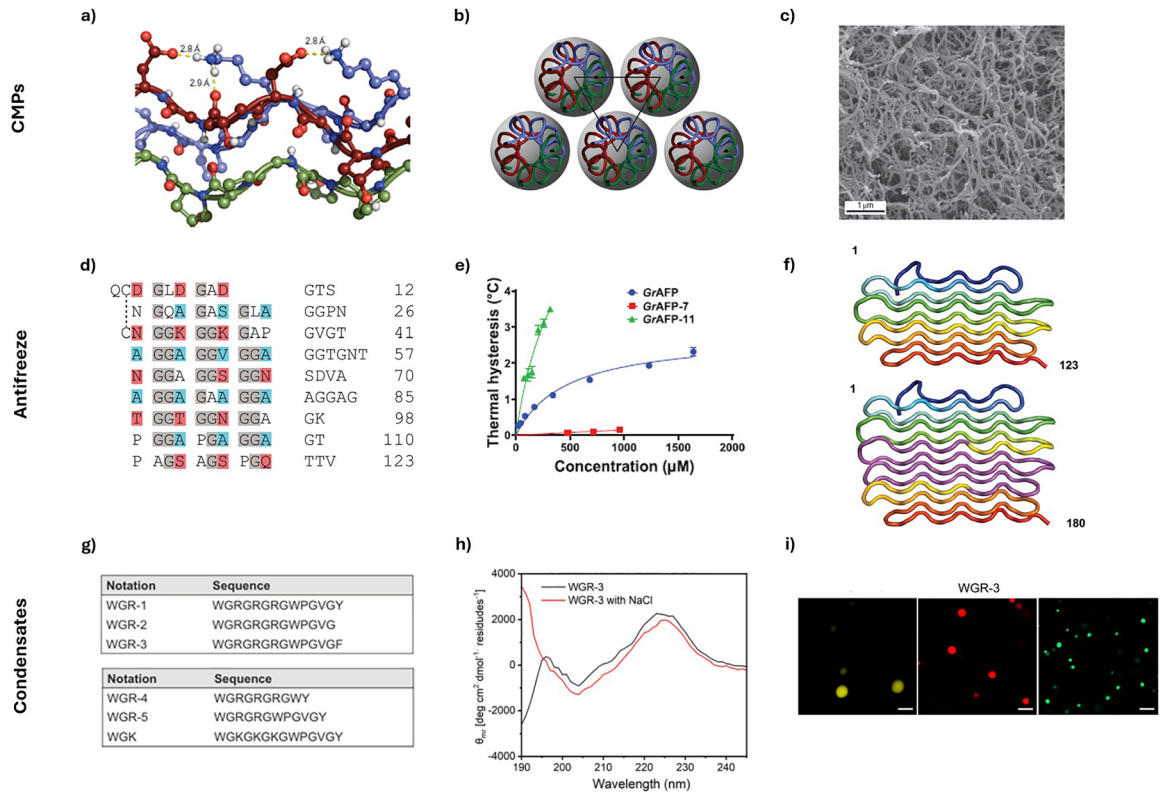

Fig. 4 (a) Collagen mimetic peptide triple helix with Asp-Lys interstrand charged pairs. (b) Possible fibril arrangement of CMP triple helices in quasi-hexagonal format. (c) Fibrous structure of assembled CMP hydrogel through SEM. Fig. 4a–c are reproduced from the ref. 39, with permission from Springer Nature, Copyright 2011. (d) Sequence of native *Gr*AFP used in *Gr*AFP mimetic constructs aligned in structural segments. In gray, the glycine-rich core residues, in blue, the hydrophobic residues, and in red, the hydrophilic residues, are both exposed to the solvent. The dotted line represents a disulfide bond. Reproduced from the ref. 75, with permission from John Wiley & Sons, Copyright 2021. (e) Thermal hysteresis of different *Gr*AFP constructs with deletion/addition of PPII helices. *Gr*AFP-11 with two extra PPII helices in the IBS has the highest TH activity, while *Gr*AFP-7 with two fewer PPII helices has almost no TH activity whatsoever. (f) Structural representation of *Gr*AFP constructs. Above the native *Gr*AFP construct and below the construct with two additional PPII helices. Fig. 4e and f are reproduced from ref. 92 from John Wiley & Sons, Copyright 2023. (g) Table of the designed peptide sequences; (h) CD analysis of WGR-3 peptide at 1 mM with and without NaCl; (i) confocal microscopy images of different fluorophores in the WGR-3 peptide droplets. Fluorescein, Rhodamine B, and GFP from left to right. Scalebar = 10 μm. Fig. 4g–i are reproduced from ref. 120.

hydrophobic interactions were also explored as possible "sticky-ends". Kar *et al.* presented interesting results where a triple helical peptide, with a hydrophobic core flanked by Gly-Pro-Hyp repeats and Phe at N-terminal and Tyr at C-terminal (Phe(Gly-Pro-Hyp)$_3$Gly-Gln-Hyp-Gly-Leu-Hyp-Gly-Leu-Hyp(Gly-Pro-Hyp)$_4$Gly-Tyr), displayed a T_m = 48 °C compared to T_m = 58.8 °C for (Gly-Pro-Hyp)$_{10}$ and could assemble several iterations of these triple helical peptides to higher-order assemblies by "sticky-ended" CH–π interactions between CH_3 groups in the Pro/Hyp rings and the ring of Phe residues.[31]

Another possibility is addressing not only canonical amino acids but also non-canonical amino acid PPII and triple helix propensity. Studies delved into changing the Gly-Pro-Yaa sequence to either Hyp, Pro, or Flp (4(*R*)-fluoro-L-proline). Their use as substitutes is related to increased PPII stability through stereochemical interactions, such as the promotion of *trans* peptide bonds by C^{γ}-exo ring pucker, promoting amino acids. When (Gly-Pro-Yaa)$_7$ contained Flp, a known promotor of said ring pucker, the stability of the triple helical peptide increased (T_m = 45 °C) in comparison with (Gly-Pro-Hyp)$_7$ (T_m = 36 °C), both still displaying a CD spectra on par with the triple helix conformation, as the presence of the *trans* peptide bonds is also key for the extended conformation of the PPII helix.[47,48]

Currently, studies on the improvement of CMP stability have opted for using the strategies mentioned thus far in tandem to obtain even more stable PPII and triple helical configurations out of their CMPs. Hentzen *et al.* showed how an ABC heterotrimeric CMP could form a triple helix by taking advantage of pairwise electrostatic interactions between a canonical and non-canonical amino acid. Using a host peptide (Gly-Pro-Hyp)$_3$-Gly-Xaa-Yaa-(Gly-Pro-Hyp)$_3$, where Xaa was Pro or (4*S*)-aminoproline and Yaa was Pro, Asp, or Glu, it was shown that the triple helix with the highest stability relied on the pairwise interaction between (4*S*)-aminoproline and Asp, with a T_m = 30 °C, while other options were all below 30 °C. The position of these residues helped to maintain the necessary stagger and register with only three salt bridge interactions per helix. The authors suggested this was driven by the interaction with a predicted distance of 2.6 Å between the ammonium and carboxylate groups (N^+–O^-) and a H-bond length of 1.7 Å, geometries that are ideal for a strong salt bridge and would promote the ABC register, since lone (4*S*)-aminoproline is a triple helix destabilizer.[59]

The very recent study of Cole *et al.* was even able to tackle one of the known issues regarding the specific chain register in heterotrimeric CMPs, when they obtained a stable heterotrimeric triple helix CMP from computational studies using an algorithm, SCEPTTr, for predicting melting temperatures of synthetic triple helices, based on previous studies.[37,60] Their H3 ABC heterotrimeric CMP, to their knowledge, has the most controlled register and specificity in triple helix self-assembly in known literature by relying on two in tandem CMP stabilization strategies, pairwise cation–π interactions (Phe-Arg, Arg-Tyr) conjoined with axial (Lys-Asp) and lateral (Arg-Asp) salt bridges, all verified by CD and NMR. When compared to H1 or H2 peptides (T_m = 23.5; 39.5 °C, respectively) also evaluated in this study, H3 had the highest T_m = 44 °C. The relationship between the interactions of the heterotrimers and their T_m was given as the presence of residues that would not interfere with the ABC heterotrimeric assembly (Ser, Pro, and Arg) in the case of H3 *vs.* H1 and H2, as all shared the same axial, lateral, and cation-π residue interactions.[61]

Several of these strategies also include alterations in the used buffers, such as pH, concentration, among others,[62,63] showing how quickly and how diverse the field of CMP stabilization strategies has become, especially when it comes to obtaining more stable mimetic assemblies of PPII strands and PPII-based triple helices.

Antifreeze proteins

PPII helix-adopting proteins are known to play crucial roles in protein–protein interactions, often contributing to the enhanced stability of supramolecular assemblies. While many PPII-adopting proteins are linked to structural purposes, like the collagen triple helix, nature came up with fascinating examples that diverge from this trend.

Antifreeze (glyco)proteins, AF(G)Ps, have been found in several organisms, from arthropods to fish, which live in polar regions in subzero temperatures. In essence, these proteins link themselves to ice crystals. They can lower the overall freezing point, in a process known as thermal hysteresis (TH), creating a separation between the melting and freezing points of ice, creating a temperature gap so ice crystals neither grow nor melt. They can also prevent the appearance of high-dimensional ice crystals, in a process known as ice recrystallization inhibition (IRI) (Fig. 3d).[64] AF(G)Ps can minimize the effects of subzero temperatures on surrounding living tissues, ensuring the survival of the organisms they originate from ref. 65–67.

