




Do polyproline II helix associations modulate biomolecular condensates?

Miguel Mompeán , Javier Oroz  and Douglas V. Laurents 

Departamento de Química Física Biológica, Instituto de Química Física Rocasolano, CSIC, Madrid, España

Keywords

biomolecular condensates; polyproline II helix; SH3 domain

Correspondence

D. V. Laurents, Departamento de Química Física Biológica, Instituto de Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, España
E-mail: dlaurents@iqfr.csic.esMiguel Mompeán and Javier Oroz are
Co-first authors(Received 7 November 2020, revised 29
March 2021, accepted 9 April 2021)

doi:10.1002/2211-5463.13163

Edited by Irene Díaz-Moreno

Biomolecular condensates are microdroplets that form inside cells and serve to selectively concentrate proteins, RNAs and other molecules for a variety of physiological functions, but can contribute to cancer, neurodegenerative diseases and viral infections. The formation of these condensates is driven by weak, transient interactions between molecules. These weak associations can operate at the level of whole protein domains, elements of secondary structure or even moieties composed of just a few atoms. Different types of condensates do not generally combine to form larger microdroplets, suggesting that each uses a distinct class of attractive interactions. Here, we address whether polyproline II (PPII) helices mediate condensate formation. By combining with PPII-binding elements such as GYF, WW, profilin, SH3 or OCRE domains, PPII helices help form lipid rafts, nuclear speckles, P-body-like neuronal granules, enhancer complexes and other condensates. The number of PPII helical tracts or tandem PPII-binding domains can strongly influence condensate stability. Many PPII helices have a low content of proline residues, which hinders their identification. Recently, we characterized the NMR spectral properties of a Gly-rich, Pro-poor protein composed of six PPII helices. Based on those results, we predicted that many Gly-rich segments may form PPII helices and interact with PPII-binding domains. This prediction is being tested and could join the palette of verified interactions contributing to biomolecular condensate formation.

Several distinct classes of weak interactions drive the formation of a score of different biomolecular condensates

As described in other articles of this special issue and recent reviews [1], a score of biomolecular condensate is essential for the efficient subcellular organization of the cell. They perform many vital physiological functions, but they are also implicated in cancer [2,3], neurodegenerative diseases [4,5] and viral infections [6]. During the last decade of the 20th century, it became clear that sphingolipids and cholesterol undergo two-

dimensional phase separation in the cell membrane to form special domains called 'lipid rafts' [7]. These 2D condensates also concentrate proteins modified with a glycosylphosphatidylinositol anchor. The specific localization and concentration of such proteins into lipid rafts play key roles in signalling and vesicle transport.

Several years after these foreshadowing findings, germline granules were discovered. They also have nonaqueous liquid properties and concentrate certain mRNAs and proteins necessary for early embryonic differentiation [8]. This pioneering study stimulated a general reappraisal of three-dimensional subcellular

Abbreviations

CPEB, cytoplasmic polyadenylation element-binding (protein); CTD, C-terminal domain; PRM, proline-rich motif; RNA pol II, RNA polymerase II; sfAFP, snow flea antifreeze protein; SH3, Src-homology 3 (domain); VASP, vasodilator-stimulated phosphoprotein; WASP, Wiskott-Aldrich syndrome protein; WWW, tryptophan-tryptophan (domain); ZnF, zinc finger.

entities such as germ bodies, Cajal bodies, stress granules and nucleoli, and now, these all have been recognized as being distinct liquid phases [1]. Their most notable features include (a) a content enriched in certain proteins, RNA and metabolites, while excluding some water; (b) rapid formation and dissociation in response to cell conditions; (c) stabilization due to weak and ephemeral interactions such as hydrophobic interactions; (d) molecules within the condensate can exchange from the condensed to the dispersed phases; and (e) fusion with condensates of the same class, but repulsion of other types of condensates. This fifth property is more highly developed in stress granules, which contain two nonmixing layers [9] and nucleoli [10], which consist of three separate layers. The complex organization of the nucleolus was proposed to constitute an assembly line for efficient ribosome biosynthesis [10].

The mutual avoidance of different types of condensates implies that their stabilizing interactions must be more sophisticated than simple hydrophobic interactions [11]. In fact, to date several different kinds of stabilizing contacts have been identified, ranging from π - π and cation- π interactions between small groups of atoms in aromatic and cationic residues [12] to hydrophobic α -helices [13] to the specific yet weak interactions among entire folded domains such as the N-terminal domain of transactive response DNA-binding protein of 43 kDa [14] or the pentamerization domain of nucleophosmin [15]. In some cases, different classes of weak contacts, namely hydrophobic, π - π and sp^2 - π interactions, as well as hydrogen bonds, combine to promote condensate formation [16]. Sp^2 interactions include those formed by delocalized electrons in backbone and side-chain amide groups and aromatic moieties. In addition to these interactions, there are cases of proline-rich PPII helices interacting with folded domains to stabilize condensates [1], and even a case of a putative PPII helix formed by nonproline residues interacting with a folded domain to drive condensate formation has been reported [17]. Here, we will review the reported cases of PPII helices contributing to condensate formation and our recent proposal that PPII helices formed by glycine-rich PPII helices may interact with each other or other partners to promote the formation of condensates [18,19].

