

<https://doi.org/10.1038/s42004-024-01268-2>

Hydrogen bonding patterns and cooperativity in polyproline II helical bundles

Check for updates

Rubén López-Sánchez, Douglas V. Laurents & Miguel Mompeán

Hydrogen bond cooperativity (HBC) plays an important role in stabilizing protein assemblies built by α -helices and β -sheets, the most common secondary structures. However, whether HBC exists in other types of protein secondary structures such as polyproline II (PPII) helices remains unexplored. This is intriguing, since PPII systems as assembling blocks are continuously emerging across multiple fields. Here, using a combination of computational chemistry tools and molecular modeling corroborated by experimental observables, we characterize the distinct H-bonding patterns present in PPII helical bundles and establish that HBC stabilizes intermolecular PPII helices as seen in other protein assemblies such as amyloid fibrils. In addition to cooperative interactions in canonical CO \cdots HN H-bonds, we show that analogous interactions in non-canonical CO \cdots H $_{\alpha}$ C $_{\alpha}$ H-bonds are relevant in Gly-rich PPII bundles, thus compensating for the inability of glycine residues to create hydrophobic cores. Our results provide a mechanistic explanation for the assembly of these bundles.

The last decade has witnessed remarkable advances in the design of proteins with unnatural folds, functions, and enzymatic activities¹. These proteins are built of α -helices and β -sheets, which are the two main classes of secondary structure in globular proteins. However, a third class of protein secondary structure, the polyproline II (PPII) helix, has been rarely utilized in protein design^{2,3}. A triple PPII helix is the basis of collagen, and there are several proteins and protein domains which are composed of bundles of Gly-rich PPII helices⁴. The abundance of collagen and the diversity of Gly-rich PPII helical bundle proteins underscores the potential usefulness of this structural element as a building block for protein design. In this regard, recent applications have exploited the extended conformation of PPII helices and their assembling potential to develop copolymers with distinct applications in different fields such as biomedicine and biotechnology⁵. Still, progress is slowed by our limited understanding of the bases of PPII helical bundle assembly and conformational stability. To address this gap in our comprehension, we and others have studied the *Hypogastrura harveyi* “snow flea” antifreeze protein (*HhAFP*) as a model system for Gly-rich PPII helical bundles. The structure of the *HhAFP* consists of a bundle of six PPII helices organized into two sheets oriented antiparallel to each other, each sheet containing three parallel PPII helices^{6,7}. The intramolecular nature of the *HhAFP* PPII bundle facilitated high-resolution structural and dynamical studies, which showed that despite a high content in glycine residues, and an apparent lack of a stable hydrophobic core^{6,8}, the conformational stability of the intramolecular PPII bundle is similar to typically globular proteins. This observation was explained by a number of distinct contributions such as the

two disulfide bonds, canonical backbone CO \cdots HN H-bonds as well as twenty-eight non-canonical backbone CO \cdots H $_{\alpha}$ C $_{\alpha}$ H-bonds in accordance with experimental^{7,8} and computational reports⁷, and could be additionally stabilized by n \rightarrow π^* interactions^{9,10}. Moreover, for non-glycine residues the presence of unconventional intramolecular interactions such as N-H \cdots H-C(sp³), π \cdots H-C(sp³) or CO \cdots H-C(sp²) between the main chain and hydrophobic side chains, both aliphatic and aromatic, has been reported in PPII helix-forming peptides¹¹. All these energetic factors may also stabilize other Gly-rich PPII bundles containing a larger number of PPII helices, such as the *Granisotoma rainieri* “springtail” antifreeze protein (*GrAFP*)¹², the larger isoform of the *HhAFP*¹³ or the anaplastic lymphoma kinase (ALK)¹⁴ receptor. The PPII bundles assembled by these proteins contain as many as nine, thirteen, and fourteen PPII helices, respectively, as revealed in the crystallographic structures^{12,14} or homology models¹³. The unfolded ensemble of the *HhAFP* short isoform is compact and contains nonlocal electrostatic interactions which might initiate folding¹⁵. The cooperative nature of this protein’s folding was established by kinetic studies⁸ and recent protein engineering studies have revealed that activity increases with the number of PPII helices¹⁶. These results are consistent with, but do not prove, hydrogen bond cooperativity (HBC) in PPII helical bundles.

The unexpectedly high conformational stability and the large number of H-bonds that are present in PPII bundles, as seen in the *HhAFP* fold, are reminiscent of amyloids, which are quasi-infinite β -sheets with networks of aligned H-bonds. In both PPII helices and β -sheet-rich amyloids, the polypeptide chain extends facilitating self-assembly. One critical

Instituto de Química Física “Blas Cabrera” - CSIC, Madrid, Spain. e-mail: dlaurents@iqf.csic.es; mmompean@iqf.csic.es

contribution to the remarkably high conformational stabilities of amyloids are cooperative H-bonding interactions, in which the alignment of CO...HN backbone H-bonding groups leads to increased polarization and strength^{17–19}. We, among others, have reported that these HBC effects are also present in Asn and Gln residues, whose amide side chains form strongly hyperpolarized H-bonds^{20–22}. Interestingly, in the PPII helices of Gly-rich proteins featuring PPII helical bundles such as the *HhAFP*, networks of aligned, canonical CO...HN H-bonds which resemble those observed in amyloids are present across the helix-helix interfaces, suggesting that HBC might also occur in these systems. In fact, HBC was previously suggested as the reason why polyglycine exhibited stronger CO...HN H-bonds when grouped in polyglycine II (PGII)-type structures, which do align multiple helices, versus collagen triple helices, which do not²³. In addition, a number of aligned non-canonical CO...H_αC_α H-bonds are also present at the helix-helix interfaces in PPII helical bundles, providing an additional source of stability^{7,8}. These CO...H_αC_α H-bonds between PPII helices have been long known in proteins like collagen²⁴ and PGII assemblies^{25,26}, and although their cooperative nature has not been established, CO...HC cooperative H-bonding interactions have been reported in model systems, such as in (HFCO)_n clusters and other carbonyl compounds^{27,28}, or in theoretical models to study the interactions between water and DMSO²⁹.

In this study, we have investigated the nature of H-bonding interactions in the context of both canonical CO...HN and non-canonical CO...H_αC_α H-bonds across helix-helix interfaces in Gly-rich PPII helical bundles. Molecular models of PPII-based assemblies based on NMR observables were built to afford a detailed characterization of all H-bonding interactions in these assemblies. By means of four independent approaches: Dispersion-corrected Density Functional Theory (DFT-D), Natural Bonding Orbital (NBO), Quantum Theory of Atoms in Molecules (QTAIM), and Symmetry-Adapted Perturbation Theory (SAPT) calculations, we provide a detailed picture of an energetic kernel that drives PPII

bundle assembly. To date, NBO methodology has served to demonstrate cooperative effects between canonical CO...HN H-bonds in amides and peptides describing other types of secondary structures^{30,31}, but not in PPII helices, nor between non-canonical CO...H_αC_α H-bonds. Regarding QTAIM, whereas both canonical and non-canonical H-bonds of the two most common secondary structures have already been studied^{32,33}, their cooperativity is not and there is no information for PPII helices. In fact, H-bonding cooperativity has only been studied using QTAIM analysis for small molecules^{34,35}. Finally, SAPT allows us to study the nature of non-covalent interactions³⁶, and although HBC has been studied in protein ligands using this methodology, it has not been applied to HBC between the different secondary structures to the best of our knowledge³⁷.

Results

Model system to study Gly-Rich PPII bundles

The experimentally resolved 3D structure of the short isoform of the *HhAFP* consists of six PPII helices, where helices I, III, and V constitute a polar face while helices II, IV, and VI form the non-polar, ice-binding surface (Fig. 1A). The large *HhAFP* isoform is predicted by AlphaFold2 (AF2) with a similar fold containing thirteen instead of six PPII helices (Fig. 1A), which is in agreement with the model proposed by Mok et al.¹³. The *GrAFP* structure also matches this bundle arrangement (Fig. 1A), with five parallel helices in one face packing antiparallel with a four-helical face¹². The dearth of side chains in these Gly-rich sequences underlines the importance of H-bonding for bundle stabilization in these and in other Gly-rich PPII bundles such as in the recently elucidated ALK structure¹⁴ (Fig. 1A).