AF(G)Ps are classified into several types based on their sequence motifs, protein secondary structure, type of activity, and natural origin. While numerous studies and reviews have explored the mechanisms and structures of various AF(G)Ps,[68–71] this review focuses specifically on those that adopt a PPII helical conformation, as these AF(G)Ps exhibit unique properties that set them apart from other antifreeze proteins. These properties are derived from their PPII conformation and PPII-based higher-order assemblies. Table 2 summarizes the major types of PPII-adopting AF(G)Ps, along with their PPII assembly, relevant motifs, their interactions, and overall geometry.

PPII-adopting AFGPs

Antifreeze glycoproteins are divided into 8 types from AFGP1 to AFGP8, with isomers ranging from 33.7 kDa to 2.6 kDa, respectively. These proteins consist of repeats of alanine and threonine (Ala-Ala-Thr*)$_n$, where the threonine residues are glycosylated with a disaccharide moiety (β-D-galactosyl-(1,3)-α-*N*-acetyl-D-galactosamine).[66,77] AFGP7 and AFGP8 comprise low molecular weight AFGPs with n = 5 and n = 4 repetitive motif units, respectively. In some instances, the Ala residues can be changed to Pro, and Thr* to Arg, as found in some polar fishes.[66] Due to the inherent flexibility of the backbone of low molecular weight AFGPs,

Table 2 PPII-adopting antifreeze (glyco)proteins, their PPII assembly, antifreeze motifs, and relevant antifreeze activity. Thr* – Threonine glycosylated with disaccharide β-D-galactosyl-(1,3)-α-*N*-acetyl-D-galactosamine

Type	PPII Assembly	Motifs	PPII residue interaction	Residues hydrophobic face	Residues hydrophilic face	Ref.
Low molecular weight AFGPs	Single PPII helix	Ala-Ala-Thr* Pro substitutions for Ala and Arg substitutions for Thr* can occur	Ala – PPII propensity Thr* – stability through water interactions	Ala-Ala	Thr*	72–74
Snow flea-like AFPs	PPII helix bundles	Gly-X1-X2	Backbone hydrogen bonds (Gly-Gly) Proline-promoting loops Disulfide bridging	Gly, Ala, Val, Ser, Thr, Pro (majority of small residues)	Arg, Lys, Asp, Asn, Ser, Thr, His (majority of hydrophilic and/or charged residues)	20, 75 and 76

especially AFGP8, there was an unresolved thread that its secondary oscillated between extended random coil conformation and PPII helix conformation,[78,79] as seen by solid-state ^{13}C NMR and Fourier-transform infrared spectroscopy (FTIR).[79] However, the use of CD[80] and then two-dimensional infrared spectroscopy (2D-IR)[72] shed light on the clear PPII helix structure of these AFGPs near and above their working temperature.

AFGP8 relies on its motif units and their functional geometry, of which its PPII helix is both essential and a consequence of the residue–residue interaction within its repetitive motif units.[74] AFGP8 has two distinct faces, a hydrophilic face and a hydrophobic face, therefore having an amphipathic conformation (Fig. 5). Such a conformation is achieved through the spatial segregation of the CH_3 groups of the AFGP backbone, present in the Ala and Thr side chains, to one side, and the disaccharide moieties to the opposing one. Some authors even proposed that the disaccharide moieties could stabilize the PPII conformation through hydrogen bonds between the amine protons of the disaccharide and the backbone carbonyl groups in Thr residues.[81] We also reason that the presence of Ala residues with their single CH_3 side chain is, in itself, a promoter of the PPII conformation this protein adopts, as evidenced in other studies.[14,82]

Knowledge regarding the ice-binding mechanism of AFGPs to ice has changed throughout the years, even if always linked intimately to its flexible PPII structure. Initially, it was thought to be irreversible,[73] but Zhang *et al.* revealed that AFGP8 can bind both reversibly and irreversibly. The reversible binding occurs through van der Waals interactions between CH_3 groups in the backbone and the disaccharides and the ice lattice, as these groups match the ice inner grooves, preventing ice recrystallization. Irreversible binding involves direct hydrogen bonding between saccharide OH-groups and ice. This dual binding mechanism highlights AFGPs' role as water molecule modulators, particularly at supercooled temperatures due to their higher hydration shell, modulating the water–water hydrogen bonds needed for the ice crystal organization.[18,67,83]

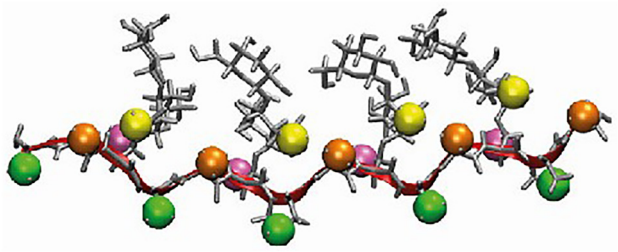

Fig. 5 3D structure of AFGP8. The PPII backbone is shown in red and the disaccharide moieties in gray, with the CH_3 groups of the backbone residues in pink, green and orange, and the CH_3 groups of the disaccharide moieties in yellow. It is proposed that the backbone constitutes the IBS through the CH_3 groups and the disaccharide moieties of the hydrophilic face. Reproduced from the ref. 73, with permission from American Chemical Society, Copyright 2018.

PPII-adopting AFPs

Among the many types of antifreeze proteins, snow flea-like antifreeze proteins (sfAFPs) comprise a subgroup of hyperactive insect AFPs that feature unique supramolecular PPII assemblies. These sfAFPs possess hyperactive antifreeze activity that is intrinsically related to the functional geometry provided by their PPII-rich structure.[84,85]

Extensive analysis of sfAFPs and their PPII-rich structure, notably for the snow flea AFP from *Hypogastrura harveyi*, has been made. From attempts at CD spectra deconvolution to distinguish random coil signatures from PPII helices, to building molecular dynamics models,[20] and using finer techniques such as X-ray diffraction[86] and NMR,[76] it was possible to obtain crucial data on the intricate structure, motifs, and interactions that govern both the structural assembly and ice-binding mechanism of these hyperactive proteins. The common characteristic of sfAFPs was their adoption of a PPII bundle conformation (Fig. 3e and f), where some can have 9 (PDB:7JJV[75]), 12, or even 13-bundled PPII helices.[19,87]

Using the snow flea AFP *Hh*sfAFP (PDB:2PNE[86]) as an example (Fig. 3e and f), this sfAFP is 81 residues in length, with 6 different segments composed of three to five Gly-X1-X2 repetitive units per segment, which is similar to the Gly-Xaa-Yaa repeats of collagen. Each segment organizes into a separate PPII helix conformation, which later assembles into a six-bundled PPII helical supramolecular structure. In each PPII strand segment, there are proline-containing loops that bring cysteine residues close together to form disulfide bridges, as is the case for Cys1–Cys29 and Cys13–Cys44. Residues at X1 are usually Gly, small side-chain residues, or Cys. Residues at X2 vary depending on the segment, being either charged/hydrophilic or hydrophobic in nature.

The stability of PPII helical bundles is influenced by disulfide bonds,[88] Gly residue interactions in the central region,

and proline-containing loops (Fig. 3e). A densely packed Gly-rich core is facilitated by two types of interhelical hydrogen bonds: non-canonical C^{α}–H···O=C and canonical NH···O=C bonds. These non-canonical hydrogen bonds provide internal stiffness, resistance to bending, and contribute to overall hydrogen bond cooperativity, stabilizing PPII helical bundles.[89,90] In sfAFPs, prolines act as spacers for PPII helix segments, and they play a crucial role in stabilizing flanking loops. The proline's ring pucker conformation creates a kink,[6,91] as in collagen, which is essential for approximating each segment/PPII helix and enabling the stabilizing core interhelical glycine interactions.

These interactions culminate in a functional geometry of sfAFPs that allows the appearance of two distinct faces, a hydrophobic face that binds to ice, and a hydrophilic face that modulates water–water interactions, on opposite sides of the protein, with a Gly-rich center (Fig. 3f).[20]

While AFGPs are thought to bind directly to the ice lattice grooves through their side chains backbone residues, sfAFPs rely on an ordered structure called water clathrate. Water clathrate is an organized water structure around hydrophobic groups of the hydrophobic face (ice-binding sites), that mimics the quasi-liquid layer adjacent to an ice lattice, facilitating the adsorption of sfAFP to ice.[70] What makes the sfAFPs interesting in comparison to other AF(G)Ps is that their hyperactivity can be related to the PPII-bundled IBS. For *Gr*AFP (PDB:7JJV[75]), the 4 PPII helices that make up its IBS allow for a TH of 2.5 °C, as the PPII-based hydrophobic face is flatter and has increased surface area than in other AFPs, thereby increasing the antifreeze activity.[92,93]

PPII-adopting AF(G)Ps mimetics

Research on mimetic counterparts of AF(G)Ps that adopt a PPII-helical conformation is still limited, with studies focusing on other AFP types.[94–101] The synthesis of glycosylated residues poses a significant challenge for developing synthetic PPII-adopting AFGPs, and PPII-adopting AFPs are less common and have only recently gained attention as research models, delaying studies of their mimetic counterparts. Despite this, very recent studies have shown how synthetic PPII-adopting AF(G)Ps can be used as sustainable cryoprotectant alternatives for biomedical applications, with an increased antifreeze activity due to their PPII conformation.[65,92,102]

PPII-adopting synthetic AFGPs

The first synthetic AFGPs (sAFGPs) appeared in the early 2000s,[80] and later solid-phase synthesis and amino acid *N*-carboxyanhydride (NCA) polymerization strategies were developed to tackle the known bottleneck regarding the use of glycosylated residues, crucial for the PPII structure of native AFGPs.[65,68,102]

Recent studies published in mid-2024 by the lab of Jessica R Kramer delved into obtaining sAFGPs based on PPII-adopting AFGPs.[65,102] They synthesized several sAFGPs with changes to the saccharide moiety in the Thr residues, length of the sAFGP, and by playing with the ratio of Ala: Thr in the repetitive motif units, which in native AFGPs is 2 : 1(Ala : Thr*) or Ala-Ala-Thr*. A CD analysis and antifreeze activity assays were conducted to verify the secondary structure of these mimetics and their IRI activity. In terms of secondary structure, they found that changing the ratio of Ala: Thr did not change the PPII conformation of the sAFGP, but changing the saccharide moiety did, cementing the idea that the β-D-galactosyl-(1,3)-α-*N*-acetyl-D-galactosamine is indeed a driver of the PPII helix in these sAFGPs, as seen in native AFGPs, as all other glycans led to a decrease in the CD spectra maximum near 217 nm. The length of the sAFGP also played a role, with a more robust PPII structure as the number of residues increased to 42 repetitive motif units or 128 residues. After this, the sAFGPs adopted a mix of PPII and α-helical conformation.