The unique structure and interactions of polyproline II helices

The PPII helix was first characterized in peptides composed of proline residues in aqueous solution in the 1950s [20]. Compared to the well-known right-handed

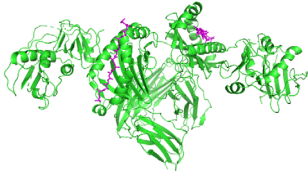
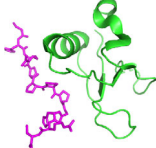
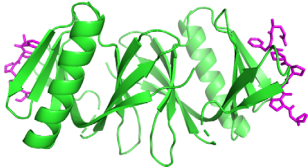
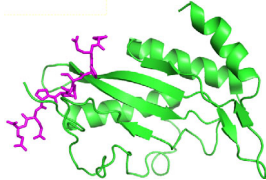


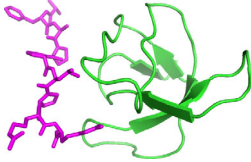
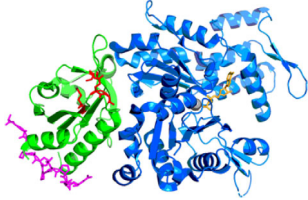
α -helix, the PPII helix is left-handed and makes one turn exactly every three residues. Replacing every third Pro residue with Gly allows three PPII helices to associate as tiny glycine can fit and form H-bonds at the small, interhelical position. This PPII triple helix is the basis of collagen, the most abundant protein in the human body. Other residues besides proline can also adopt the PPII helical conformation. For example, peptides composed of charged residues such as lysine or glutamic acid form PPII helices [21]. More interestingly, segments rich in glycine residues can form bundles of PPII helices as observed in a number of natural proteins, for example acetophenone carboxylase [22]. Many 'loop' segments in globular proteins [23] and 'unfolded' segments of intrinsically disordered proteins [24,25] actually adopt the PPII conformation.

An isolated PPII helix is stiff and exposes both side chains and backbone to solvent. These properties favour association with other proteins. For example, in the striking structure of human acetylcholinesterase, the tetramerization motif consists of a central PPII helix bound to four α -helices [26]. Moreover, PPII helices frequently bind folded domains to mediate protein/protein interactions. Currently, eight classes of PPII helix-binding proteins/protein domains have been reported: (a) the class II major histocompatibility complex (class II MHC), (b) the glycine-tyrosine-phenylalanine domain, (c) the enabled VASP homology domain, (d) the ubiquitin E2 variant (UEV) domain, (e) the tryptophan-tryptophan (WW) domain, (f) the octamer repeat of aromatic residues domain, (g) the Src-homology 3 (SH3) domain and (h) profilin. Many of these domains are small, have the N and C termini close together to facilitate modular protein architectures with domain repeats and use exposed aromatic residues to bind a PPII helix. Some of their characteristics are shown in Table 1. In the following paragraphs, we highlight the roles of some of these domains in some particularly fascinating biomolecular condensates.

A PPII helix at the C terminus of RNA polymerase mediates the formation of both superenhancer condensates and nuclear paraspeckle condensates for splicing

One triumph of structural biology was the elucidation of the conformation and mechanism of RNA polymerase II (RNA pol II) and transcription factors [27], for which Roger Kornberg received the Nobel Prize in 2006. These discoveries currently aid the understanding of viral RNA polymerases, such as the SARS-CoV-2 replicase. However, one domain of RNA pol II

Table 1. Characteristics of PPII-binding domains.

Domain /Protein	No. of Residues	Biological Role	Binds to	Structure ^a	Reference
Class II MHC	1214	Displays pathogen-derived peptides for T-cell activation	Non-human protein fragments		[67,68]: PDB 1SEB
GYF domain	62	CD2 signaling, immune cell activation	SHRPPPPGHRV		[69,70]: PDB:1L2Z
EVH domain	115	Actin-based motile and neural development processes	EFPPPPPT		[71]: PDB:1QC6
UEV domain	145	Cytokinesis, viral particle budding	PSAP; PTAP		[72]: PDB: 1M4Q
WW domain	35–40	Cytoskeleton, Hippo signaling. Many others	PPxY; LPxY; PPR; PGM motifs; PR; phosphoS-P; phosphoT-P		[73]: PDB 1K9Q
OCRE domain	55	Alternative splicing regulation; binds spliceosome protein SmN	RPPPPGIR		[74,75]: PDB: 5MF9
SH3 domain	60	Cytoskeleton regulation, condensate formation	R,KxxPxxP; PxxPxR,K		[76]: PDB: 1PRL
Profilin	139	Promotes actin filament formation	S,A,T,G(PPPP..)L		[77,78]: PDB: 1CJF

^aFor each structure, the PPII helical ligand is shown in magenta and the binding domain is shown in green. For the WW domain, the two conserved Trp residues are shown in blue. In the case of profilin, residues whose mutation is linked to ALS are shown in red and actin is shown in blue.

initially eluded elucidation: over 350 residues at the C terminus were missing in the X-ray structure. The sequence of this absent C-terminal domain (CTD) consists of 52 repeats of a consensus seven-residue sequence: YSPTSPS. Robert Woody, a top expert in circular dichroism, used this spectroscopic technique to reveal that the conformational ensemble of RNA pol II CTD is mainly statistical coil with a significant proportion of PPII helix [28]. He and his team predicted that transcription factors with WW or SH3 domains could use this PPII helix to bind RNA pol II.