In order to investigate the nature of H-bonding interactions across the different PPII bundles described above, we have designed a model system that recreates the interactions seen in the various assemblies. This model relies on replicating the PPII helical stretch spanning residues 31–39 from

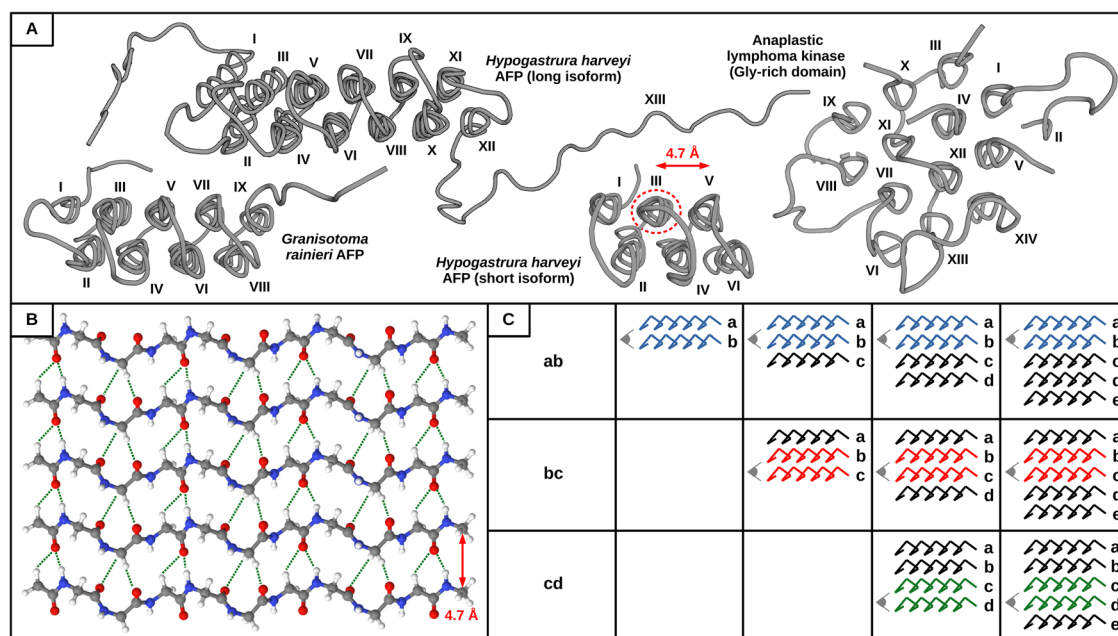


Fig. 1 | H-bonding in polyproline II helical bundle domains. **A** Experimentally determined structures from the Protein Data Bank (PDB) for *HhAFP* short isoform (PDB ID: 3BOG), *GrAFP* (PDB ID: 7JJV) and ALK Gly-rich domain (PDB ID: 7LS0/7LRZ), and predicted structure from AF2 for *HhAFP* long isoform (AF2 ID: D7PBP2). Roman numerals denote the distinct helices in each protein. The dashed red circle indicates the PPII helix cloned for the computational model. **B** Gly-rich PPII helical bundle computational model: red spheres represent oxygen; blue, nitrogen; gray, carbon, and white, hydrogen. Shown with a green dashed line are the potential hydrogen bonds that may exist based on the proximity between the

hydrogen donor (i.e., N or C) and hydrogen acceptor (i.e., O) atoms. **C** Scheme utilized for the calculation of the different parameters between PPII helices in the different interfaces. Blue PPII helices give rise to the *ab* interface; the red ones, to the *bc* interface, and the green ones, to the *cd* interface. The interaction energies are probed at these colored interfaces, i.e., where the *n*-mers are split for the HBS calculations, both DFT-D and SAPT. The rest of the computational parameters are calculated directly between the colored PPII helices through the NBO and QTAIM analyses. For added clarity, the gray eye symbol indicates the interface under study for each case.

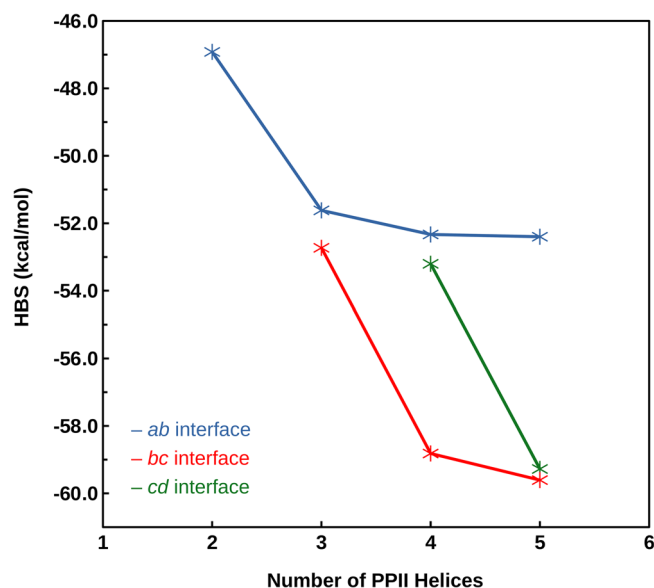


Fig. 2 | HBS between PPII helices in the different interfaces of the computational model at the M06-2X/6-31 + G(d) level of theory. The *ab* interface is indicated in blue; the *bc* interface, in red, and the *cd* interface, in yellow, as the color scheme depicted in Fig. 1C. Note that the HBS values decrease significantly as more PPII helices are added; this is a hallmark of HBC.

the central PPII helix in the polar face (i.e., helix III, as depicted in Fig. 1A) of the short *HhAFP* isoform. These nine residues correspond to three turns of the PPII helix that are properly aligned to form H-bonds with the helices above (helix I) and below (helix V), which lie on the same face and are spaced by 4.7 Å (Fig. 1A). The nine-residue stretch corresponding to residues 31–39 was cloned (i.e., repeated in space) four times to build a five-helical bundle. This leads to five monomeric, nine-residue helices lying within the same plane that are spaced by 4.7 Å from each other. The N- and C-termini of each PPII helix were capped with acetyl and methylamine moieties, respectively, to recreate the sequence continuity beyond the three turns that form the stretch. To isolate the pure PPII architecture from any other driving force for assembly arising from hydrophobic, polar or charged side chains, as well as to generalize the model for all types of Gly-rich PPII bundles, as those shown in Fig. 1A, all non-glycine residues, i.e., H32, N35, and N38, were mutated to glycine. Therefore, each PPII helix will contain 41 heavy atoms, giving rise to a total of 205 for the complete bundle (Fig. 1B). In order to analyze the interactions inherent to these structures, no type of solvent is included in the computational model.

Cooperative bundle assembly of PPII helices and H-Bonding patterns

Interaction energy calculations reveal enhanced H-Bonding strength and cooperativity in growing PPII helical bundles. Building on the general model representing Gly-rich PPII bundles elaborated in the previous section, we next sought to characterize the interaction energy between every pair of PPII helices in such an assembly. In this model system, H-bonds drive PPII bundle assembly as there are no nonpolar, charged or aromatic groups which would give rise to hydrophobic, charge-charge or cation- π interactions (Fig. 1B). Thus, the interaction energy between two monomeric helices provides a direct estimation of the H-bonding strength (HBS). On this basis, we employed DFT-D to compute the interaction energies for dimers ($a\cdots b$), trimers ($a\cdots bc$ and $ab\cdots c$), tetramers ($a\cdots bcd$, $ab\cdots cd$ and $abc\cdots d$) and pentamers ($a\cdots bcde$, $ab\cdots cde$, $abc\cdots de$, $abcd\cdots e$) of PPII helices, according to the scheme shown in Fig. 1C. Note that for an *n*-mer system, there are (*n*-1) interacting interfaces and thus (*n*-1) HBS possible values. The only interaction that will not be analyzed is $abcd\cdots e$ since, as *de* is the last

Table 1 | HBS between PPII helices in the different interfaces of the computational model at the M06-2X/6-31 + G(d) level of theory

HBS (kcal·mol ⁻¹)	Interface	Number of PPII Helices			
		2	3	4	5
PPII helical bundle	<i>ab</i>	-46.92	-51.62	-52.33	-52.40
	<i>bc</i>	-	-52.73	-58.82	-59.61
	<i>cd</i>	-	-	-53.20	-59.28

interface and *e* is the last PPII helix, the limit of the model has been reached and it is not possible to study how the addition of monomers affects the different computational parameters. The analysis of the computed HBS values shows that as PPII helices are added to the assembly, the interaction energies between monomers are strengthened. That is, the HBS between helices *a* and *b* in a five-helix bundle, $abcde$, is stronger than in a four-helix bundle, $abcd$, which is stronger than in a three-helix bundle, abc , which in turn is stronger than in a two-helix bundle ab , (i.e., $|HBS(a\cdots b)_{abcde}| > |HBS(a\cdots b)_{abcd}| > |HBS(a\cdots b)_{abc}| > |HBS(a\cdots b)_{ab}|$, Fig. 2 and Table 1). In this regard, we recall that we measure the interaction energy between the first two PPII helices (called *a* and *b*) as more and more helices are added on top of them, and therefore it is not the total interaction energy of the whole assembly, but rather that between the first two PPII helices what is being quantified. The same holds true for the HBS in the rest of the interfaces as helices are added to the assembly (i.e., $|HBS(b\cdots c)_{abcde}| > |HBS(b\cdots c)_{abcd}| > |HBS(b\cdots c)_{abc}|$ and $|HBS(c\cdots d)_{abcde}| > |HBS(c\cdots d)_{abcd}|$, Fig. 2 and Table 1). This is proof of cooperative H-bonding interactions or HBC, as seen in amyloids^{17–22}. To uncover the source of this cooperativity, we employed various computational methods to analyze the underlying H-bonding Interactions as described in the following sections.