When testing the IRI activity, they understood that while the ratio changes of Ala : Thr did not change the PPII structure, it led to a decrease in IRI when not in the native Ala : Thr (2 : 1) ratio. However, when running computational studies, they saw that these polymerized sAFGPs did not always have the Ala-Ala-Thr triplet in this format and still had relevant IRI, meaning that the ratios of Ala : Thr are relevant, but the position they take in the triplet is not. It was also shown that in terms of IRI, using the 57 residue PPII-adopting mimetic in comparison with the 170-residue mimetic that adopts a mix of PPII/α-helix conformations, did not alter the IRI activity significantly, with both exceeding the less than 10% mean grain size of ice crystals. While the authors claimed that these results meant that the PPII helix is not a necessity for sAFGP IRI activity, we argue that this is not the case, as there is still a presence of PPII conformation in the CD spectra.[65]

While this study implies that the antifreeze activity of AFGPs is not related to their PPII structure but to a bulk effect of the residues within the backbone, several of the presented examples show that the PPII conformers have relevant IRI, with a smaller backbone, which is arguably more efficient in mimetic formulations.

PPII-adopting synthetic AFPs

At the time of writing this review article, there are still few papers that investigate PPII-adopting synthetic AFPs, or as we name them, synthetic snow flea-like AFPs (SsAFPs). Some studies have delved into the antifreeze activities of several canonical amino acids,[83,103] but the study of Scholl and Davies shows how the presence of the PPII helix is essential for the hyperactive antifreeze activity of SsAFPs.[92] In this study, three hypotheses were tested on synthetic variants (Fig. 4d–f) of *Gr*AFP, mutating residues in the IBS, lengthening the PPII helices in the IBS, and adding/deleting PPII helices in the IBS. A mutation Ala for Lys (Ala46Lys) in the IBS showed a decrease in the TH activity. Doing a double mutation (Ala46Lys/Ala82Lys) almost eliminated the TH activity completely. This was said to be related to a decrease in the flatness of the IBS or a lack of hydrophobic residues to interact with the ice lattice through water clathrate, as previously shown in natural sfAFPs. Lengthening the PPII helices also showed a 5-fold decrease in TH, which was related to the occurrence of a less rigid higher-order PPII bundle structure. The rigidity has been shown to be essential to maintain flatness

and the correct IBS conformation. However, when testing the addition/deletion of helices in the IBS (Fig. 4e and f), it was shown that the addition of an extra helix led to an increase in surface area and flatness, contributing to more interactions with the ice lattice and a 3-fold increase in TH. These results cement the relevance of the PPII helix in the hyperactivity of these SsAFPs. Another study conducted in 2009 even showed how there could be similarities in the antifreeze activity of these sfAFPs and collagen, since they share a similar repetitive motif unit (Gly-X1-X2 *vs.* Gly-Xaa-Yaa). While the mechanism was not clearly described at the time, one of the gelatin hydrolysate-containing fractions showed pertinent IRI activity in concentrated sucrose solutions, which could be possibly linked to a matching pattern with the ice lattice for better binding.[104]

The findings presented in this chapter show the critical role of the PPII helix and PPII-based supramolecular assemblies in the antifreeze activity of mimetic AF(G)Ps, by addressing the pioneering studies conducted on these synthetic variants and demonstrating their efficacy as structural models for proteins that exploit the functional geometry of the PPII helix. The research done thus far reveals that the PPII helix significantly enhances antifreeze activity, offering valuable insights into the design and development of more effective antifreeze mimetics.

Natural condensates

Condensates are membraneless organelles formed through liquid–liquid phase separation (LLPS) that play a crucial role in cellular dynamics by selectively concentrating proteins, ribonucleic acids, and other biomolecules. These dynamic, self-organizing structures emerge spontaneously within cells, acting as distinct microenvironments that facilitate various biochemical processes and regulatory functions. Unlike traditional membrane-bound organelles, condensates exhibit fluid-like properties and can rapidly assemble, disassemble, or reshape in response to cellular signals and alterations in environmental factors.[3,105,106] While this adaptability is essential for normal cellular function, it also introduces a potential vulnerability. Under certain circumstances, these phase-separated compartments can promote pathological conditions, including cancer progression and viral infections. Indeed, dysregulation of LLPS-driven organelle formation has been implicated in numerous disease states, underscoring the critical balance required in maintaining proper phase separation dynamics for overall cellular health and function.[105]

A crucial aspect of condensate behavior is its specificity and selective interactions. These interactions include, but are not limited to, electrostatic forces, hydrophobic effects, π–π stacking, cation–π interactions, and hydrogen bonding. The balance and specificity of these interactions determine the miscibility or immiscibility of different condensate types.[107]

As mentioned before, the challenging identification of PPII structures often leads to their oversight, particularly in regions classified as intrinsically disordered, being catalogued as random coil or just as ''a helical conformation''. This difficulty in detection may explain why a direct relationship between the presence of populations with PPII structures and the formation of biomolecular condensates has not yet been firmly established.[7,8]

However, recent structural analyses have revealed significant populations of PPII helices in intrinsically disordered proteins that exhibit a propensity for phase separation.[108,109] This discovery has led to the hypothesis that these conformational elements may play a crucial role in modulating the formation and properties of biomolecular condensates.

In a recent study,[108] the authors propose the hypothesis that glycine-rich segments, frequently found in proteins involved in biomolecular condensates, can adopt PPII helical conformations. This research focuses on the third Arg-Gly-Gly (RGG) region of the FUS protein. The RGG motifs integrate the RNA-binding domain of FUS, which plays a crucial role in RNA-dependent phase separation by enhancing RNA binding affinity, participating in multivalent interactions, and modulating condensate properties.[110–112]

Using NMR and molecular dynamics simulations, the researchers found evidence of significant PPII helical populations in a polypeptide corresponding to the third (Arg-Gly-Gly) region of FUS. To further support their findings, they designed a model peptide, based on the third RGG region of FUS, called RGGmini (Ac-Arg-Gly-Gly-Tyr-Gly-Gly-Arg-Gly-Gly-Trp-Gly-Gly-Arg-Gly-Gly-Tyr-NH_2). CD of this peptide suggested a PPII population of about 40%. Additional evidence for PPII helical structure came from FTIR and NMR data on RGGmini. Key NMR observations included Nuclear Overhauser Effect (NOE) correlations between side chain hydrogens of amino acids spaced three residues apart, consistent with a PPII helical structure. Based on these findings, the authors propose the hypothesis that glycine-rich segments of disordered proteins may form PPII helices, which could help stabilize biomolecular condensates. The same hypothesis could also be extended to the proline-rich sequence involved in LLPS processes, with some papers revealing the presence of populations with a PPII structure.[113]

The well-known C-terminal domain (CTD) of RNA polymerase II (RNA pol II) is composed of 52 tandem repeats of the heptapeptide consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Fig. 3g), forming a unique structural motif that creates biomolecular condensates (Fig. 3h).[114] Despite being described as a random coil, its CD spectrum (Fig. 3i) shows characteristic signs of PPII structures, suggesting their presence. This hypothesis is corroborated by Janke *et al.*,[109] in which the authors made a prediction using NMR data of the CTD region heptads 27–52 of human RNA polymerase.[109] Although predominantly a statistical coil (62%), this region contains a significant proportion of PPII helix (25%). To make this prediction, C^{α}, C^{β}, ^{15}N, ^{1}HN, and CO shifts were used as inputs for the δ2D algorithm to predict secondary structure population.[115] This prediction method will be discussed in greater detail later in the chapter ''Characterization of PPII structures''.

The CTD of RNA pol II plays a crucial role in the formation of biomolecular condensates, such as superenhancer and nuclear paraspeckle condensates, which are essential for the processes of transcription and splicing. The unique structure of the CTD facilitates interactions with transcription factors containing

WW or SH3 domains. These interactions are key to recruiting the Mediator complex, which in turn helps organize phase-separated superenhancer condensates. This organization significantly enhances transcription rates.[116,117]

Mimetic condensates

Significant efforts have been made to identify and understand biomolecular condensate systems due to their potential applications. To this end, numerous peptide mimetic systems have been reported, aiming to study how to engineer the properties of these systems.[118–120]

The studies conducted do not mention the formation of PPII structures by mimetic peptides. However, given that the focus is primarily on the nature of the amino acids comprising the peptides and their respective interactions, it is possible that this type of structure often goes unnoticed and is simply categorized as ''disordered regions''.