One such protein is Mediator, which is in fact a large set of protein factors that interact with RNA pol II and DNA regulator sequences distant from the transcribed gene to strongly enhance or repress transcription [29]. When bound to Mediator, the CTD of RNA pol II adopts two PPII helices connected by a turn [30]. Work over the last 15 years has shown that the Mediator complex is essential for forming a phase-separated ‘superenhancer’ condensate, which dramatically increases transcription rates and plays important physiological roles as well as in cancer [31].

Despite the elegance and importance of this mechanism, Sharp and others wondered whether the RNA Pol II CTD could be moonlighting. They realized that the CTD heptad repeat is rich in Ser, Tyr and Thr residues, which can be phosphorylated, and might endow the RNA Pol II CTD with the ability to interact with a second set of partners. In 2019, they reported that following phosphorylation by CDK7/CDK9, the RNA Pol II CTD stops interacting with Mediator and the enhancer complex and instead nucleates the formation of a second type of biocondensate, called nuclear paraspeckles, by binding to arginine-rich splicing factors of the spliceosome [32]. This phosphorylated form of RNA Pol II also adopts a PPII helix [33] or a mixed PPII/extended conformation [34] when bound to partners. These findings are reminiscent of an early scene from the film ‘Cinema Paradiso’ (<https://www.youtube.com/watch?v=qMqE1Fayk28>) where the village priest signals Alfredo, the cinema technician, to mark the kissing scenes for cutting and splicing. In an analogous fashion, the phosphorylation of RNA pol II CTD (the priest’s bell) disrupts the association of the enhancer complex (kissing couples) and induces the formation of the spliceosome/nuclear paraspeckles (Alfredo and his scissors).

Two-faced profilin binds actin and PPII helices to modulate condensate formation

Profilin is a small protein with separate binding sites for actin and PPII helices. It is well known for

promoting changes in the actin cytoskeleton for cell development and motion. First, profilin binds to an actin monomer and promotes ADP→ATP exchange. Then, profilin–actin pairs are channelled into growing actin filaments as profilin binds to PPII helices in proteins such as WASP and VASP. In Huntington’s disease, toxic monomers and aggregates of the N-terminal fragment of the huntingtin protein (Htt NTF) disrupt normal phase separation processes and induce necrosis and apoptosis [35]. These Htt NTFs contain an expanded polyQ tract followed by polyproline segments, which adopt polyproline II helices with occasional kinks [36]. Interestingly enough, profilin can bind to the PPII tracts of Htt NTF and reduce their cytotoxicity [37,38].

Profilin is also essential for the formation of an extensive actin filament network at dendritic spines [39], which is essential for memory consolidation [40]. This actin filament network is highly dynamic, turning over within minutes [41], so how can some memories last a lifetime? Cytoplasmic polyadenylation element-binding protein (CPEB) whose functional aggregation is essential for long-term memory [42] is present in P-body-like neuronal granules at dendritic spines [43]. CPEB’s folded RNA recognition motif and ZnF domains retain and repress mRNAs, which are key for memory consolidation. Following stimulation, CPEB’s N-terminal disordered region, which contains polyglutamine and polyproline segments, forms a highly stable functional amyloid [44] leading to the release and activation of the retained mRNAs. As a working hypothesis, we recently proposed that after amyloid formation, the PPII helices by human CPEB3’s polyproline tracts could be favourably positioned to interact with profilin and orchestrate a permanent fortification of the local actin cytoskeleton [45].

SH3 domains reveal how affinity and avidity control aqueous ↔ liquid condensate ↔ solid-phase transitions

The SH3 domain is the most common and versatile PPII helix-binding element [46]. Like most PPII-binding motifs, the SH3 domain is modular in nature, as many proteins have a few or several SH3 domains strung close together along their sequence. For example, the protein Nck contains three SH3 domains and binds to six proline-rich motifs (PRMs), which putatively adopt PPII helices, in a second protein called N-WASP to form part of the glomerular filtration barrier in kidney podocytes. In 2012, Li *et al.* [47] exploited this system to uncover how changing the strength of the SH3 domain/PRM interaction, as well as the