Identification of canonical and non-canonical H-Bonding patterns stabilizing Gly-Rich PPII helical bundles.

Having established that H-bonding in Gly-rich PPII assembly is cooperative, akin to amyloid assembly, we next ought to characterize which H-bonding interactions are present. We started by identifying the distinct potential stabilizing interactions by visual inspection of our model (Fig. 1B). For a PPII helix composed of three residues *i*, *i* + 1 and *i* + 2, which corresponds to exactly one turn of PPII helix, the carbonyl oxygen of the first residue, *i*, from one PPII helix interacts with the amide proton H_N of the second residue, *j* + 1, in the next PPII helix to form a canonical $CO\cdots HN$ H-bond (Fig. 3A). In addition to this canonical interaction, we also identified the possible participation of the carbonyl oxygen from residue *i* in a non-canonical $CO\cdots H_\alpha C_\alpha$ interaction with the H_α from residue *j* in the second PPII helix. That is, the carbonyl oxygen from the first residue *i* in one helix participates in both the formation of a non-canonical H-bond with the H_α from the first residue *j* in the mating helix, and a canonical H-bond with the H_N from the second residue *j* + 1. This non-canonical $CO\cdots H_\alpha C_\alpha$ interaction is denoted as a “front” H-bond (since it would be found in the same plane as each face of *HhAFP*) (Fig. 3A). Furthermore, we advance the existence of two additional non-canonical $CO\cdots H_\alpha C_\alpha$ H-bonds in Gly-rich PPII bundles that we here have denoted as an “inner” (since it would be found inside the core of *HhAFP*) and “outer” (since it would be found outside the core of *HhAFP*) (Fig. 3A). The supposed inner and outer H-bonds involve the two H_α protons of the last (i.e., third) residue in one PPII helix (i.e., residue *i* + 2). Particularly, in the “inner” H-bond one H_α proton would interact with the carbonyl oxygen from residue *j* + 2, whereas in the “outer” H-bond the second H_α proton from residue *i* + 2 would interact with the carbonyl oxygen from the second residue in the mating helix (i.e., *j* + 1, Fig. 3A). Since the results based on the HBS between PPII helices do not allow one to discern whether cooperativity holds every type of H-bond considered, we resorted to NBO and QTAIM analyses.

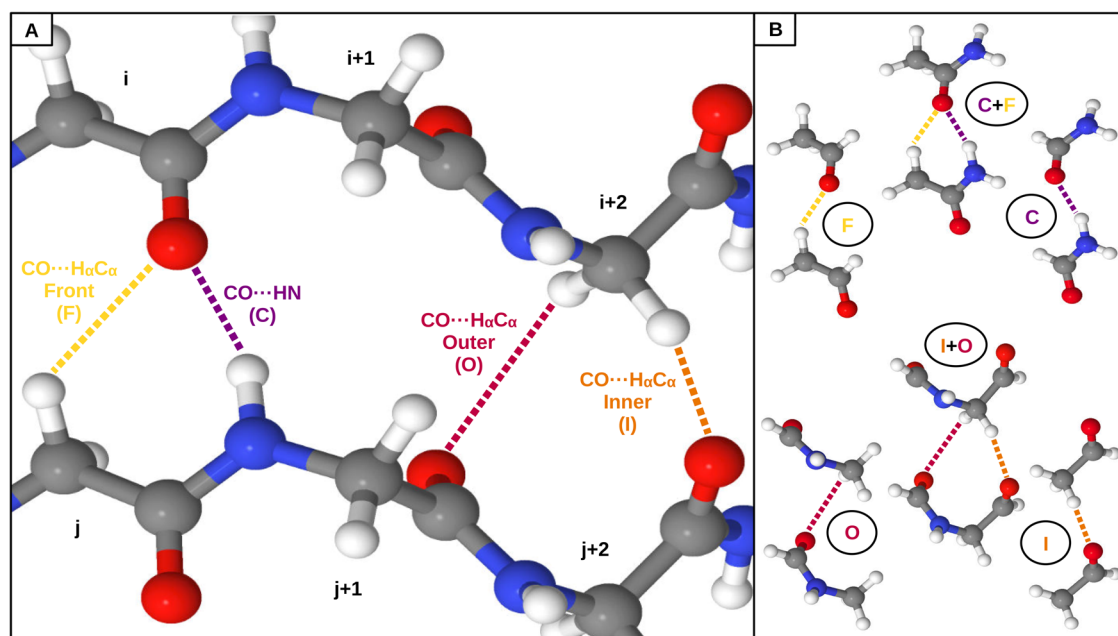


Fig. 3 | H-bonds in polyproline II helical bundle assemblies. A Possible H-bonds/interactions between PPII helices in a 2D PPII helical bundle: CO...HN (purple), CO...H α C α front (yellow), CO...H α C α inner (orange) and CO...H α C α outer (magenta). **B** Simplified structures for CO...HN (formamide) and CO...H α C α front

(acetamide), CO...HN + CO...H α C α front (acetamide), CO...H α C α inner (acet-aldehyde), CO...H α C α outer (N-methylformamide) and CO...(H α) $_2$ C α inner+outer (N-(2-oxoethyl)formamide) H-bonds/interactions.

NBO and QTAIM analyses independently corroborate the existence and HBC of N-H...O = C and C α -H α ...O = C inner H-bonds. Regarding the NBO analysis (Tables 2 and 3 and Fig. 4), the occupancy of the σ^* orbitals of the N-H bonds, $q(\sigma^*_{\text{NH}})$, and the stabilization energy corresponding to the electronic delocalization from not only the lone pairs (n) of the oxygen in those orbitals, but also from the σ and π orbitals of the C = O bonds, $E(2)_{\text{n(O)}/\sigma(\text{CO})/\pi(\text{CO}) \rightarrow \sigma^*(\text{NH})}$, indicate the progressive strengthening of the CO...HN H-bonds as the PPII helical assembly grows. This electron donation confers partial anionic character on the electron-acceptor species, then enhancing the capacity of the CONH units to act as an electron-donor in the next H-bond, which will be stronger. As for the CO...H α C α inner H-bonds, the $q(\sigma^*_{\text{CaHa}})$ and the $E(2)_{\text{n(O)}/\sigma(\text{CO})/\pi(\text{CO}) \rightarrow \sigma^*(\text{CaHa})}$ increase as PPII helices are added to the bundle, in a homologous way to what happened with CO...HN H-bonds, also involving the $\sigma(\text{CO})$ and $\pi(\text{CO})$ orbitals (Tables 2 and 3 and Fig. 4). These results are supported by the QTAIM analysis (supplementary material).

Front H-bonds exist but show no HBC and outer H-bonds do not exist, according to NBO and QTAIM analysis. For the CO...H α C α front H-bonds, while $E(2)_{\text{n(O)}/\sigma^*(\text{CaHa})}$ slightly increases as PPII helices are added to the bundle, the $q(\sigma^*_{\text{CaHa}})$ decreases (Tables 2 and 3 and Fig. 4). This decrease of $q(\sigma^*_{\text{CaHa}})$ can be explained on the basis that the $q(\sigma^*_{\text{NH}})$ follows the opposite trend. Both types of σ^* orbitals receive electrons from the same oxygen, so if the amount of electrons reaching one of the two different σ^* orbitals increases as the assembly grows, the amount reaching the other σ^* orbital will decrease. This phenomenon occurs even if the CONH units, which also give rise to the canonical H-bonds, have more and more electrons available to form the new H-bonds of the next interfaces, which is the reason why $E(2)_{\text{n(O)}/\sigma^*(\text{CaHa})}$ slightly increases as PPII helices are added to the assembly, since it depends on $q(\text{nO})$ and not on $q(\sigma^*_{\text{CaHa}})$ (Eq. 3). Finally, for the CO...H α C α outer interactions, the $q(\sigma^*_{\text{CaHa}})$ values remain practically constant as the PPII helical bundle grows and the negligible $E(2)_{\text{n(O)}/\sigma^*(\text{CaHa})}$ values do not have a clear tendency to change (Tables 2 and 3 and Fig. 4), suggesting their non-existence and hence non-cooperativity. These results are supported by QTAIM analyses (supplementary material).