A potential example of this hypothesis is the study of Leshem *et al.*, in which the authors aimed to establish sequence-structure–function relationships for LLPS promoter peptides in order to understand the molecular mechanisms and driving forces behind the formation of biomolecular condensates.[120] To this end, they designed and studied a library of six minimalistic peptide sequences (Fig. 4g) to observe which ones promoted LLPS and formed biomolecular condensates (Fig. 4i).

The secondary structure of the peptides was analyzed using CD (spectrum of the WGR-3 peptide, as an example, represented in Fig. 4h) and described as ''consistent with a disordered peptide in random coil conformation''. However, by examining the respective spectra (provided in the article's supplementary information) for some of the peptides, we can observe shifts in the minimum and maximum spectrum for the random coil conformation, with values consistent with PPII structures. An even stronger indicator is the presence of a positive peak in the spectra, with a maximum between 220 nm and 230 nm, which is characteristic of PPII conformation.[91,121] Based on these factors, it is plausible to hypothesize that unidentified PPII populations may exist in some of these mimetic peptides.

Another interesting theory is put forward by Edun *et al.*[122] This theory proposes the reverse scenario, in which it is the LLPS phenomenon that induces protein folding in IDPs. Specifically, this case is reported for PPI and PPII structures. For this study, the model peptide PR20 (composed of 20 proline-arginine repeats) and polyethylene glycol, as a molecular crowder to induce LLPS, were used. Using FTIR and 2D-IR to monitor structural changes, researchers recorded that LLPS formation was accompanied by changes in amide-I spectra consistent with folding into helical polyproline structures. More specifically, 2D-IR spectroscopy revealed new peak pairs indicating the formation of polyproline I (PPI) and PPII helices. Finally, by comparing the experimental findings with 2DIR spectroscopic simulations of both these structures, the results were corroborated, and it was estimated that the ratio of PPI to PPII in droplets was 1.09 to 1.00.[122]

Although this is the first reported case of LLPS inducing folding, these results are particularly intriguing in the context of this review. They prompt us to consider whether the formation of condensates might favor specific types of polyproline structures (in this case, not exclusively PPII). This perspective suggests a potential mutual relationship: not only might PPII populations contribute to the formation of condensates (as previously proposed), but the condensates themselves could also promote the adoption of PPII structures. This reciprocal interaction could lead to a mutual stabilization and induction effect between PPII structures and condensate formation.

Characterization of PPII structures

This chapter explores the various techniques employed by researchers to detect, analyze, and understand PPII helices (Table 3). Experimental and computational methods are explored, highlighting their strengths and limitations. We will also describe the main challenges in identifying PPII helices (compared to α-helices and β-sheets), particularly focusing on spectroscopic methods.[91,123]

Bioinformatics tools

Nowadays, there are databases and tools that specialize in making PPII predictions, such as PPIIPRED,[6] which utilizes a bidirectional recurrent neural network to process protein sequence data, multiple sequence alignments, disorder predictions, and information on residue charges.

When a protein's 3D structure is available, researchers can also utilize dedicated software like PolyprOnline,[11] which focuses on PPII helices. This tool offers a comprehensive database and multiple assignment methods, enabling a comparative analysis across proteins. Therefore, this software provides quantitative data on PPII content and standardizes PPII identification.

Although bioinformatics tools provide rapid and practical means for analysis, their results must be interpreted with caution. AI-based predictors like AlphaFold are optimized for folded structures and may not accurately represent intrinsically disordered or flexible regions that exist as dynamic ensembles. PPII-specific tools such as PPIIPRED and PolyprOnline, while designed for PPII prediction and analysis, rely on training data and structural databases of static conformations, which can limit their ability to capture conformational variability. Consequently, confidence scores may underestimate uncertainty in these flexible regions, and overreliance on single static models risks oversimplifying the inherent structural heterogeneity.[6,124–126]

Therefore, careful interpretation combined with experimental corroboration becomes essential, particularly in design workflows for CMPs and condensate-forming peptides, where accurate structural representation directly impacts functional outcomes.

The experimental identification of PPII structures is accomplished through various spectroscopic techniques. The next section will detail the primary methods used to characterize PPII conformations.

Table 3 Comprehensive characterization toolbox for Polyproline II (PPII) helices and assemblies

Method/techniques	Main readout/signature	Strengths	Limitations	Recommended for	Ref.
Computational (AlphaFold, PPIIPRED, PolyprOnline)	Predicted secondary structure; PPII content quantification; comparative protein databases	Rapid, practical, and easily accessible; utilizes large datasets; standardized assignments	Relies on static conformations; confidence scores limited for flexible regions/May misrepresent disordered/flexible regions; needs experimental validation	Preliminary screening; comparative analysis; folded proteins	6,11, 124–126
Circular dichroism (CD)	Quick, accessible; monitors conformational/thermal changes; suitable for secondary structure overview	Quick, accessible; monitors conformational/ thermal changes; suitable for secondary structure overview	Ambiguous with higher-order (triple helix/fiber) structures; lacks residue-specific resolution; averages whole protein/population	Initial structure assessment; folding/ Unfolding studies	7,39,121, 127–132
Vibrational circular dichroism (VCD) and Raman optical activity (ROA)	VCD: Shift in the amide I carbonyl stretch frequency, at $\sim$1623 cm^{-1}, about 23 cm^{-1} lower than random coil; ROA: strong positive extended amide III band at $\sim$1318 cm^{-1}	High 3D structure sensitivity; distinguishes PPII from random coil; quantitative in complex mixtures	Less commonly available; requires specialized equipment/expertise; glycan/ solvent may interfere; signal/ noise poor for small or dilute samples	Complex proteins; mixtures; detailed conformational analysis	132–142
Nuclear magnetic resonance (NMR)	^{15}N chemical shifts: $\sim$1.1 ppm at 10 °C; coupling constants: $^3J_{HN^{\alpha}}$ $\sim$6.5 Hz and $^1J_{C^{\alpha}H^{\alpha}}$ $\sim$142.6 Hz; NOE patterns: strong sequential $d\alpha N(i,i+1)$ and typically lack $dNN(i,i+1)$ and $d^{\alpha}N(i,i+4)$ (contrary to α-helices)	High-resolution/residue specific; identifies flexible/ disordered regions; multiple PPII indicators	Requires advanced experiments/expertise for flexibility and dynamics; chemical shift differences can complicate assignment; challenging for short/ heterogeneous peptides	Intrinsically disordered proteins/peptides (LLPS-driven condensates); quantitative conformational analysis; distinguishing PPII from random coil structures	1,4,14,76, 115,132, 143–149
Scattering/crystallography/EM	3D atomic structures from X-ray/neutron scattering, electron microscopy (EM); fiber diffraction	Precise atomic/molecular structure; reveals supramolecular organization	Typically requires crystalline/ fiber samples; less suitable for dynamic/disordered peptides	Well-ordered proteins/ fibers; collagen/CMP fibrils; stable complexes	39, 150–153

Circular dichroism (CD)

CD stands out as the most straightforward and widely accessible technique for identifying PPII helices. CD is particularly valuable for providing a quick overview of the secondary structure in a given sample, being sensitive to conformational changes, such as temperature-induced structural transitions.

This method relies on the characteristic spectral features of PPII structures, which include a positive band (distinct from random coil) between 220 and 230 nm and a negative band between 195 and 210 nm.[121,127,128] It is important to note that slight variations in the positions of the peaks can occur depending on the composition of the peptides and the experimental conditions.

One of the main challenges in characterizing PPII helices by CD arises from the possible presence of higher-order structures, such as PPII triple helices or even fibers composed of triple helices (as occurs in collagen). These structural elements, depending on their degree of prevalence, also cause deviation in the CD spectrum relative to the secondary PPII helix structure.

Based on Lopes *et al.*[121] research, the PPII helix (from poly-L-proline) exhibits a maximum positive signal at 225 nm and a minimum typically between 195–205 nm. In the case of PPII triple helices (diluted collagen samples), the spectral profile is very similar, but both the maximum and minimum are slightly shifted to lower wavelengths, appearing at 220–222 nm and 195–200 nm, respectively.[121] However, given that these differences in wavelength are not significant, for CMPs, and considering that spectral variations may already result from the inherent sequence differences, it is not feasible to differentiate between isolated PPII helices and CMP-based collagen triple helices using this method.