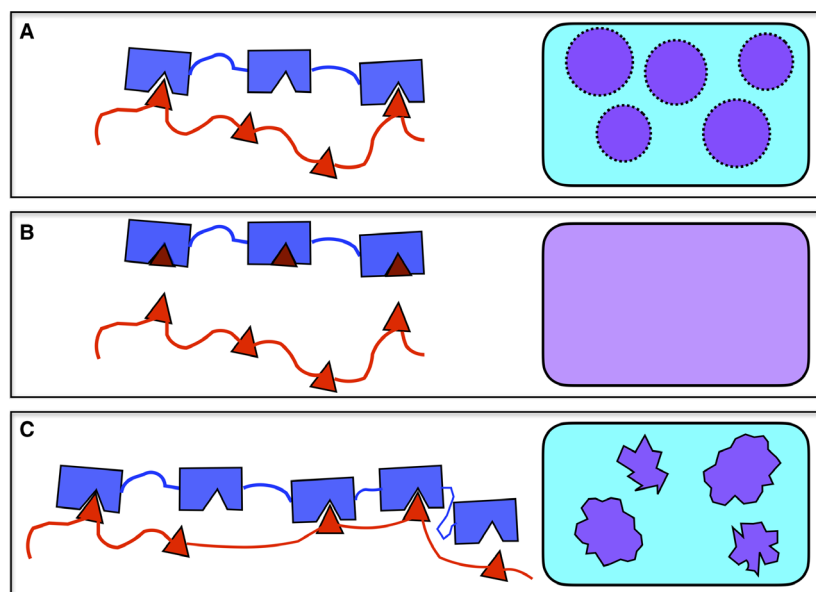


Fig. 1. Formation of soluble oligomers, phase-separated liquid microdroplets and hydrogels by SH3 domains + PPII helices. (A) Three modular SH3 domains (blue) with moderate affinity for four PPII helices linked on the same polypeptide chain (red) form phase-separated liquid microdroplets (right, purple). (B) The addition of high-affinity PPII helical monomers (dark red) leads to the displacement of the moderate-affinity ligands (red) from the SH3 domains (blue) and the dissolution of the microdroplets. (C) Increasing the number of linked SH3 domains (blue) and moderate-affinity PPII helices (red) to five each leads to the formation of rigid hydrogels (right, purple). Figure based on the results of Ref. [47].

number of interacting SH3 domains and PRM, affects phase transitions. They discovered that when a single SH3 domain binds to a high-affinity PRM, a stable heterodimer is formed, which remains in the aqueous phase (Fig. 1A). In contrast, when an aqueous solution containing polypeptides with three SH3 domains (SH3)₃ with moderate affinity for PRM is mixed with a second solution containing polypeptides with four PRM (PRM)₄, liquid/liquid phase separation occurs as the polypeptides combine to form condensates (Fig. 1B). It is fascinating that the addition of a monovalent, high-affinity PRM to this system causes the condensate to break up, as the tight binding ligand displaces the low-affinity, medium-avidity (PRM)₄ polypeptide. As the reader might have already guessed, moderate-affinity systems with more SH3 domains and PRMs, such as (SH3)₅ plus (PRM)₅, produced semi-solid, gel-like condensates (Fig. 1C) [47].

Recognition of noncanonical segments by an SH3 domain can also modulate condensation formation

Whereas most SH3 domain ligands contain proline residues, SH3 domains that bind class I ligands can also recognize an RKXXYXXY motif, where X is a small, nonproline residue [48]. As such proline-less segments are common, their interactions with SH3 domains could impact condensate formation. Certain mRNAs are transported along dendrites in a type of condensate called ‘neuronal granules’ by the protein hnRNPA2. The mRNA cargo is released when the

kinase Fyn is incorporated into the neuronal granule at its destination and phosphorylates the hnRNPA2 [49]. Fyn uses its SH3 domain to bind to hnRNPA2. However, hnRNPA2 does not contain the PXXP motif recognized by SH3 domains. Intrigued by this observation, Amaya, Ryan and Fawzi recently used NMR spectroscopy to study the interaction between the C-terminal disordered region of hnRNPA2 with Fyn [17]. They found evidence that Fyn may bind to residues Y₃₃₅GGRSRY₃₄₁, in the ‘disordered’ region of hnRNPA2, right at the C terminus of hnRNPA2. Interestingly, if this segment were to adopt a PPII helix, the R and Y side chains would be favourably positioned to form cation- π interactions. This suggests that segments lacking PRMs may be able to adopt the PPII helical conformation and bind to SH3 domains to contribute to biomolecular condensate formation and dissociation.

Proteins formed by bundles of Gly-rich PPII helices have diverse biological functions

Whereas the polyproline II helix is commonly associated with proline residues, it can also be adopted by proteins rich in glycine. These proteins share a conformation based on that of polyglycine peptides, which adopt a network of PPII helices stabilized by interhelical N-H||O=C [50] and C α -H||O=C hydrogen bonds [51]. In the case of acetophenone carboxylase, the β -subunit contains a complete hexagonal bundle with six PPII helices surrounding a central seventh PPII helix

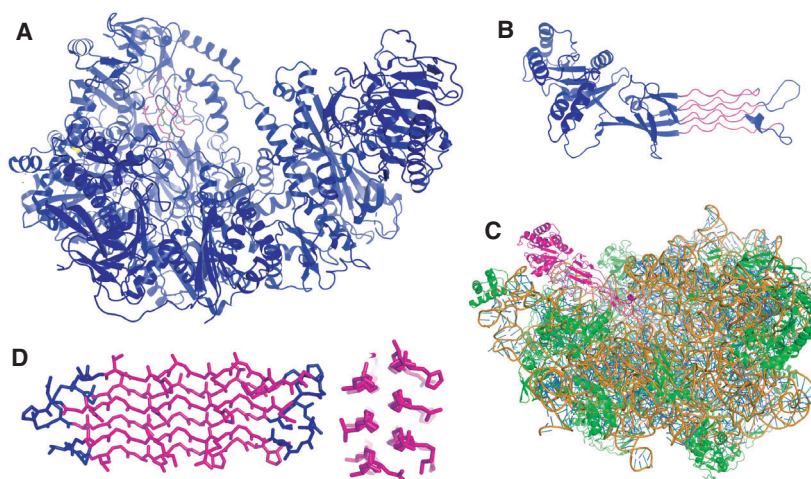


Fig. 2. Glycine-rich PPII helical bundles are found in a variety of proteins. (A) Ribbon diagram of the *Aromatoleum aromaticum* acetophenone carboxylase core complex (PDB 5L9W). α -helices, β -strands and loops are shown in blue, and PPII helices are shown in magenta except for the central PPII helix, which is shown in green. (B) X-ray crystal structure of the *Bacillus subtilis* Obg GTPase (PDB: 1LNZ) with α -helices, β -strands and loops are shown in blue, and PPII helices are shown in magenta. (C) Cryo-EM structure of the *E. coli* Obg GTPase bound to the large ribosomal subunit, where it acts to mimic a tRNA and block ribosome assembly (PDB: 1LN7). Here, the entire Obg GTPase is shown in magenta, ribosomal proteins are shown in green and rRNA is shown in gold (backbone) and blue (bases) (PDB: 4CSU). (D) The *Hypogastrum* 'snow flea' antifreeze protein contains six PPII helices (magenta) connected by short loops (blue). The cross section (right) shows the polar (right side) and nonpolar (left) faces (PDB: 3BOG).