Contribution of the different H-bonds to PPII helical bundle stability

The presence of a large number of non-canonical CO...H α C α H-bonds is important for the stability of PPII helical bundles^{7,8}. In order to quantify the extent to which the distinct H-bonds contribute to the interaction energy between PPII helices, we have followed a divide-and-conquer approach by which mimicking molecular fragments resembling all types of H-bonds are subjected to independent HBS calculations^{38–40} (Fig. 3B). Within this framework, each CO...HN canonical H-bond is represented by a pair of formamide molecules. Front and inner non-canonical CO...H α C α H-bonds are mimicked by a pair of acetaldehyde molecules with different spatial orientations. Finally, outer non-canonical CO...H α C α interactions are mimicked using pairs of N-methylformamide molecules. It is necessary to emphasize that the mimetic molecules that give rise to the different interactions maintain the same position, and therefore the same distance, as in the complete PPII helices. In the following subsections, where we employ mimetics rather than full helices, we change the nomenclature from “PPII helices” to “layers of molecules” and “bundle assembly (of PPII helices) growth” for “stack (of layers of molecules) growth”. The term “interface” will now signify the region where mimetic molecules interact, rather than complete PPII helices.

Strengthened canonical N-H...O = C H-bonds versus electron-limited non-canonical C α -H α ...O = C front interactions in expanding stacks. Regarding canonical H-bonds, the H-bonding strength or HBS among the different interfaces between formamide moieties (Fig. 3B) is strengthened as the stack grows (Fig. 5 and Table 4). These observations are reminiscent of the complete model of PPII bundles, and evince the cooperative reinforcement in CO...HN H-bonds between PPII helices. In the same way, Fig. 5 and Table 4 also show that, when the stack of layers of the acetaldehyde molecules that mimic CO...H α C α front H-bonds (Fig. 3B) grows, the HBS of the different interfaces is strengthened. One could interpret this as evidence of the cooperative reinforcement between non-canonical front H-bonds, which is in conflict with the previous NBO and QTAIM analyses of the full-length system that seem to indicate that they do not show HBC because

Table 2 | Mean occupancy (q) of the corresponding σ^* orbitals for each CO...HN, CO...H_αC_α front, CO...H_αC_α inner and CO...H_αC_α outer H-bond/interaction between PPII helices in the different interfaces at the M06-2X/6-31 + G(d) level of theory

$q(\sigma^*)$ (10^{-2} electrons)	Interface	Number of PPII Helices			
		2	3	4	5
CO...HN	<i>ab</i>	3.590	3.601	3.607	3.609
	<i>bc</i>	-	4.154	4.188	4.196
	<i>cd</i>	-	-	4.226	4.255
CO...H _α C _α Front	<i>ab</i>	1.806	1.795	1.794	1.794
	<i>bc</i>	-	1.787	1.775	1.776
	<i>cd</i>	-	-	1.805	1.794
CO...H _α C _α Inner	<i>ab</i>	1.944	1.961	1.960	1.958
	<i>bc</i>	-	2.016	2.034	2.032
	<i>cd</i>	-	-	1.964	1.986
CO...H _α C _α Outer	<i>ab</i>	1.357	1.357	1.357	1.358
	<i>bc</i>	-	1.610	1.606	1.607
	<i>cd</i>	-	-	1.642	1.639

Table 3 | Mean stabilization energy ($E(2)$) of the corresponding $n(O)/\sigma(CO)/\pi(CO) \rightarrow \sigma^*(NH/C_{\alpha}H_{\alpha})$ electron delocalization for each CO...HN, CO...H_αC_α front, CO...H_αC_α inner and CO...H_αC_α outer H-bond/interaction between PPII helices in the different interfaces at the M06-2X/6-31 + G(d) level of theory

$E(2)$ (kcal·mol ⁻¹)	Interface	Number of PPII Helices			
		2	3	4	5
CO...HN	<i>ab</i>	16.350	16.558	16.590	16.605
	<i>bc</i>	-	21.525	21.793	21.833
	<i>cd</i>	-	-	21.548	21.825
CO...H _α C _α Front	<i>ab</i>	1.405	1.423	1.430	1.430
	<i>bc</i>	-	1.293	1.310	1.318
	<i>cd</i>	-	-	1.788	1.815
CO...H _α C _α Inner	<i>ab</i>	5.783	5.840	5.837	5.837
	<i>bc</i>	-	6.690	6.743	6.740
	<i>cd</i>	-	-	6.450	6.527
CO...H _α C _α Outer	<i>ab</i>	0.293	0.287	0.287	0.287
	<i>bc</i>	-	0.180	0.173	0.173
	<i>cd</i>	-	-	0.213	0.210

the carbonyl oxygen's electron density is shared with the much more favored canonical H-bonds. Perhaps the most intuitive way to visualize the effect that CO...HN H-bonds have on CO...H_αC_α front H-bonds, as well as to adequately quantify both interactions, is to calculate the HBS corresponding to acetamide molecules since they mimic both H-bonds at the same time (Fig. 3B). When the stack grows, the HBS of the different interfaces is reinforced, and so does the HBS of both the canonical and non-canonical front H-bonds separately, albeit less than the sum of both H-bonds isolated at each interface (Fig. 5 and Table 4). As explained earlier, this situation can be attributed to the limited quantity of electrons that oxygens may donate to the canonical and the non-canonical front H-bonds.

Strengthened non-canonical C_α-H_α...O=C inner H-bonds versus dubious non-canonical C_α-H_α...O=C outer interactions in expanding stacks. Concerning non-canonical CO...H_αC_α inner H-

bonds, while it is true that the HBS of the different interfaces between pairs of acetamide molecules that mimic these interactions (Fig. 3B) become stronger upon addition of layers to the stack, it does so over a range of energy values that might fall within the uncertainty of the computational level of theory (Fig. 5 and Table 4). The values of the HBS between the N-methylformamide molecules that represent the non-canonical CO...H_αC_α outer interactions (Fig. 3B) are null within the error of the calculation (Fig. 5 and Table 4). Furthermore, these negligible interaction energies corresponding to the different interfaces do not show any tendency to increase or decrease, nor any evidence of cooperativity therefore, as the stack of mimicking molecules grows. Moreover, it is necessary to take into account that since these two putative non-canonical H-bonds share a common C_α, one interaction cannot be isolated from the other to obtain the HBS that would actually correspond to them. Consequently, we used pairs of N-(2-oxoethyl)formamide molecules to adequately mimic both potential H-bonds together (Fig. 3B), which gave rise to considerable HBS values showing a notable strengthening of the interactions across the different interfaces as the stack grows (Fig. 5 and Table 4). Additional QTAIM and SAPT analyses and results are shown in Supplementary Figs. 1–10 and Supplementary Tables 1–9. These interaction energy values are due almost exclusively to non-canonical inner H-bonds, not to the combination with the apparently non-existent non-canonical outer H-bonds that were initially proposed, as the previous NBO and QTAIM analyses of the full PPII bundle model showed. The findings of this section are in agreement with the NBO and QTAIM analyses of the mimicking systems.

Summing interaction energy contributions in PPII helical bundles provides insight into electron delocalization effects. Since the set of different molecules of acetamide (i.e., CO...HN + H_αC_α front H-bonds) and N-(2-oxoethyl)formamide (i.e., CO...(H_α)₂C_α inner+outer H-bonds/interactions) shown in Fig. 3B are not a bona fide representation of the complete PPII bundle, it would not be expected that the sum of the different contributions to HBS would give the total HBS between full-length PPII helices. Indeed, Table 4 shows that the sum of the the CO...HN and the different CO...H_αC_α interaction energies produce HBS values initially smaller than the ones obtained when calculating the total HBS between full-length PPII helices shown in Table 1. However, as layers are added to the stack, the HBS values calculated as the sum of contributions exceed the global ones (Tables 4 and 1), demonstrating once again the presence of HBC in PPII helices and suggesting that in complete PPII helices, electron density would be delocalized not only in adjacent helices, but also along the peptide chain itself, generating macrodipoles previously reported for these protein secondary structures^{41,42}.