Temperature-dependent CD studies can also be helpful, as PPII structures tend to show more significant spectral changes with temperature than random coils. Specifically, there is a notable decrease in the intensity of the positive band (maximum) and an increase in the intensity of the negative band (minimum). This behavior contrasts with that of random coils, which display more subtle alterations in their CD spectra across a range of temperatures.[7,129]

CD is particularly useful for distinguishing whether CMPs assemble as isolated PPII helices or as triple helices. This is typically evaluated by monitoring the CD signal at the maximum wavelength of the CMP (usually near 225 nm) while gradually increasing the temperature. For CMPs forming a triple helix, the CD profile exhibits a pronounced and cooperative thermal melting transition, which corresponds to the disruption of stabilizing interchain hydrogen bonds within the triple-helical structure. This transition is marked by a distinct melting temperature (T_m), which serves as an indicator of the thermal stability of the triple helix and can vary depending on factors such as peptide composition and environmental conditions. In contrast, if the peptide adopts only a PPII conformation, the CD signal changes gradually with temperature and no cooperative melting transition is observed.[39,130,131]

Despite its utility, CD has limitations. It provides average information on the content of secondary structures in proteins, without distinguishing between different regions or offering detailed information on specific residues. Therefore, in order to overcome some of these shortcomings and make a more complete analysis, researchers can employ complementary spectroscopic techniques.[132]

Vibrational circular dichroism (VCD) and Raman optical activity (ROA)

Complementing conventional CD, vibrational circular dichroism (VCD) and Raman optical activity (ROA) offer enhanced detection of PPII structures.[132]VCD, utilizing infrared light for vibrational transitions, provides detailed structural information in the amide I and II regions[133134]. It distinguishes PPII from random coil conformations based on a characteristic shift in the amide I carbonyl stretch frequency, with PPII structures exhibiting this vibration at $\sim$1623 cm^{-1}, about 23 cm^{-1} lower than random coil.[132,135,136]

ROA, measuring the difference in Raman scattering intensity using circularly polarized light, is highly sensitive to three-dimensional molecular structure. It excels in detecting PPII helices due to their distinctive spectral features, particularly a strong positive extended amide III band at $\sim$1318 cm^{-1}.[132,137–141] This allows ROA to differentiate PPII from random coil structures and identify PPII content in proteins with complex structural compositions.[142]

Nuclear magnetic resonance (NMR)

NMR's ability to provide high-resolution data on flexible and dynamic protein regions makes it especially valuable for characterizing PPII conformations, which are frequently found in intrinsically disordered proteins. However, the absence of intramolecular hydrogen bonds in PPII helices limits the applicability of standard ^{1}H-NMR spectroscopy in distinguishing them from disordered conformations. To overcome this limitation, more advanced NMR techniques are required for accurate identification and analysis of PPII structures.

It is also important to note that although NMR identification of PPII structures often relies on parameters specific to certain residues, such as proline, glycine, and alanine, this review takes a broader approach.[14,76,143] To encompass a wide range of PPII-forming sequences, we will focus on NMR parameters that are universally indicative of PPII structures, rather than those tied to particular amino acids. This strategy allows for a comprehensive analysis of PPII helices across diverse peptide and protein compositions, utilizing general structural indicators such as backbone chemical shifts, coupling constants, and relaxation measurements.

NMR chemical shifts

One of the most straightforward NMR approaches for identifying PPII structures is through the analysis of chemical shifts. The chemical shifts of backbone atoms (^{1}H, ^{13}C, and ^{15}N) are highly sensitive to local conformation. Therefore, with the chemical shifts of these nuclei, it is then possible to use methods such as δ2D to estimate secondary structure populations, including PPII helices.[115] This method utilizes a large database of 1772 proteins, with known chemical shifts and structures. This database is divided into libraries for different types of secondary structure, including PPII fragments longer than three residues (2419 residues).[115,144]

In 2002, Sik Lok Lam and Victor L. Hsu[145] provided a detailed NMR characterization of a model PPII helix and demonstrated how various NMR parameters, especially ^{15}N chemical shifts, can be used to identify and quantify these structures in peptides and proteins. The researchers established that ^{15}N chemical shifts exhibit notable changes from random coil values in isolated PPII helices with an average $\Delta\delta^{15}$N of 1.1 ppm at 10 °C. This can be explained due to the high sensitivity of the ^{15}N chemical shifts to local conformational changes and hydrogen bonds. Systematic chemical deviations were also recorded for the C^α, C^β, and C' nuclei (0.3 ppm, 0.2 ppm, and 0.1 ppm, respectively). However, since these deviation values are small, these parameters are consequently less effective than the ^{15}N shifts in distinguishing PPII helices from random coil structures. The authors demonstrated a robust correlation between ^{15}N chemical shift deviations and CD spectroscopy data, specifically the molar ellipticities at 220 nm at different temperatures.[145]

NMR coupling constants (*J*-couplings)

This data was used in conjunction with the coupling constants $^3J\mathrm{HN}^\alpha$ and $^1J\mathrm{C}^\alpha\mathrm{H}^\alpha$ to study protein secondary structure. $^3J\mathrm{HN}^\alpha$ is sensitive to the phi (φ) torsion angle, while $^1J\mathrm{C}^\alpha\mathrm{H}^\alpha$ reflects local geometry around the C^α atom. For the PPII peptide model used at 10 °C, they measured average values of $^3J\mathrm{HN}^\alpha$ (6.5 ± 0.3 Hz) and $^1J\mathrm{C}^\alpha\mathrm{H}^\alpha$ (142.6 ± 0.5 Hz), consistent with PPII helical structures.[145] The combination of these constants with ^{15}N chemical shift data provides an accurate method for identifying PPII helices and distinguishing them from other secondary structures with greater accuracy and confidence.[4,145–147]

NMR nuclear Overhauser effect (NOE) signals

The presence or absence of NOE signal patterns can also play a crucial role in identifying PPII helices. Strong sequential $d_{\alpha N}(i,i+1)$ NOEs are characteristic of PPII helices, reflecting the extended nature of the backbone where adjacent residues are positioned favorably for through-space interactions between the alpha proton of one residue and the amide proton of the subsequent residue. Medium-range $d_{\alpha N}(i,i+2)$ and $d_{\alpha N}(i,i+3)$ NOEs are often also observed, indicating the helical arrangement of residues and providing evidence for the characteristic left-handed helical twist of PPII structures. Notably, PPII helices typically lack $d_{NN}(i,i+1)$ NOEs and longer-range $d_{\alpha N}(i,i+4)$ NOEs, which serve to distinguish them from α-helices.[1,4,132,148,149]

In proline-rich sequences, $d_{\alpha\delta}(i,i+1)$ NOEs between the alpha proton of one residue and the delta protons of the subsequent proline can provide additional evidence for PPII conformation.[149]

In addition to characteristic NOE patterns, the quantitative analysis of NOE intensities provides valuable information about internuclear distances in PPII helices. However, in certain cases, the dynamic nature of these structures complicates this analysis.[4,149]

As described, several NMR parameters are available for detecting PPII helices. While each parameter has its own limitations when used individually, their combined application provides a powerful and efficient approach for distinguishing and characterizing these challenging structures.

Additional techniques

In addition to the described methods, X-ray scattering, neutron scattering, and electron microscopy (EM) are powerful techniques for characterizing higher-order fibrillar assemblies based on PPII structures.

X-ray and neutron scattering provide detailed insights into molecular packing, periodicities, and supramolecular motifs, making them ideal for deciphering the hierarchical organization of collagen-mimetic peptides and related assemblies. However, X-ray scattering requires well-ordered samples and may provide limited structural information for highly flexible regions, while neutron scattering demands substantial sample quantities and access to specialized facilities, which restricts its widespread use.[150,151]

EM offers direct visualization of fibril morphology and supramolecular architecture in near-native states, complementing scattering techniques by capturing heterogeneous and dynamic assemblies across multiple length scales. Although EM generally yields lower atomic resolution compared to crystallography and requires careful sample preparation to minimize artifacts, it excels in studying structurally diverse and non-crystalline fibrillar systems.[39,152,153]

Conclusions and future perspectives

The PPII helix has emerged as a crucial structural motif in supramolecular assemblies, playing significant roles in both natural and mimetic systems. This review has explored the fundamental principles of PPII helices through their occurrence in natural assemblies, exemplified by collagen, AF(G)Ps, and LLPS-driven condensates, as well as the development of mimetics based on these examples. We have also delved into the advanced characterization methods that have been fundamental in detecting and analyzing PPII helices, such as CD and NMR, among others, and critical optimization strategies for the detection of sole PPII helices or their presence in supramolecular assemblies.

The significant prevalence of the PPII structure in globular proteins, its distinct structural dihedral angles, and polyvalence have been highlighted, noting that changes in these angles could result in any of the other common protein secondary structures. The inherent flexibility of this structure, while posing challenges for characterization efforts, is also related to the functionality of proteins harboring this secondary structure, as it facilitates interactions with other proteins or binding sites and even between themselves.

PPII helices play a vital role in various biological processes and structures. In antifreeze glycoproteins it provides the essential geometry for ice binding, ensuring their normal function. PPII helices also serve as building blocks for supramolecular assemblies, exemplified by hyperactive antifreeze proteins that form PPII bundles, enhancing their ice-binding capabilities. Perhaps the most well-known PPII-based supramolecular structure is the collagen triple helix, where three PPII helices intertwine to form a right-handed supercoil. Furthermore, PPII helices contribute to the formation of natural condensates through LLPS, highlighting their importance in cellular organization. These diverse roles underscore the significance of the PPII helix across several fields, such as biomimetics, with the advent of collagen-mimetic peptides (CMPs), synthetic AF(G)Ps, and mimetic condensates.

CMPs have been evolving to better impart PPII helices mimics for improved folding into a triple helix, with the development of several strategies. Synthetic AF(G)Ps, while still in their infancy, have been shown to be more effective than other structural conformational mimics due to their PPII helices. While studies in mimetic condensates mentioned thus far do not always explicitly imply the presence of a PPII structure in their peptide mimics, a deeper analysis of their data could allow for a distinction between ''disordered regions'' and clear PPII structures. This review has discussed propositions for better analyzing spectra of peptides or proteins that might adopt PPII conformations, along with the current adopted method of protein structure software co-analysis as an alternative to the sole use of analytical methods.

PPII helices can now be seen as a crucial secondary structure that holds immense potential for several fields, including nanotechnology and materials science, with applications in biomimetic materials, drug delivery systems, nanoelectronics, and sensors. While some of these fields are still quite recent, progress has been made in understanding and utilizing PPII helix-based assemblies, even if the predictability of assembly formation, stability, scalability for practical applications, and precise control over assembly properties are areas that require further investigation. Addressing these challenges will be crucial for realizing the full potential of PPII-based supramolecular assemblies in various applications.