[22] (PDB 5L9W; Fig. 2A). This PPII domain is mostly buried inside this enzyme. Based on this crystal structure and sequence analysis, [52] Schühle and Heider proposed that this fold would be shared by all bacterial enzymes of the hydantoinase/carboxylase family.

The GTP-binding protein Obg, a completely unrelated protein, also contains a glycine-rich PPII helical bundle, but in this case, the crystal structure showed that the bundle is mainly solvent-exposed and consists of six PPII helices arranged in two layers (PDB 1LNZ; Fig. 2B) [53]. This essential domain is conserved from bacteria to humans, and remarkably, it acts as a tRNA mimic that binds the large ribosomal subunit to regulate ribosomal assembly (PDB 4CSU; Fig. 2C) [54].

A remarkable domain consisting of ten PPII helices is found at the tip of the spike protein of T-even (except T4) bacteriophages [55]. These helices are arranged in three layers, with the two central PPII helices being composed of six and seven consecutive glycine residues, similar to the structure of polyglycine [50]. Bacteriophages are extremely abundant; for every grain of sand on Earth, there are about a trillion phages [56]. This means that this glycine-rich helical bundle fold is exceptionally plentiful in the biosphere. Whereas glycine-rich PPII bundle proteins are rare relative to those composed by α -helices and β -sheets, their abundance, presence across different biological kingdoms and diversity of folds, which likely evolved

independently, evince the success of this protein structure family.

Glycine-rich PPII helical bundles are also observed in the *Collembola Hypogastruridae* 'snow flea' antifreeze proteins [57]. Initially isolated in Canada, similar *Collembola* proteins have been recently isolated in the Middle East [58], Iceland [59] and Japan [60]. As in Obg, their structures consist of a flat, solvent-exposed, two-layer network of PPII helices (PDB 3BOG; Fig. 2D). However, the snow flea antifreeze proteins can have between six and 13 PPII helices. One face is rich in exposed Ala residues, poor in charged residues and has been reported to bind to nascent ice crystals [61]. A similar fold has also been proposed to be present in a high glycine-tyrosine hair keratin-associated protein [62]. Based on sequence comparison, CD and FTIR spectroscopy results and MD simulations, these researchers advanced a structural model featuring four glycine-rich PPII helices stabilized by two disulfide bonds.

In summary, proteins or protein domains composed of Gly-rich PPII helical bundles have been found in diverse proteins across the kingdoms of life. Both the widespread nature of this fold and the observation that it is overlooked by the most popular secondary structure classification algorithms [22,23] lead us to suggest that it may also be present yet unrecognized in other biological contexts.

Could Gly-rich PPII helices contribute to condensate formation?

The several examples reviewed here highlight how PPII helices can interact with folded globular domains to stabilize biomolecular condensates. Is it possible that condensate formation could also be promoted by Gly-rich PPII helices associating with each other? Condensate formation by fused in sarcoma (FUS) protein is driven by cation- π interactions between its RGG motifs and G/S,Y,G/S repeats [12]. The abundance of glycine and the spacing of Arg residues in the former match the template seen in folded PPII helical bundle proteins such as 'snow flea' antifreeze protein (sfAFP) (Fig. 2D, [57]), so RGG motifs or similar (RGGF)_N repeats in the CTD of nucleolin might form PPII helices. While there are well-developed tools based on NMR chemical shift deviations to identify α -helices and β -sheets and to measure their population in proteins, none were available for PPII helices.

Spurred by this need, we recently characterized ¹³C,¹⁵N-labelled *Hypogastrura harveyi* sfAFP by NMR spectroscopy [18]. This 81-residue protein is composed of 46% glycine and 10% alanine residues, yet it adopts a well-ordered, brick-shaped structure consisting of six long PPII helices [61]. We found that PPII helical bundles have a signature set of ¹³C α , ¹³C β , ¹³CO and ¹H α chemical shift deviations [18]. Most remarkably, when a glycine residue is at an internal position, one of its two ¹H α nuclei shows a highly anomalous chemical shift value. The formation of weak C α -H||O=C hydrogen bonds by these ¹H nuclei can account for this extraordinary value.

Through H/D exchange and {¹H}-¹⁵N dynamic measurements, we found that sfAFP has a conformational stability and rigidity similar to those of globular proteins composed of α -helices and β -sheets. This is perplexing if we consider that glycine residues enjoy a high conformational entropy in the unfolded state, and like hippies called to join a military formation, are loath to give up their freedom and become fixed in a folded protein [63]. Stabilizing contributions from disulfide bonds, backbone N-H||O=C and C α -H||O=C hydrogen bonds and hydrophobic interactions at a dimer interface [18,64], as well as n \rightarrow π^* interactions [65], are key to overcoming this entropy effect. These findings should facilitate the use of PPII helical bundles as a new structural element for protein design, which has made remarkable achievements in recent years despite being limited to α -helices and β -sheets [66]. Finally, whereas the form of sfAFP with six PPII helices is the most abundant, natural samples also contain a longer isoform composed of 13 PPII helices with a similar, though longer disposition [51,58].