Summary of interactions

The computational parameters studied here are summarized in Table S10 as the mean variation for each of the proposed H-bonds of the system's five PPII helices. One should interpret this mean variation as the average of the differences between the computed parameter X for the last and first value for the interfaces ab ($X_{abcde} - X_{ab}$), bc ($X_{abcde} - X_{abc}$) and cd ($X_{abcde} - X_{abcd}$). The NBO and QTAIM analyses, as well as the DFT-D calculation of interaction energies (HBS), lead us to suggest the existence and cooperativity of the CO...HN and CO...H_αC_α inner H-bonds; the existence, but non-cooperativity, of the CO...H_αC_α front H-bonds, and the non-existence, and therefore non-cooperativity, of the CO...H_αC_α outer H-bonds. The distance between the H-bond acceptor, the carbonyl oxygen, and the H-bond donor (H_N or H_α) is another way to gauge the strengthening and cooperativity, or the lack, of these four types of possible H-bonds. In particular, non-canonical CO...H_αC_α outer H-bonds, whose existence cannot be established here, feature the longest distances (Table S11 and Fig. S11). Moreover, the shortest distances are seen for the CO...HN and CO...H_αC_α inner H-bonds and these distances decrease further as the bundle grows (Table S11 and Fig. S11), which is consistent with their HBC. In the case of non-canonical CO...H_αC_α

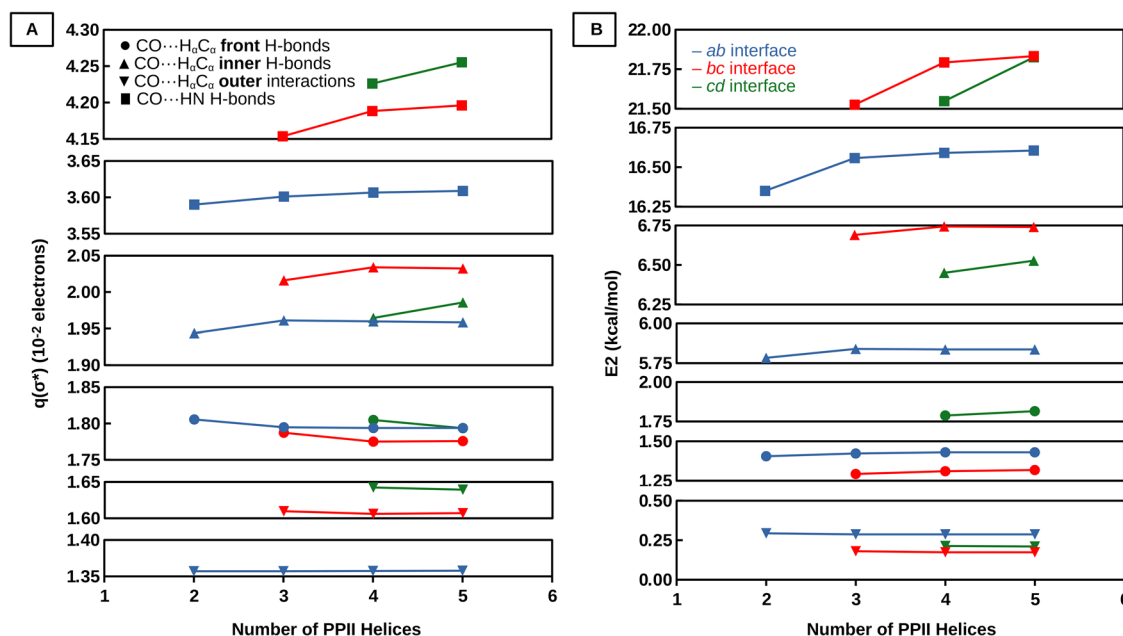


Fig. 4 | Orbital occupancy and stabilization energies evince HBC in PPII helical bundles. **A** Mean occupancy (q) of the corresponding $\sigma^*(\text{NH}/\text{C}_\alpha\text{H}_\alpha)$ orbitals and **B** mean stabilization energy ($E(2)$) of the corresponding $n(\text{O})/\sigma(\text{CO})/\pi(\text{CO}) \rightarrow \sigma^*(\text{NH}/\text{C}_\alpha\text{H}_\alpha)$ electron delocalization for each CO...HN (squares),

CO...H_αC_α front (circles), CO...H_αC_α inner (triangles) and CO...H_αC_α outer (inverted triangles) H-bond/interaction between PPII helices. The *ab* interface for the different systems is indicated in blue; the *bc* interface, in red, and the *cd* interface, in green. M06-2X/6-31 + G(d) level of theory.

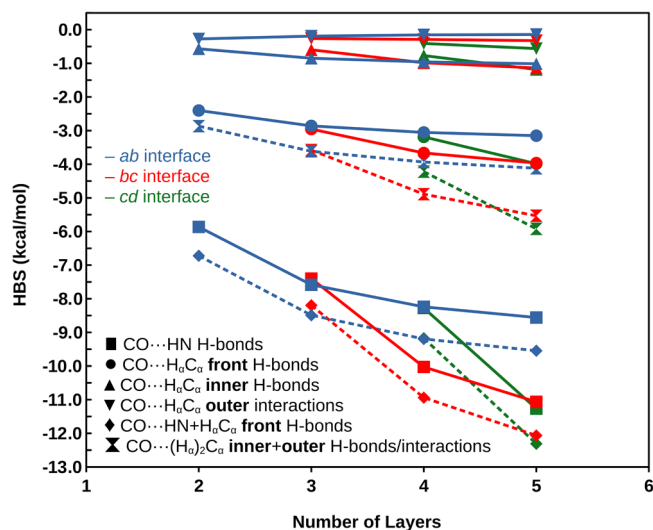


Fig. 5 | HBC in CO...HN and CO...Ha-Cα H-bonds. Mean HBS between the molecules that reproduce each CO...HN (formamide, squares), CO...H_αC_α front (acetaldehyde, circles), CO...HN + CO...H_αC_α front (acetamide, diamonds, and dashed line), CO...H_αC_α inner (acetaldehyde, triangles), CO...H_αC_α outer (N-methylformamide, inverted triangles) and CO...(H_α)₂C_α inner+outer (N-(2-oxoethyl)formamide, hourglass and dashed line) H-bond/interaction. As in the Fig. 1C color scheme, the *ab* interface for the different systems is indicated in blue; the *bc* interface, in red, and the *cd* interface, in green.

front H-bonds, intermediate H-bonding distances without any trend to increase or decrease as the assembly grows are obtained (Table S11 and Fig. S11); this is consistent with the lack of HBC seen in the NBO, QTAIM and DFT-D results.

SAPT calculations are somewhat more precise, much more computationally costly, and provide insight into the contributions of the distinct interaction energy components to HBC. Our SAPT results, obtained at a higher level of theory than the rest of the methods, confirm that HBC exists

in the cases already detected by DFT-D, NBO and QTAIM calculations, and evidence that it arises mostly from electrostatic effects and not induction or dispersion (Supplementary Tables S5, S6, S7 and Figs. S3, S5, S6, S7, S8, S9, S10). We refer the interested reader to the supplementary material for a specialized, in-depth description of the different interactions analyzed in this study.

Corroboration of the computational model

To further corroborate these findings, we have examined the experimental and computational NMR shift changes produced by H-bond formation. In solution, Gly-rich peptides have a tendency to adopt the extended PPII helix conformation over random coil or other structures^{43–45}. Nevertheless, these situations cannot be easily distinguished spectroscopically since the two ¹H_α chemical shifts of glycine residues are degenerate. However, we discovered that the two ¹H_α of glycine residues in PPII bundles do display different chemical shifts, providing the first set of conformational chemical shifts, $\Delta\delta$, for Gly-rich PPII bundles⁷. Experimental conformational shifts for ¹³C_α and ¹³CO of $\Delta\delta(\text{C}_\alpha)_{\text{exp}} = -0.59$ (± 0.65) ppm and $\Delta\delta(\text{CO})_{\text{exp}} = -0.20$ (± 1.12) ppm were obtained, while ¹H_α nuclei in CO...H_αC_α H-bonded feature a characteristic conformational shift of $\Delta\delta(\text{H}_\alpha)_{\text{exp}} = -0.59$ (± 0.32) ppm. In our computational model (Fig. 1B), conformational chemical shifts for these nuclei were predicted to test our model's ability to serve as a bona fide atomistic representation of PPII bundles. By comparing the corresponding shift for a nuclei in an extended PPII, isolated conformation with respect to that PPII in a bundle, the following DFT-GIAO (Gauge-Including-Atomic-Orbital) theoretical conformational shifts were obtained: $\Delta\delta(\text{C}_\alpha)_{\text{theo}} = -0.59$ ppm, $\Delta\delta(\text{CO})_{\text{theo}} = -0.31$ ppm and $\Delta\delta(\text{H}_\alpha)_{\text{theo}} = -0.61$ ppm, for H-bonded nuclei. The similarity of the experimental and computational conformational shifts provides additional support for the existence of these H-bonds.