As the understanding of these structures deepens and new characterization and synthesis methods are developed, we can expect to see increasingly sophisticated and functional materials based on PPII helices. Continued research and innovation in this area will undoubtedly lead to exciting discoveries and advancements, marking the versatility and unique properties of PPII helices, which make them a promising subject for interdisciplinary research.

Abbreviations

2D-IR	Two-dimensional infrared spectroscopy
AFGP(s)	Antifreeze glycoprotein(s)
AFM	Atomic force microscopy
AFP(s)	Antifreeze protein(s)
CD	Circular dichroism
CMP(s)	Collagen mimetic peptide(s)
DLS	Dynamic light scattering
DSC	Differential scanning calorimetry
EM	Electron microscopy
FTIR	Fourier-transform infrared spectroscopy
FUS	Fused in sarcoma (Protein)
GFP	Green fluorescent protein
HPLC	High-performance liquid chromatography
IBS	Ice-binding site

IDPs	Intrinsically disordered proteins
IRI	Ice recrystallization inhibition
LLPS	Liquid–liquid phase separation
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
PPII	Polyproline II
ROA	Raman optical activity
SAXS	Small angle X-ray scattering
SEM	Scanning electron microscopy
SFafp(s)	Snow flea-like antifreeze protein(s)
SPPS	Solid-phase peptide synthesis
ssAFP(s)	Synthetic snow flea-like AFP(s)
sAFGP	Synthetic AFGP(s)
TEM	Transmission electron microscopy
TH	Thermal hysteresis
T_m	Melting temperature
VCD	Vibrational circular dichroism
WAXS	Wide angle X-ray scattering

Author contributions

Francisco Seco: conceptualization (equal); writing – original draft (equal); writing – review and editing (equal). Sandro Amador: conceptualization (equal); writing – original draft (equal); writing – review and editing (equal). Felipe Conzuelo: funding acquisition; writing – review and editing. Ana C. Baptista: supervision; funding acquisition; writing – review and editing. Leonor Morgado: funding acquisition; writing – review and editing. Ana S. Pina: conceptualization (lead); supervision; funding acquisition; writing – review and editing.

Conflicts of interest

The authors state there are no conflicts of interest to declare.

Data availability

No primary research results, software or code have been included, and no new data were generated or analysed as part of this review.

Acknowledgements

A. S. P. and F.C. acknowledge the support from Fundação para a Ciência e Tecnologia (FCT), through MOSTMICRO-ITQB R&D Unit (UIDB/040612/2020, UIDP/04612/2020), LS4FUTURE Associated Laboratory (LA/P/0087/2020). A.S.P. and F.C. are grateful to FCT for the researcher contracts under the Scientific Employment Stimulus–Individual Calls 2021 and 2022, specifically 2021.01283.CEECIND/CP1657/CT0004 for A.S.P. and 2022.05842.CEECIND/CP1725/CT0001 for F.C. L.M. also acknowledge the support from FCT for the projects (i) UIDP/04378/2020 and UIDB/04378/2020 (Research Unit on Applied Molecular Biosciences—UCIBIO) and (ii) LA/P/0140/2020 (Associate Laboratory Institute for Health and Bioeconomy—i4HB). A. C. B. acknowledge the funding by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., in the scope of the projects LA/P/0037/2020, UIDP/50025/2020 and UIDB/50025/2020 of the Associate Laboratory Institute of Nanostructures, Nanomodelling and Nanofabrication – i3N. The consortium acknowledges the funding from ''la Caixa'' Foundation under the grant agreement (LCF/TR/CI23/56000027), and the InnOValley Proof of Concept under the grant agreement IOVPoC-2022-07 funded by Oeiras Municipality and co-funded by MOSTMICRO-ITQB.

Notes and references

1 B. J. Stapley and T. P. Creamer, *Protein Sci.*, 1999, **8**, 587–595.
2 A. A. Adzhubei and M. J. E. Sternberg, *J. Mol. Biol.*, 1993, **229**, 472–493.
3 Y. Mansiaux, A. P. Joseph, J.-C. Gelly and A. G. de Brevern, *PLoS One*, 2011, **6**, e18401.
4 P. Kumar and M. Bansal, *J. Struct. Biol.*, 2016, **196**, 414–425.
5 P. Wilhelm, B. Lewandowski, N. Trapp and H. Wennemers, *J. Am. Chem. Soc.*, 2014, **136**, 15829–15832.
6 K. T. O'Brien, C. Mooney, C. Lopez, G. Pollastri and D. C. Shields, *R. Soc. Open Sci.*, 2020, **7**, 191239.
7 M. Kjaergaard, A. Nørholm, R. Hendus-Altenburger, S. F. Pedersen, F. M. Poulsen and B. B. Kragelund, *Protein Sci.*, 2010, **19**, 1555–1564.
8 S. Ambadipudi and M. Zweckstetter, *Expert Opin. Drug Discovery*, 2016, **11**, 65–77.
9 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.
10 T. J. Narwani, H. Santuz, N. Shinada, A. Melarkode Vattekatte, Y. Ghouzam, N. Srinivasan, J.-C. Gelly and A. G. de Brevern, *Amino Acids*, 2017, **49**, 705–713.
11 R. Chebrek, S. Leonard, A. G. de Brevern and J.-C. Gelly, *Database*, 2014, 2014, bau102–bau102.
12 M. V. Cubellis, F. Caillez, T. L. Blundell and S. C. Lovell, *Proteins:Struct., Funct., Bioinf.*, 2005, **58**, 880–892.
13 J. A. Vila, H. A. Baldoni, D. R. Ripoll, A. Ghosh and H. A. Scheraga, *Biophys. J.*, 2004, **86**, 731–742.
14 M. Moradi, V. Babin, C. Sagui and C. Roland, *Biophys. J.*, 2011, **100**, 1083–1093.
15 J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis and J. M. Jumper, *Nature*, 2024, **630**, 493–500.
16 J. Rey, S. Murail, S. de Vries, P. Derreumaux and P. Tuffery, *Nucleic Acids Res.*, 2023, **51**, W432–W437.
17 M. Michaelis, L. Cupellini, C. Mensch, C. C. Perry, M. Delle Piane and L. Colombi Ciacchi, *Chem. Sci.*, 2023, **14**, 8483–8496.