This suggests that two PPII helical bundles composed of several PPII helices may combine at the ends to form a larger PPII helical bundle. On a speculative note, such a (PPII)₆ + (PPII)₆ association event might also contribute to condensate formation.

Conclusions

Biomolecular condensates play essential roles in multiple physiological functions such as ribosome synthesis, transcription, splicing, stress response, RNA transport and translation regulation. On the other hand, they are also implicated in cancer, neurodegenerative diseases and viral infections. Biomolecular condensate formation and dissociation are governed by several classes of weak, transient interactions, one of which could well be the weak binding of small domains to PPII helices. Since recent work has shown that PPII helices and helical bundles can be formed by polypeptides poor in proline, their role in condensate formation may well be larger than what has previously been recognized.

Acknowledgements

JO is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2018-026042-I) and a Leonardo Fellow from the BBVA Foundation (Grant Number BBM-TRA-0203). MM is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2019-026574-I) and a 'La Caixa' Foundation Junior Leader Fellow (LCR/BQ/PR19/11700003). This study was supported by Grant BTC-PID2019-109306RB-I00 (DVL) from the Spanish Ministry of Science and Innovation.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

MM, JO and DVL researched the literature and wrote the paper.

Data availability

There are no related papers.

References

- 1 Banani SF, Lee HO, Hyman AA and Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285–298.

- 2 Yoshida K, Sanada M, Shiraiishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M *et al.* (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69.
- 3 Bouchard JJ, Otero JH, Scott DC, Szulc E, Martin EW, Sabri NA, Granata D, Marzahn MR, Lindorff-Larsen K *et al.* (2018) Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol Cell* **72**, 19–36.e8.
- 4 Colombrita C, Zennaro E, Fallini C, Weber M, Sommacal A, Buratti E, Silani V and Ratti A (2009) TDP-43 is recruited to stress granules in conditions of oxidative insult. *J Neurochem* **111**, 1051–1061.
- 5 Wegmann S, Eftekharzadeh B, Tepper K, Zoltowska KM, Bennett RE, Dujardin S, Laskowski PR, MacKenzie D, Kamath T, Commins C *et al.* (2018) Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J* **37**, e98049.
- 6 Zhou Y, Su JM, Samuel CE and Ma D (2019) Measles virus forms inclusion bodies with properties of liquid organelles. *J Virol* **93**, e00948–19.
- 7 Simons K and Ikonen E (1997) Functional rafts in cell membranes. *Nature* **387**, 569–572.
- 8 Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoegge C, Gharakhani J, Jülicher F and Hyman AA (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732.
- 9 Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A and Parker R (2016) ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* **164**, 487–498.
- 10 Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV and Brangwynne CP (2016) Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697.
- 11 Riback JA, Zhu L, Ferrolino MC, Tolbert M, Mitrea DM, Sanders DW, Wei MT, Kriwacki RW and Brangwynne CP (2020) Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214.
- 12 Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnke M, Maharana S, Lemaitre R, Pozniakovskiy A, Drechsel D *et al.* (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.
- 13 Conicella AE, Zerze GH, Mittal J and Fawzi NL (2016) ALS mutations disrupt phase separation mediated by α -helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549.
- 14 Wang A, Conicella AE, Schmidt HB, Martin EW, Rhoads SN, Reeb AN, Nourse A, Ramirez Montero D, Ryan VH, Rohatgi R *et al.* (2018) A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J* **37**, e97452.
- 15 Mitrea DM, Cika JA, Guy CS, Ban D, Banerjee PR, Stanley CB, Nourse A, Deniz AA and Kriwacki RW (2016) Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife* **5**, e13571.
- 16 Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J and Fawzi NL (2019) Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nature Struct Mol Biol* **26**, 637–648.
- 17 Amaya J, Ryan VH and Fawzi NL (2018) The SH3 domain of Fyn kinase interacts with and induces liquid-liquid phase separation of the low-complexity domain of hnRNPA2. *J Biol Chem* **293**, 19522–19531.
- 18 Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M and Laurents DV (2018) The singular NMR fingerprint of a polyproline II helical nundle. *J Am Chem Soc* **140**, 16988–17000.
- 19 Mompeán M, McAvan BS, Félix SS, Treviño M, Oroz J, Pantoja-Uceda D, Cabrita EJ, Doig AJ, Laurents DV *et al.* (2020) Glycine rich segments adopt polyproline II helices which may contribute to biomolecular condensate formation. *bioRxiv*. <https://doi.org/10.1101/2020.07.30.229062> (in press at Arch. Biochem. Biophys.) “[PREPRINT]”
- 20 Cowan PM, McGavin S and North ACT (1955) The polypeptide chain configuration of collagen. *Nature* **176**, 1062–1064.
- 21 Tiffany ML and Krimm S (1968) New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6**, 1379–1382.
- 22 Warkentin E, Weidenweber S, Schühle K, Demmer U, Heider J and Ermler U (2017) A rare polyglycine type II-like helix motif in naturally occurring proteins. *Proteins* **85**, 2017–2023.
- 23 Adzhubei AA and Sternberg MJ (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* **229**, 472–493.
- 24 Stapley BJ and Creamer TP (1998) A survey of left handed polyproline II helices. *Prot Sci* **8**, 587–595.
- 25 Rath A, Davidson AR and Deber CM (2005) The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers* **80**, 179–185.
- 26 Dvir H, Harel M, Bon S, Liu WQ, Vidal M, Garbay C, Sussman JL, Massoulié J and Silman I (2004) The synaptic acetylcholinesterase tetramer assembles around a polyproline II helix. *EMBO J* **23**, 4394–4405.
- 27 Cramer P, Bushnell DA and Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.5 angstrom resolution. *Science* **292**, 1863–1875.