Discussion

The stability of Gly-rich PPII helical bundle domains is puzzling due to the high flexibility of glycine⁸. The stabilizing action of disulfide bonds¹⁶, the hydrophobic effect⁷, a limited denatured state conformational entropy¹⁵, and an important number of non-canonical C_αH_α...O = C H-bonds^{7,8} favor

Table 4 | Mean HBS at the M06-2X/6-31 + G(d) level of theory

HBS (kcal·mol ⁻¹)	Interface	Number of Layers			
		2	3	4	5
CO···HN	<i>ab</i>	-5.86	-7.58	-8.24	-8.56
	<i>bc</i>	-	-7.40	-10.03	-11.06
	<i>cd</i>	-	-	-8.27	-11.27
CO···H _α C _α Front	<i>ab</i>	-2.40	-2.86	-3.05	-3.15
	<i>bc</i>	-	-2.96	-3.67	-3.97
	<i>cd</i>	-	-	-3.19	-3.98
CO···HN + H _α C _α Front	<i>ab</i>	-6.72 [-0.827]	-8.49 [-10.44]	-9.20 [-11.30]	-9.55 [-11.71]
	<i>bc</i>	-	-8.19 [-10.35]	-10.94 [-13.69]	-12.07 [-15.03]
	<i>cd</i>	-	-	-9.17 [-11.46]	-12.31 [-15.25]
CO···H _α C _α Inner	<i>ab</i>	-0.57	-0.85	-0.96	-1.01
	<i>bc</i>	-	-0.60	-0.98	-1.14
	<i>cd</i>	-	-	-0.77	-1.18
CO···H _α C _α Outer	<i>ab</i>	-0.27	-0.19	-0.16	-0.14
	<i>bc</i>	-	-0.26	-0.29	-0.33
	<i>cd</i>	-	-	-0.41	-0.56
CO···(H _α) ₂ C _α Inner+Outer	<i>ab</i>	-2.87 [-0.84]	-3.61 [-1.04]	-3.93 [-1.11]	-4.12 [-1.16]
	<i>bc</i>	-	-3.58 [-0.86]	-4.89 [-1.27]	-5.53 [-1.47]
	<i>cd</i>	-	-	-4.22 [-1.18]	-5.92 [-1.75]
Sum	<i>ab</i>	-35.49 (76%)	-44.80 (87%)	-48.58 (93%)	-50.54 (96%)
	<i>bc</i>	-	-43.51 (83%)	-58.43 (99%)	-64.86 (109%)
	<i>cd</i>	-	-	-49.36 (93%)	-67.00 (113%)

HBS between the molecules that reproduce each CO···HN (formamide), CO···H_αC_α front (acetaldehyde), CO···HN + H_αC_α front (acetamide), CO···H_αC_α inner (acetaldehyde), CO···H_αC_α outer (N-methylformamide) and CO···(H_α)₂C_α inner+outer (N-(2-oxoethyl)formamide) H-bond/interaction, and summation of all the contributions in the different interfaces. The summation of the contributions is the HBS resulting from the four CO···HN + H_αC_α front H-bonds and the three CO···(H_α)₂C_α inner+outer H-bonds/interactions. Shown in parentheses is the percentage of total HBS between full-length PPII helices that each value of the summation of the different contributions represents. Shown in brackets is the HBS summation of the isolated CO···HN and CO···H_αC_α front H-bonds (for the CO···HN + H_αC_α front H-bonds) and of the isolated CO···H_αC_α inner H-bonds and CO···H_αC_α outer interactions (for the CO···(H_α)₂C_α inner+outer H-bonds/interactions).

these proteins' folded conformation. Here, we have rationally designed a computational model to scrutinize potential stabilizing contributions in Gly-rich PPII helical bundles that could explain these assemblies' remarkable stability. Four possible H-bond types between individual helices; namely, canonical NH···O=C H-bonds and three types of putative C_αH_α···O=C H-bonds called front, inner and outer were investigated.

Based on the DFT-D results and corroborative evidence from the NBO, QTAIM and SAPT calculations, we have established the existence and cooperativity of canonical NH···O=C and inner C_αH_α···O=C H-bonds, the existence but non-cooperativity of non-canonical front C_αH_α···O=C H-bonds, and the non-existence and hence non-cooperativity of C_αH_α···O=C outer H-bonds. Regarding the relative merits of DFT-D versus SAPT calculations, we find that the marginally higher precision of the latter does not justify its much greater computation costs. Considering that many previous studies of amyloids showed cooperativity between canonical NH···O=C H-bonds^{17–22}, it is not surprising that these interactions also show cooperativity between PPII helices, as previously proposed²³. Furthermore, in both amyloids³⁸ and PPII helical bundles, the NH electron acceptor in canonical H-bonds receives most of the available electron density from the O=C electron donor, which also forms front C_αH_α···O=C H-bonds. This prevents front C_αH_α···O=C non-canonical H-bonds from being cooperative. In contrast, inner C_αH_α···O=C non-canonical H-bonds, which do not have an analogy in amyloids, do exist and show cooperativity and their discovery is a major finding of this study. Our results strongly suggest that C_αH_α···O=C outer H-bonds do neither exist nor afford stability. This reconciles with experimental evidence where glycine residues engage both H_α in H-bonds only in the collagen-like structures²⁴, but just one H_α in Gly-rich PPII bundles^{7,8}.

While the major contribution to the PPII bundle stability comes from canonical H-bonds, the non-canonical ones also make significant contributions. In the interaction between two helices in our model before experiencing cooperativity, i.e., helices *a* and *b* without considering *c*, *d*, and *e*, each CO···HN H-bond represented by two formamide molecules contributes ≈6 kcal/mol (Table 4 and Fig. 5). This increases to ≈7 kcal/mol when also taking into account the extra ≈1 kcal/mol contributed by the CO···H_αC_α front H-bonds as calculated using a pair of acetamide molecules (Table 4 and Fig. 5). Furthermore, if we additionally consider the ≈3 kcal/mol arising from the CO···H_αC_α inner interaction, adequately represented by a pair of N-(2-oxoethyl)formamide molecules (Table 4 and Fig. 5), we reach an HBS of ≈10 kcal/mol per PPII helical turn, as opposed to the initial ≈6 kcal/mol corresponding only to the canonical H-bonds. Therefore, the important role of non-canonical H-bonds in the stability of PPII bundles is highlighted.

These results, supported by ¹H and ¹³C NMR chemical shift data, establish canonical NH···O=C and non-canonical C_αH_α···O=C inner cooperative H-bonds as significant stabilizers of Gly-rich PPII helical bundle domains, especially the long, extensive ones. Indeed, *H. harveyi* produces two different sized isoforms of its AFP: one with six and with thirteen PPII helices (Fig. 1A)⁴⁶. It is pertinent that the longer isoform, which is more active than the shorter one, is stable despite having only one disulfide bridge compared to two in the shorter isoform. We advance that HBC would make a larger stabilizing contribution in the case of the longer *Hh*AFP isoform and could help account for its conformational stability. Moreover, a recent investigation of a homologous AFP from *G. raimieri* found that the insertion of two additional PPII helices augmented biological activity¹⁶.

Conclusions

In the past 20 years, several proteins with domains composed of glycine-rich polyproline II helix bundles have been uncovered, but the basis of their conformational stability is still not completely understood⁴⁷. Here we conclude that the improved comprehension of the basis of PPII helical bundle stability provided by the discovery of HBC could also facilitate the use of Gly-rich PPII helices for protein design. In this regard, our findings raise important questions which should be addressed in future research. First, HN, C_αH_α and CO dipoles in isolated PPII helices are aligned such that they produce a macrodipole^{41,42}, in analogy to that found in α-helices⁴⁸. Based on this and previous reports of charge stabilization of macrodipoles in isolated PPII helices^{41,42}, we advance that proteins or protein domains composed of PPII helical bundles may be stabilized by adding complementary charges to the helices' termini. Secondly, regarding the doubt as to whether the parallel or antiparallel alignment of the PPII helices in bundles is more favorable⁴⁹, our results suggest the later as it would juxtaposition complementary macrodipoles. Thirdly, our approach paves the way towards future studies of PPII helical networks with H-bond networks extending in multiple angles, which may lead to summation or cancelation of macrodipole effects.

Methods

Dispersion-corrected Density Functional Theory (DFT-D)

Density Functional Theory (DFT) calculations were carried out with the high-nonlocality M06-2X functional⁵⁰ that has proven suitable to describe non-covalent interactions and is particularly amenable for extensively H-bonded biological systems^{21,51,52}. As for the basis set, we chose the 6-31 + G(d) as an appropriate balance between accuracy and computational cost for the systems treated here that contain a large number of atoms. Following literature precedents dealing with large biomolecular systems containing multiple H-bonding interactions^{53–55}, the D3 dispersion correction scheme⁵⁶ was applied to all the calculations. Every interaction energy was counterpoise (CP)-corrected following the method proposed by Boys and Bernardi to consider the inherent basis set superposition error⁵⁷. Finally, it is worth emphasizing that, in order to know the intrinsic energy of the systems under study, all quantum mechanical calculations have been carried out without the inclusion of any implicit or explicit solvent.