18 P. Pandey and S. S. Mallajosyula, *Phys. Chem. Chem. Phys.*, 2019, **21**, 3903–3917.
19 C. L. Scholl, M. Holmstrup, L. A. Graham and P. L. Davies, *Sci. Rep.*, 2023, **13**, 8880.
20 F.-H. Lin, L. A. Graham, R. L. Campbell and P. L. Davies, *Biophys. J.*, 2007, **92**, 1717–1723.
21 K. Tarnutzer, D. Siva Sankar, J. Dengjel and C. Y. Ewald, *Sci. Rep.*, 2023, **13**, 4490.
22 J.-M. Bourhis, N. Mariano, Y. Zhao, K. Harlos, J.-Y. Exposito, E. Y. Jones, C. Moali, N. Aghajari and D. J. S. Hulmes, *Nat. Struct. Mol. Biol.*, 2012, **19**, 1031–1036.
23 H. Zhang, X. Ji, P. Li, C. Liu, J. Lou, Z. Wang, W. Wen, Y. Xiao, M. Zhang and X. Zhu, *Sci. China: Life Sci.*, 2020, **63**, 953–985.
24 X. Wang, J. C. Schwartz and T. R. Cech, *Nucleic Acids Res.*, 2015, **43**, 7535–7543.
25 V. F. Thompson, D. R. Wieland, V. Mendoza-Leon, H. I. Janis, M. A. Lay, L. M. Harrell and J. C. Schwartz, *J. Biol. Chem.*, 2023, **299**(10), 105237.
26 R. Sopakayang, R. De Vita, A. Kwansa and J. W. Freeman, *J. Theor. Biol.*, 2012, **293**, 197–205.
27 J. Engel and H. P. Bächinger, *Top. Curr. Chem.*, 2005, **247**, 7–33.
28 M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958.
29 J. A. Fallas and J. D. Hartgerink, *Nat. Commun.*, 2012, **3**, 1087.
30 S. P. Boudko and H. P. Bächinger, *Sci. Rep.*, 2016, **6**, 37831.
31 K. Kar, S. Ibrar, V. Nanda, T. M. Getz, S. P. Kunapuli and B. Brodsky, *Biochemistry*, 2009, **48**, 7959–7968.
32 B. Sahariah and B. K. Sarma, *J. Phys. Chem. B*, 2021, **125**, 13394–13405.
33 C. Tang, K. Zhou, Y. Zhu, W. Zhang, Y. Xie, Z. Wang, H. Zhou, T. Yang, Q. Zhang and B. Xu, *Food Hydrocoll.*, 2022, **131**, 107748.
34 D. Kodr, T. Fiala and H. Wennemers, *Tetrahedron Lett.*, 2024, **138**, 154964.
35 T. Fiala, E. P. Barros, M.-O. Ebert, E. Ruijsenaars, S. Riniker and H. Wennemers, *J. Am. Chem. Soc.*, 2022, **144**, 18642–18649.
36 N. J. Bulleid, J. A. Dalley and J. F. Lees, *EMBO J.*, 1997, **16**, 6694–6701.
37 A. A. Jalan, D. Sammon, J. D. Hartgerink, P. Brear, K. Stott, S. W. Hamaia, E. J. Hunter, D. R. Walker, B. Leitinger and R. W. Farndale, *Nat. Chem. Biol.*, 2020, **16**, 423–429.
38 S. K. Holmgren, K. M. Taylor, L. E. Bretscher and R. T. Raines, *Nature*, 1998, **392**, 666–667.
39 L. E. R. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang and J. D. Hartgerink, *Nat. Chem.*, 2011, **3**, 821–828.
40 S. E. Paramonov, V. Gauba and J. D. Hartgerink, *Macromolecules*, 2005, **38**, 7555–7561.
41 C. G. Fields, C. M. Lovdahl, A. J. Miles, V. L. M. Hageini and G. B. Fields, *Biopolymers*, 1993, **33**, 1695–1707.
42 C. G. Fields, B. Grab, J. L. Lauer, A. J. Miles, Y. Yu and G. B. Fields, *Lett. Pept. Sci.*, 1996, **3**, 3–16.
43 J. Ottl and L. Moroder, *J. Am. Chem. Soc.*, 1999, **121**, 653–661.
44 J. Ottl, R. Battistuta, M. Pieper, H. Tschesche, W. Bode, K. Kühn and L. Moroder, *FEBS Lett.*, 1996, **398**, 31–36.
45 A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry*, 2000, **39**, 14960–14967.
46 A. V. Persikov, J. A. M. Ramshaw and B. Brodsky, *J. Biol. Chem.*, 2005, **280**, 19343–19349.
47 S. K. Holmgren, L. E. Bretscher, K. M. Taylor and R. T. Raines, *Chem. Biol.*, 1999, **6**, 63–70.
48 L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 777–778.
49 A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry*, 2005, **44**, 1414–1422.
50 C.-H. Chiang, Y.-H. Fu and J.-C. Horng, *Biomacromolecules*, 2017, **18**, 985–993.
51 J. P. Orgel, T. J. Wess and A. Miller, *Structure*, 2000, **8**, 137–142.
52 E. Y. Jones and A. Miller, *Biopolymers*, 1987, **26**, 463–480.
53 F. W. Kotch and R. T. Raines, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 3028–3033.
54 I. C. Tanrikulu and R. T. Raines, *J. Am. Chem. Soc.*, 2014, **136**, 13490–13493.
55 J. A. Fallas, V. Gauba and J. D. Hartgerink, *J. Biol. Chem.*, 2009, **284**, 26851–26859.
56 V. Gauba and J. D. Hartgerink, *J. Am. Chem. Soc.*, 2007, **129**, 15034–15041.
57 H.-S. Chiu and J.-C. Horng, *J. Phys. Chem. B*, 2021, **125**, 7351–7358.
58 C.-C. Chen, W. Hsu, K.-C. Hwang, J. R. Hwu, C.-C. Lin and J.-C. Horng, *Arch. Biochem. Biophys.*, 2011, **508**, 46–53.
59 N. B. Hentzen, V. Islami, M. Köhler, R. Zenobi and H. Wennemers, *J. Am. Chem. Soc.*, 2020, **142**, 2208–2212.
60 D. R. Walker, S. A. H. Hulgan, C. M. Peterson, I.-C. Li, K. J. Gonzalez and J. D. Hartgerink, *Nat. Chem.*, 2021, **13**, 260–269.
61 C. C. Cole, D. R. Walker, S. A. H. Hulgan, B. H. Pogostin, J. W. R. Swain, M. D. Miller, W. Xu, R. Duella, M. Misiura, X. Wang, A. B. Kolomeisky, G. N. Philips and J. D. Hartgerink, *Nat. Chem.*, 2024, **16**, 1698–1704.
62 S. A. H. Hulgan and J. D. Hartgerink, *Biomacromolecules*, 2022, **23**, 1475–1489.
63 J. A. Fallas, L. E. R. O'Leary and J. D. Hartgerink, *Chem. Soc. Rev.*, 2010, **39**, 3510.
64 L. L. C. Olijve, K. Meister, A. L. DeVries, J. G. Duman, S. Guo, H. J. Bakker and I. K. Voets, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 3740–3745.
65 A. C. Deleray, S. S. Saini, A. C. Wallberg and J. R. Kramer, *Chem. Mater.*, 2024, **36**, 3424–3434.
66 M. Urbańczyk, J. Góra, R. Latajka and N. Sewald, *Amino Acids*, 2017, **49**, 209–222.
67 W. Zhang, H. Liu, H. Fu, X. Shao and W. Cai, *J. Phys. Chem. B*, 2022, **126**, 10637–10645.
68 J. Garner and M. M. Harding, *ChemBioChem*, 2010, **11**, 2489–2498.
69 C. Chen, Y. Zhang, J. Sun, Y. Liu, Y. Qin, Z. Ling, W. Liu and W. Li, *Fuel*, 2022, **327**, 125060.
70 S. Cui, W. Zhang, X. Shao and W. Cai, *J. Chem. Inf. Model.*, 2022, **62**, 5165–5174.

71 W. Jiang, F. Yang, X. Chen, X. Cai, J. Wu, M. Du, J. Huang and S. Wang, *J. Futur. Foods*, 2022, **2**, 203–212.
72 G. Giubertoni, K. Meister, A. L. DeVries and H. J. Bakker, *J. Phys. Chem. Lett.*, 2019, **10**, 352–357.
73 K. Mochizuki and V. Molinero, *J. Am. Chem. Soc.*, 2018, **140**, 4803–4811.
74 Y. Sun, G. Giubertoni, H. J. Bakker, J. Liu, M. Wagner, D. Y. W. Ng, A. L. Devries and K. Meister, *Biomacromolecules*, 2021, **22**, 2595–2603.
75 C. L. Scholl, S. Tsuda, L. A. Graham and P. L. Davies, *FEBS J.*, 2021, **288**, 4332–4347.
76 M. Á. Treviño, D. Pantoja-Uceda, M. Menéndez, M. V. Gomez, M. Mompeán and D. V. Laurents, *J. Am. Chem. Soc.*, 2018, **140**, 16988–17000.
77 M. M. Harding, P. I. Anderberg and A. D. J. Haymet, *Eur. J. Biochem.*, 2003, **270**, 1381–1392.
78 A. N. Lane, L. M. Hays, N. Tsvetkova, R. E. Feeney, L. M. Crowe and J. H. Crowe, *Biophys. J.*, 2000, **78**, 3195–3207.
79 N. M. Tsvetkova, B. L. Phillips, V. V. Krishnan, R. E. Feeney, W. H. Fink, J. H. Crowe, S. H. Risbud, F. Tablin and Y. Yeh, *Biophys. J.*, 2002, **82**, 464–473.
80 Y. Tachibana, G. L. Fletcher, N. Fujitani, S. Tsuda, K. Monde and S.-I. Nishimura, *Angew. Chem., Int. Ed.*, 2004, **43**, 856–862.
81 C. Her, Y. Yeh and V. V. Krishnan, *Biomolecules*, 2019, **9**, 235.
82 A. M. Brown and N. J. Zondlo, *Biochemistry*, 2012, **51**, 5041–5051.
83 C. Chen, M. Liu, X. Wu, Q. Yang, F. Yao, H. Zhang and J. Li, *J. Phys. Chem. C*, 2023, **127**, 9327–9335.
84 L. A. Graham, M. E. Boddington, M. Holmstrup and P. L. Davies, *Sci. Rep.*, 2020, **10**, 3047.
85 X. Chen, J. Wu, X. Li, F. Yang, D. Huang, J. Huang, S. Wang and V. Guyonnet, *npj Sci. Food*, 2022, **6**, 10.
86 B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff and S. B. H. Kent, *J. Am. Chem. Soc.*, 2008, **130**, 9695–9701.
87 Y.-F. Mok, F.-H. Lin, L. A. Graham, Y. Celik, I. Braslavsky and P. L. Davies, *Biochemistry*, 2010, **49**, 2593–2603.
88 M. Treviño, R. López-Sánchez, M. R. Moya, D. Pantoja-Uceda, M. Mompeán and D. V. Laurents, *Biophys. J.*, 2022, **121**, 4560–4568.
89 C. S. Rodríguez and D. V. Laurents, *Arch. Biochem. Biophys.*, 2024, **756**, 109981.
90 R. López-Sánchez, D. V. Laurents and M. Mompeán, *Commun. Chem.*, 2024, **7**, 191.
91 J.-C. Horng and R. T. Raines, *Protein Sci.*, 2006, **15**, 74–83.
92 C. L. Scholl and P. L. Davies, *FEBS Lett.*, 2023, **597**, 538–546.
93 G. Todde, C. Whitman, S. Hovmöller and A. Laaksonen, *J. Phys. Chem. B*, 2014, **118**, 13527–13534.
94 H. Qi, Y. Gao, L. Zhang, Z. Cui, X. Sui, J. Ma, J. Yang, Z. Shu and L. Zhang, *Engineering*, 2024, **34**, 164–173.
95 L. F. Kong, A. A. Qatran Al-Khdhairawi and B. A. Tejo, *Biocatal. Agric. Biotechnol.*, 2020, **23**, 101447.
96 E. Gandini, M. Sironi and S. Pieraccini, *J. Mol. Graph. Model.*, 2020, **100**, 107680.
97 Z. Muganlı, S. Saeidiharzand, R. Rekuviene, V. Samaitis, A. Jankauskas, A. Koşar, G. Gharib and A. Sadaghiani, *Adv. Mater. Interfaces*, 2023, **10**(14), 2300021.
98 R. J. de Haas, R. P. Tas, D. van den Broek, C. Zheng, H. Nguyen, A. Kang, A. K. Bera, N. P. King, I. K. Voets and R. de Vries, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, 2017.
99 M. Calvaresi, S. Höfinger and F. Zerbetto, *Biomacromolecules*, 2012, **13**, 2046–2052.
100 M. J. Kuiper, C. J. Morton, S. E. Abraham and A. Gray-Weale, *eLife*, 2015, **4**, 1–14.
101 R. Miyata, Y. Moriwaki, T. Terada and K. Shimizu, *Heliyon*, 2021, **7**, e07953.
102 T. J. McPartlon, C. T. Osborne and J. R. Kramer, *Biomacromolecules*, 2024, **25**, 3325–3334.
103 J. Ma, X. Zhang, Z. Cui, M. Zhao, L. Zhang and H. Qi, *J. Mater. Chem. B*, 2023, **11**, 4042–4049.
104 S. Wang and S. Damodaran, *J. Agric. Food Chem.*, 2009, **57**, 5501–5509.
105 J. J. Bouchard, J. H. Otero, D. C. Scott, E. Szulc, E. W. Martin, N. Sabri, D. Granata, M. R. Marzahn, K. Lindorff-Larsen, X. Salvatella, B. A. Schulman and T. Mittag, *Mol. Cell*, 2018, **72**, 19–36.
106 Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez and J. H. Waite, *Nat. Chem. Biol.*, 2015, **11**, 488–495.
107 M. Mompeán, J. Oroz and D. V. Laurents, *FEBS Open Bio*, 2021, **11**, 2390–2399.
108 M. Mompeán, B. S. McAvan, S. S. Félix, M. Á. Treviño, J. Oroz, R. López-Sánchez, D. Pantoja-Uceda, E. J. Cabrita, A. J. Doig and D. V. Laurents, *Arch. Biochem. Biophys.*, 2021, **704**, 108867.
109 A. M. Janke, D. H. Seo, V. Rahmanian, A. E. Conicella, K. L. Mathews, K. A. Burke, J. Mittal and N. L. Fawzi, *Biochemistry*, 2018, **57**, 2549–2563.
110 D. Maltseva, S. Chatterjee, C.-C. Yu, M. Brzezinski, Y. Nagata, G. Gonella, A. C. Murthy, J. C. Stachowiak, N. L. Fawzi, S. H. Parekh and M. Bonn, *Nat. Chem.*, 2023, **15**, 1146–1154.
111 E. Bertrand, C. Demongin, I. Dobra, J. C. Rengifo-Gonzalez, A. S. Singatulina, M. V. Sukhanova, O. I. Lavrik, D. Pastré and L. Hamon, *Sci. Rep.*, 2023, **13**, 7772.
112 L. R. Ganser, A. G. Niaki, X. Yuan, E. Huang, D. Deng, N. A. Djaja, Y. Ge, A. Craig, O. Langlois and S. Myong, *Structure*, 2024, **32**, 177–187.
113 X. Zhang, M. Vigers, J. McCarty, J. N. Rauch, G. H. Fredrickson, M. Z. Wilson, J.-E. Shea, S. Han and K. S. Kosik, *J. Cell Biol.*, 2020, **219**, e202006054.
114 M. Boehning, C. Dugast-Darzacq, M. Rankovic, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq and M. Zweckstetter, *Nat. Struct. Mol. Biol.*, 2018, **25**, 833–840.
115 C. Camilloni, A. De Simone, W. F. Vranken and M. Vendruscolo, *Biochemistry*, 2012, **51**, 2224–2231.
116 M. J. Macias, S. Wiesner and M. Sudol, *FEBS Lett.*, 2002, **513**, 30–37.
117 P. Cramer, D. A. Bushnell and R. D. Kornberg, *Science*, 2001, **292**, 1863–1876.
118 C. Brasnett, A. Kiani, S. Sami, S. Otto and S. J. Marrink, *Commun. Chem.*, 2024, **7**, 151.
119 D. M. Mitrea, M. Mittasch, B. F. Gomes, I. A. Klein and M. A. Murcko, *Nat. Rev. Drug Discovery*, 2022, **21**, 841–862.