- 28 Bienkiewicz EA, Moon Woody A and Woody RW (2000) Conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments. *J Mol Biol* **297**, 119–133.
- 29 Kornberg RD (2005) Mediator and the mechanism of transcriptional activation. *Trends in Biochem Sci* **30**, 235–239.
- 30 Robinson PJJ, Bushnell DA, Trnka MJ, Burlingame AL and Kornberg RD (2012) Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci USA* **109**, 17931–17935.
- 31 Hnisz D, Shrinivas K, Young RA, Chakraborty AK and Sharp PA (2017) A phase separation model for transcriptional control. *Cell* **169**, 13–23.
- 32 Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille JH, Afeyan LK, Zamudio AV, Shrinivas K *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548.
- 33 Verdecia MA, Bowman ME, Lu KP, Hunter T and Noel JP (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol* **7**, 639–643.
- 34 Gabrega C, Shen V, Shuman S and Lima CD (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol Cell* **11**, 1549–1561.
- 35 Ramdzan YM, Trubetskoy MM, Ormsby AR, Newcombe EA, Sui X, Tobin MJ, Bongiovanni MN, Gras SL, Dewson G, Miller J *et al.* (2017) Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis. *Cell Rep* **19**, 919–927.
- 36 Falk AS, Bravo-Arredondo JM, Varkey J, Pacheco S, Langen R and Siemer AB (2020) Structural model of the proline-rich domain of huntingtin exon-1 fibrils. *Biophys J* **119**, 2019–2028.
- 37 Shao J, Welch WJ, Diprospero NA and Diamond MI (2008) Phosphorylation of profilin by ROCK1 regulates polyglutamine aggregation. *Mol Cell Biol* **28**, 5196–5208.
- 38 Posey AE, Ruff KM, Harmon TS, Crick SL, Li A, Diamond MI and Pappu RV (2018) Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers. *J Biol Chem* **293**, 3734–3746.
- 39 Michaelsen-Preusse K, Zessin S, Grigoryan G, Scharkowski F, Feuge J, Remus A and Korte M (2016) Neuronal profilins in health and disease: relevance for spine plasticity and Fragile X syndrome. *Proc Natl Acad Sci USA* **113**, 3365–3370.
- 40 Fukazawa Y, Saitoh Y, Ozawa F, Ohta Y, Mizuno K and Inokuchi K (2003) Hippocampal LTP is accompanied by enhanced F-actin content within the dendritic spine that is essential for late LTP maintenance *in vivo*. *Neuron* **38**, 447–460.
- 41 Star EN, Kwiatkowski DJ and Murthy VN (2002) Rapid turnover of actin in dendritic spines and its regulation by activity. *Nat Neurosci* **5**, 239–246.
- 42 Si K, Lindquist S and Kandel ER (2003) A neuronal isoform of the aplysia CPEB has prion-like properties. *Cell* **115**, 879–891.
- 43 Ford L, Ling E, Kandel ER and Fioriti L (2019) CPEB3 inhibits translation of mRNA targets by localizing them to P bodies. *Proc Natl Acad Sci USA* **116**, 18078–18087.
- 44 Hervas R, Rau MJ, Park Y, Zhang W, Murzin AG, Fitzpatrick J, Scheres S and Si K (2020) Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in *Drosophila*. *Science* **367**, 1230–1234.
- 45 Ramírez de Mingo D, Pantoja-Uceda D, Hervás R, Carrión Vázquez M and Laurents DV (2020) Preferred conformations in the intrinsically disordered region of human CPEB3 explain its role in memory consolidation. *bioRxiv*. <https://doi.org/10.1101/2020.05.12.091587>. “[PREPRINT]”
- 46 Kurochkina N and Guha U (2013) SH3 domains: modules of protein-protein interactions. *Biophys Rev* **5**, 29–39.
- 47 Li P, Banjade S, Cheng H-C, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF *et al.* (2012) Phase transitions in the assembly of multivalent signaling proteins. *Nature* **483**, 336–340.
- 48 Kang H, Freund C, Duke-Cohan JS, Musacchio A, Wagner G and Rudd CE (2000) SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55. *EMBO J* **19**, 2889–2899.
- 49 White R, Gonsior C, Krämer-Albers EM, Stöhr N, Hüttelmaier S and Trotter J (2008) Activation of oligodendroglial Fyn kinase enhances translation of mRNA transported in hnRNP A2-dependent RNA granules. *J Cell Biol* **181**, 579–586.
- 50 Crick FH and Rich A (1955) Structure of polyglycine II. *Nature* **176**, 780–781.
- 51 Ramachandran GN, Sasisekharan V and Ramakrishnan C (1966) Molecular structure of polyglycine II. *Biochim Biophys Acta* **112**, 168–170.
- 52 Schühle K and Heider J (2012) Acetone and butanone metabolism of the denitrifying bacterium "*Aromatoleum aromaticum*" demonstrates novel biochemical properties of an ATP-dependent aliphatic ketone carboxylase. *J Bacteriol* **194**, 131–141.
- 53 Buglino J, Shen V, Hakimian P and Lima CD (2002) Structural and biochemical analysis of the Obg GTP binding protein. *Structure* **10**, 1581–1592.
- 54 Feng B, Mandava CS, Guo Q, Wang J, Cao W, Li N, Zhang Y, Zhang Y, Wang Z, Wu J *et al.* (2014)