The geometry of the PPII helices in the bundle was optimized by keeping the C_αs frozen, but the amide group atoms and hydrogen atoms were allowed to undergo energy minimization to optimize H-bonding interactions and to account for the lack of hydrogens in the PDB structure used as a starting point (PDB ID: 3BOG). The energies of the monomers were computed with the same structure that they present in the optimized bundle to ensure that the interaction energy strictly corresponds to the established H-bonds. In this way, the HBS of XY interface can be calculated using Eq. 1:

$$HBS(XY) = E(X_i) - E(Y_i) \quad (1)$$

For example, the HBS of bc interface for the case of five PPII helices would be obtained through Eq. 2:

$$HBS(bc)_{abcde} = E(abcde) - E(ab) - E(cde) \quad (2)$$

For the layers of the mimetic molecules, which maintain the same geometry as the corresponding atoms in the optimized PPII helical bundle, the same process was followed as with the complete PPII helices to calculate the HBS relative to each type of H-bond. The rest of the computational parameters from NBO and QTAIM analyses can be obtained directly at the interface of interest for a given number of PPII helices, while in SAPT calculations the interaction energy is determined without computing the total energy of the monomers or n-mers. Each of the points that appear in the different graphs of this study correspond to the images presented in Fig. 1C.

Natural Bonding Orbital (NBO)

The NBO analysis⁵⁸ has been performed at the same level of theory, M06-2X/6-31 + G(d), to test the H-bonding nature of the interactions between glycine residues in the Gly-rich PPII bundles studied here. In the context of the NBO methodology, H-bonds are characterized as an electron delocalization from a lone pair (n) of the H-acceptor atom (e.g., an O atom) onto the sigma antibonding (σ*) orbital of the covalent bond between the H-donor atom and the H atom itself (e.g., an N-H or C-H bond). This interaction can be quantitatively described through the stabilization energy, $E(2)$, which is estimated based on the Second Order Perturbation Theory:

$$E(2) = \Delta E_{(ij)} = q_i \frac{F_{ij}^2}{E_i - E_j} \quad (3)$$

where q_i is the occupancy of the electron-donor orbital i , E_i and E_j are the energies of i and j , respectively, and F_{ij} is the Fock or Kohn-Sham matrix element between both NBOs. The higher the value of q_j and $E(2)$, the greater the $i \rightarrow j$ electron transfer, leading to a greater stabilization of the molecule.

Quantum Theory of Atoms in Molecules (QTAIM)

Canonical CO...HN and non-canonical CO...H_αC_α H-bonds in the Gly-rich PPII bundles have been also analyzed using the QTAIM approach^{59,60} as described in detail in the supplementary material.

Symmetry-Adapted Perturbation Theory (SAPT)

Canonical CO...HN and non-canonical CO...H_αC_α interactions in the Gly-rich PPII bundles have been further explored using the SAPT methodology^{61,62} as described in detail in the supplementary material.

Theoretical chemical shift calculations

Nuclear Magnetic Resonance (NMR) shielding tensors of both PPII helices isolated and in bundles were calculated using the gauge-including-atomic-orbital (GIAO) method^{63,64} at the M062X/6-31 + G(d) computational level too. The corresponding chemical shifts, δ , are obtained as the difference between isotropic chemical shieldings in the nucleus of interest in our model and in the compound used as a reference, σ_{iso} and $\sigma_{iso,ref}$ respectively:

$$\delta = \sigma_{iso} - \sigma_{iso,ref} \quad (4)$$

The conformational chemical shifts, $\Delta\delta$, of a protein are the experimental chemical shifts, δ_2 , relative to values expected for statistical coil, δ_1 , which translates into the difference between two different chemical shifts of the same nucleus for which the same reference compound has been used:

$$\Delta\delta = \delta_2 - \delta_1 = \sigma_{iso,2} - \sigma_{iso,ref} - \sigma_{iso,1} + \sigma_{iso,ref} = \sigma_{iso,2} - \sigma_{iso,1} \quad (5)$$

Computationally, in DFT-GIAO calculations structures are treated as static instead of dynamic species. Bearing this in mind, and considering that in solution Gly-rich sequences adopt extended PPII helix conformations over statistical coil^{43–45}, the theoretical $\Delta\delta$ values have not been calculated using a statistical coil state as the reference. Instead, $\Delta\delta$ values have been obtained as the difference between the σ_{iso} of nuclei when the PPII helix is in the model bundle ($\sigma_{iso,2}$), geometrically optimized as described in the DFT-D calculations section, with respect to the PPII helix in isolation ($\sigma_{iso,1}$), freezing backbone atoms to maintain the positions seen in the crystallographic structure used as a starting point to create the model system (PDB ID: 3BOG), while ¹H nuclei that are absent in this X-ray structure were optimized.

The Quantum Chemistry package used to perform the geometry optimizations, and single point and NMR spectroscopic shielding calculations was Gaussian 16⁶⁵. The software packages used for the visualization and modeling of the PPII helical-bundles were GaussView 6⁶⁶, Pymol 2.3.0⁶⁷ and Jmol 14⁶⁸. For the NBO analysis, NBO Version 3.1⁶⁹ was used as

implemented in Gaussian 16. The QTAIM analysis was carried out using the Multiwfn 3.8 package⁷⁰. Finally, SAPT calculations were performed using the Psi4 software⁷¹.

Data availability

All relevant data are included in the manuscript main text and supplementary material.

Received: 11 December 2023; Accepted: 6 August 2024;