120 A. Baruch Leshem, S. Sloan-Dennison, T. Massarano, S. Ben-David, D. Graham, K. Faulds, H. E. Gottlieb, J. H. Chill and A. Lampel, *Nat. Commun.*, 2023, **14**, 421.
121 J. L. S. Lopes, A. J. Miles, L. Whitmore and B. A. Wallace, *Protein Sci.*, 2014, **23**, 1765–1772.
122 D. N. Edun, M. R. Flanagan and A. L. Serrano, *Chem. Sci.*, 2021, **12**, 2474–2479.
123 A. L. Rucker and T. P. Creamer, *Protein Sci.*, 2002, **11**, 980–985.
124 L. M. F. Bertoline, A. N. Lima, J. E. Krieger and S. K. Teixeira, *Front. Bioinf.*, 2023, **3**, 1–8.
125 M. Varadi, M. Tsenkov and S. Velankar, *Proteins:Struct., Funct., Bioinf.*, 2025, **93**, 400–410.
126 C. J. Wilson, W.-Y. Choy and M. Karttunen, *Int. J. Mol. Sci.*, 2022, **23**, 4591.
127 A. F. Drake, G. Siligardi and W. A. Gibbons, *Biophys. Chem.*, 1988, **31**, 143–146.
128 M. G. Venugopal, J. A. M. Ramshaw, E. Braswell, D. Zhu and B. Brodsky, *Biochemistry*, 1994, **33**, 7948–7956.
129 V. Kubyshkin, J. Bürck, O. Babii, N. Budisa and A. S. Ulrich, *Phys. Chem. Chem. Phys.*, 2021, **23**, 26931–26939.
130 O. D. Krishna and K. L. Kiick, *Biomacromolecules*, 2009, **10**, 2626–2631.
131 K. E. Drzewiecki, D. R. Grisham, A. S. Parmar, V. Nanda and D. I. Shreiber, *Biophys. J.*, 2016, **111**, 2377–2386.
132 B. Bochicchio and A. M. Tamburro, *Chirality*, 2002, **14**, 782–792.
133 E. R. Kellenbach, R. K. Dukor and L. A. Nafie, *Spectrosc. Eur.*, 2007, **19**, 15–18.
134 J. Kapitán, V. Baumruk, V. Gut, J. Hlaváček, H. Dlouhá, M. Urbanová, E. Wünsch and P. Maloň, *Collect. Czechoslov. Chem. Commun.*, 2005, **70**, 403–409.
135 R. K. Dukor and T. A. Keiderling, *Biopolymers*, 1991, **31**, 1747–1761.
136 J. S. Cobb, V. Zai-Rose, J. J. Correia and A. V. Janorkar, *ACS Omega*, 2020, **5**, 8403–8413.
137 C. D. Syme, E. W. Blanch, C. Holt, R. Jakes, M. Goedert, L. Hecht and L. D. Barron, *Eur. J. Biochem.*, 2002, **269**, 148–156.
138 M. Furuta, T. Fujisawa, H. Urago, T. Eguchi, T. Shingae, S. Takahashi, E. W. Blanch and M. Unno, *Phys. Chem. Chem. Phys.*, 2017, **19**, 2078–2086.
139 E. W. Blanch, L. A. Morozova-Roche, D. A. E. Cochran, A. J. Doig, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 2000, **301**, 553–563.
140 E. W. Blanch, A. C. Gill, A. G. O. Rhie, J. Hope, L. Hecht, K. Nielsen and L. D. Barron, *J. Mol. Biol.*, 2004, **343**, 467–476.
141 C. Mensch, P. Bultinck and C. Johannessen, *ACS Omega*, 2018, **3**, 12944–12955.
142 F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *Structure*, 2005, **13**, 1409–1419.
143 A. Urbanek, M. Popovic, C. A. Elena-Real, A. Morató, A. Estaña, A. Fournet, F. Allemand, A. M. Gil, C. Cativiela, J. Cortés, A. I. Jiménez, N. Sibille and P. Bernadó, *J. Am. Chem. Soc.*, 2020, **142**, 7976–7986.
144 W. M. Borcherds and G. W. Daughdrill, in *Intrinsically Disordered Proteins*, ed. E. B. T.-M. E. Rhoades, Academic Press, 2018, vol. 611, pp. 101–136.
145 S. L. Lam and V. L. Hsu, *Biopolymers*, 2003, **69**, 270–281.
146 K. Pagano, F. Fogolari, A. Corazza, P. Viglino and G. Esposito, *J. Biomol. NMR*, 2007, **39**, 213–222.
147 J. M. Schmidt and F. Löhr, in *Protein Structure*, ed. E. Faraggi, IntechOpen, Rijeka, 2012, ch. 4, pp. 95–120.
148 C. Feng, Z. Wang, G. Li, X. Yang, N. Wu and L. Wang, *BioMed Res. Int.*, 2022, **2022**, 9015123.
149 X. Zhang, B. E. Ramirez, X. Liao and T. G. H. Diekwisch, *PLoS One*, 2011, **6**, e24952.
150 H. D. Middendorf, R. L. Hayward, S. F. Parker, J. Bradshaw and A. Miller, *Biophys. J.*, 1995, **69**, 660–673.
151 A. E. Langkilde, K. L. Morris, L. C. Serpell, D. I. Svergun and B. Vestergaard, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 882–895.
152 D. E. Birk, M. V. Nurminskaya and E. I. Zycband, *Dev. Dyn.*, 1995, **202**, 229–243.
153 B. D. Quan and E. D. Sone, *Methods in Enzymology*, Elsevier Inc, 2013, 1st edn, vol. 532, pp. 189–205.