- Structural and functional insights into the mode of action of a universally conserved Obg GTPase. *PLoS Biol* **12**, e1001866.
- 55 Dunne M, Denyes JM, Arndt H, Loessner MJ, Leiman PG and Klumpp J (2018) Salmonella phage S16 tail fiber adhesin features a rare polyglycine rich domain for host recognition. *Structure* **26**, 1573–1582.
- 56 Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH and Prangishvili D (2008) Exploring the prokaryotic virosphere. *Res Microbiol* **159**, 306–313.
- 57 Graham LA and Davies PL (2005) Glycine-rich antifreeze proteins from snow fleas. *Science* **310**, 461.
- 58 Bissoyi A, Reicher N, Chasnitsky M, Arad S, Koop T, Rudich Y and Braslavsky I (2019) Ice nucleation properties of ice binding proteins from snow fleas. *Biomolecules* **9**, 532.
- 59 Graham LA, Boddington ME, Holmstrup M and Davies PL (2020) Antifreeze protein complements cryoprotective dehydration in the freeze-avoiding springtail *Megaphragura arctica*. *Sci Rep* **10**, 3047.
- 60 Scholl CL, Tsuda S, Graham LA and Davies PL (2021) Crystal waters on the nine polyproline type II helical bundle springtail antifreeze protein from *Granisotoma rainieri* match the ice lattice. *FEBS J* **288**, 4332–4347.
- 61 Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA and Kent SB (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J Am Chem Soc* **130**, 9695–9701.
- 62 Singh RS, Palmer JC, Pudney PD, Paul PK, Johannessen C, Debenedetti PG, Raut J, Lee K, Noro M and Tiemessen D (2017) Molecular modeling and structural characterization of a high glycine-tyrosine hair keratin associated protein. *Phys Chem Chem Phys* **19**, 8575–8583.
- 63 D'Aquino JA, Gómez J, Hilser VJ, Lee KH, Amzel LM and Freire E (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25**, 143–156.
- 64 Gates ZP, Baxa MC, Yu W, Riback JA, Li H, Roux B, Kent SB and Sosnick TR (2017) Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. *Proc Natl Acad Sci USA* **114**, 2241–2246.
- 65 Newberry RW and Raines RT (2017) The $n \rightarrow \pi^*$ interaction. *Acc Chem Res* **50**, 1838–1846.
- 66 Baker D (2019) What has *de novo* protein design taught us about protein folding and biophysics? *Prot Sci* **28**, 678–683.
- 67 Mok YF, Lin FH, Graham LA, Celik Y, Braslavsky I and Davies PL (2010) Structural basis for the superior activity of the large isoform of snow flea antifreeze protein. *Biochemistry* **49**, 2593–2603.
- 68 Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi YL, Stauffacher C, Strominger JL and Wiley DC (1994) Complex of the human MHC class II glycoprotein HLA-DR1 and the bacterial superantigen SEB. *Nature* **368**, 711–718.
- 69 Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Wiley DC (1996) Crystallographic analysis of endogenous peptides associated with Hla-Dr1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci USA* **93**, 734.
- 70 Freund C, Dötsch V, Nishizawa K, Reinherz EL and Wagner G (1999) The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat Struct Biol* **6**, 656–660.
- 71 Freund C, Kuhne R, Yang H, Park S, Reinherz EL and Wagner G (2002) Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J* **21**, 5985–5995.
- 72 Fedorov AA, Fedorov E, Gertler F and Almo SC (1999) Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. *Nat Struct Biol* **6**, 661–665.
- 73 Pornillos O, Alam SL, Rich RL, Myszka DG, Davis DR and Sundquist WI (2002) Structural and functional interactions of the Tsg101 UEV domain. *EMBO J* **21**, 2397–2406.
- 74 Pires JR, Taha-Nejad F, Toepert F, Ast T, Hoffmüller U, Schneider-Mergener J, Kühne R, Macias MJ and Oschkinat H (2001) Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J Mol Biol* **314**, 1147–1156.
- 75 Martin BT, Serrano P, Geralt M and Wüthrich K (2016) Nuclear magnetic resonance structure of a novel globular domain in RBM10 containing OCRE, Octamer Repeat Sequence Motif. *Structure* **24**, 158–164.
- 76 Mourão A, Bonnal S, Soni K, Warner L, Bordonné R, Valcárcel J and Sattler M (2016) Structural basis for the recognition of spliceosomal SmN/B/B' proteins by the RBM5 OCE domain in splicing regulation. *eLife* **5**, e14707.
- 77 Feng S, Chen JK, Yu H, Simon JA and Schreiber SL (1994) Two binding orientations for peptides to src sh3 domain: development of a general model for sh3-ligand interactions. *Science* **266**, 1241–1247.
- 78 Mahoney NM, Rozwarski DA, Fedorov E, Fedorov AA and Almo SC (1999) Profilin binds proline-rich ligands in two distinct amide backbone orientations. *Nat Struct Biol* **6**, 666–671.