Published online: 30 August 2024

References

1. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019).
2. Baker, E. G., Bartlett, G. J., Porter Goff, K. L. & Woolfson, D. N. Miniprotein design: past, present, and prospects. *Acc. Chem. Res.* **50**, 2085–2092 (2017).
3. Shiga, S. et al. Domain-swapping design by polyproline rod insertion. *ChemBioChem* **20**, 2454–2457 (2019).
4. Mompeán, M., Oroz, J. & Laurents, D. V. Do polyproline II helix associations modulate biomolecular condensates? *FEBS Open Bio* **11**, 2390–2399 (2021).
5. Lamers, B. A. G. et al. Oligodimethylsiloxane-oligoproline block co-oligomers: the interplay between aggregation and phase segregation in bulk and solution. *J. Am. Chem. Soc.* **143**, 4032–4042 (2021).
6. Pentelute, B. L. et al. X-Ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc.* **130**, 9695–9701 (2008).
7. Treviño, M. Á. et al. The singular NMR fingerprint of a polyproline II helical bundle. *J. Am. Chem. Soc.* **140**, 16988–17000 (2018).
8. Gates, Z. P. et al. 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 (2017).
9. Bartlett, G. J., Newberry, R. W., VanVeller, B., Raines, R. T. & Woolfson, D. N. Interplay of hydrogen bonds and N→π* interactions in proteins. *J. Am. Chem. Soc.* **135**, 18682–18688 (2013).
10. Newberry, R. W. & Raines, R. T. The N→π* interaction. *Acc. Chem. Res.* **50**, 1838–1846 (2017).
11. Zhai, L., Otani, Y. & Ohwada, T. Uncovering the networks of topological neighborhoods in β-strand and amyloid β-sheet structures. *Sci. Rep.* **9**, 10737 (2019).
12. Scholl, C. L., Tsuda, S., Graham, L. A. & Davies, P. L. 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 (2021).
13. Mok, Y.-F. et al. Structural basis for the superior activity of the large isoform of snow flea antifreeze protein. *Biochemistry* **49**, 2593–2603 (2010).
14. Li, T. et al. Structural basis for ligand reception by anaplastic lymphoma kinase. *Nature* **600**, 148–152 (2021).
15. Treviño, M. Á. et al. Insight into polyproline II helical bundle stability in an antifreeze protein denatured state. *Biophys. J.* **121**, 4560–4568 (2022).
16. Scholl, C. L. & Davies, P. L. Protein engineering of antifreeze proteins reveals that their activity scales with the area of the ice-binding site. *FEBS Lett.* **597**, 538–546 (2023).
17. Viswanathan, R., Asensio, A. & Dannenberg, J. J. Cooperative hydrogen-bonding in models of antiparallel β-sheets. *J. Phys. Chem. A* **108**, 9205–9212 (2004).
18. Tsemekhman, K., Goldschmidt, L., Eisenberg, D. & Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* **16**, 761–764 (2007).
19. Mompeán, M. et al. Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett.* **6**, 2608–2615 (2015).
20. Plumley, J. A. & Dannenberg, J. J. The importance of hydrogen bonding between the glutamine side chains to the formation of amyloid VQIVYK parallel β-sheets: an ONIOM DFT/AM1 study. *J. Am. Chem. Soc.* **132**, 1758–1759 (2010).
21. Mompeán, M., Nogales, A., Ezquerra, T. A. & Laurents, D. V. Complex system assembly underlies a two-tiered model of highly delocalized electrons. *J. Phys. Chem. Lett.* **7**, 1859–1864 (2016).
22. Mompeán, M., Baralle, M., Buratti, E. & Laurents, D. V. An amyloid-like pathological conformation of TDP-43 is stabilized by hypercooperative hydrogen bonds. *Front. Mol. Neurosci.* **9**, 125 (2016).
23. Plumley, J. A., Tsai, M. & Dannenberg, J. J. Aggregation of capped hexaglycine strands into hydrogen-bonding motifs representative of pleated and rippled β-sheets, collagen, and polyglycine I and II crystal structures. a density functional theory study. *J. Phys. Chem. B* **115**, 1562–1570 (2011).
24. Bella, J. & Berman, H. M. Crystallographic evidence for Ca–H···O = C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 734–742 (1996).
25. Ramachandran, G. N., Sasisekharan, V. & Ramakrishnan, C. Molecular structure of polyglycine II. *Biochim. Biophys. Acta* **112**, 168–170 (1966).
26. Krimm, S. & Bandekar, J. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* **38**, 181–364 (1986).
27. Kar, T. & Scheiner, S. Comparison of cooperativity in CH···O and OH···O hydrogen bonds. *J. Phys. Chem. A* **108**, 9161–9168 (2004).
28. Blatchford, M. A., Raveendran, P. & Wallen, S. L. Spectroscopic studies of model carbonyl compounds in CO₂: evidence for cooperative C – H···O interactions. *J. Phys. Chem. A* **107**, 10311–10323 (2003).
29. Li, Q., An, X., Gong, B. & Cheng, J. Spectroscopic and theoretical evidence for the cooperativity between red-shift hydrogen bond and blue-shift hydrogen bond in DMSO aqueous solutions. *Spectrochim. Acta A. Mol. Biomol. Spectrosc.* **69**, 211–215 (2008).
30. Esrafilii, M. D., Behzadi, H. & Hadipour, N. L. Theoretical study of N–H···O hydrogen bonding properties and cooperativity effects in linear acetamide clusters. *Theor. Chem. Acc.* **121**, 135–146 (2008).
31. Ludwig, R. Cooperative hydrogen bonding in amides and peptides. *J. Mol. Liq.* **84**, 65–75 (2000).
32. Parthasarathi, R., Raman, S. S., Subramanian, V. & Ramasami, T. Bader's electron density analysis of hydrogen bonding in secondary structural elements of protein. *J. Phys. Chem. A* **111**, 7141–7148 (2007).
33. Arnold, W. D. & Oldfield, E. The chemical nature of hydrogen bonding in proteins via NMR: J-Couplings, chemical shifts, and AIM theory. *J. Am. Chem. Soc.* **122**, 12835–12841 (2000).
34. Alkorta, I. et al. Cooperativity in multiple unusual weak bonds. *Theor. Chem. Acc.* **126**, 1–14 (2010).
35. Venkataramanan, N. S. & Suvitha, A. Nature of bonding and cooperativity in linear DMSO clusters: a DFT, AIM and NCI analysis. *J. Mol. Graph. Model.* **81**, 50–59 (2018).
36. Sánchez-Sanz, G., Trujillo, C., Alkorta, I. & Elguero, J. Understanding regium bonds and their competition with hydrogen bonds in Au₂:HX complexes. *ChemPhysChem* **20**, 1572–1580 (2019).
37. Jezierska, A. & Panek, J. J. Cooperativity of hydrogen bonding network in microsolvated biotin, the ligand of avidin class proteins. *J. Mol. Model.* **25**, 361 (2019).
38. Scheiner, S. Contributions of NH···O and CH···O hydrogen bonds to the stability of β-sheets in proteins. *J. Phys. Chem. B* **110**, 18670–18679 (2006).

39. Scheiner, S. Theoretical analysis of the contributions made by CH••OH bonds to protein structure. *Curr. Org. Chem.* **14**, 106–128 (2010).
40. Scheiner, S. Dissection of the factors affecting formation of a CH•••O H-bond. A case study. *Crystals* **5**, 327–345 (2015).
41. Kuemin, M., Schweizer, S., Ochsenfeld, C. & Wennemers, H. Effects of terminal functional groups on the stability of the polyproline II structure: a combined experimental and theoretical study. *J. Am. Chem. Soc.* **131**, 15474–15482 (2009).
42. Huang, K.-Y. & Horng, J.-C. Impacts of the terminal charged residues on polyproline conformation. *J. Phys. Chem. B* **123**, 138–147 (2019).
43. Ohnishi, S., Kamikubo, H., Onitsuka, M., Kataoka, M. & Shortle, D. Conformational preference of polyglycine in solution to elongated structure. *J. Am. Chem. Soc.* **128**, 16338–16344 (2006).
44. Bykov, S. & Asher, S. Raman studies of solution polyglycine conformations. *J. Phys. Chem. B* **114**, 6636–6641 (2010).
45. Mompeán, M. et al. Glycine rich segments adopt polyproline II helices: implications for biomolecular condensate formation. *Arch. Biochem. Biophys.* **704**, 108867 (2021).
46. Graham, L. A. & Davies, P. L. Glycine-rich antifreeze proteins from snow fleas. *Science* **310**, 461–461 (2005).
47. Rodríguez, C. S. & Laurents, D. V. Architectonic principles of polyproline II helix bundle protein domains. *Arch. Biochem. Biophys.* **756**, 109981 (2024).
48. Wada, A. The Alpha-Helix as an Electric Macro-Dipole. *Adv. Biophys.* **9**, 1–63 (1976).
49. Crick, F. H. C. & Rich, A. Structure of polyglycine II. *Nature* **176**, 780–781 (1955).
50. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).
51. Mompeán, M. et al. Molecular mechanism of the inhibition of TDP-43 amyloidogenesis by QBP1. *Arch. Biochem. Biophys.* **675**, 108113 (2019).
52. Martín-Pintado, N. et al. Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.* **135**, 5344–5347 (2013).
53. Arey, J. S., Aeberhard, P. C., Lin, I.-C. & Rothlisberger, U. Hydrogen bonding described using dispersion-corrected density functional theory. *J. Phys. Chem. B* **113**, 4726–4732 (2009).
54. DiLabio, G. A., Johnson, E. R. & Otero-de-la-Roza, A. Performance of conventional and dispersion-corrected density-functional theory methods for hydrogen bonding interaction energies. *Phys. Chem. Chem. Phys.* **15**, 12821–12828 (2013).
55. Boese, A. D. Density functional theory and hydrogen bonds: are we there yet? *ChemPhysChem* **16**, 978–985 (2015).
56. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
57. Boys, S. F. & Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19**, 553–566 (1970).
58. Reed, A. E., Curtiss, L. A. & Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88**, 899–926 (1988).
59. Bader, R. F. W. Atoms in Molecules. *Acc. Chem. Res.* **18**, 9–15 (1985).
60. Bader, R. F. W. *Atoms in Molecules: A Quantum Theory* (Clarendon Press, 1990).
61. Jeziorski, B., Moszynski, R. & Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **94**, 1887–1930 (1994).
62. Hohenstein, E. G. & Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **133**, 014101 (2010).
63. Ditchfield, R. Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* **27**, 789–807 (1974).
64. Wolinski, K., Hinton, J. F. & Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **112**, 8251–8260 (1990).
65. Frisch, M. J. et al. Gaussian 16, Rev. A.01 (Wallingford, CT, 2016).
66. Dennington, R., Keith, T. A. & Millam, J. M. GaussView, version 6.0.16 (Semichem Inc., Shawnee Mission, KS, USA, 2016).
67. DeLano, W. L. et al. Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).
68. Jmol Development Team. (2016).
69. Glendening, E. D., Reed, A. E., Carpenter, J. E., Weinhold, F. NBO Version 3.1 (Gaussian Inc., Pittsburgh, 2003).
70. Lu, T. & Chen, F. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **33**, 580–592 (2012).
71. Smith, D. G. A. et al. PSI4 1.4: open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **152**, 184108 (2020).

Acknowledgements

This work has been funded by grants PID2019-109306RB-I00 & PID2022-137806OB-I00 (D.V.L.) and PID2020-113907RA-I00 (M.M.) from MCIN/AEI/10.13039/501100011033. Co-funded by the European Union (ERC, 101042403 – BiFOLDOME). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors also gratefully acknowledge the SGA1-CSIC for its assistance and help while using the DRAGO Supercomputer.

Author contributions

Funding acquisition: D.V.L. and M.M.; Conceptualization: R.L.S., D.V.L., and M.M.; Data acquisition and analysis: R.L.S. and M.M.; Data Interpretation: R.L.S. and M.M.; Manuscript preparation: R.L.S., D.V.L., and M.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42004-024-01268-2>.

Correspondence and requests for materials should be addressed to Douglas V. Laurents or Miguel Mompeán.

Peer review information *Communications Chemistry* thanks Konstantin Röder and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024