JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

The Singular NMR Fingerprint of a Polyproline II Helical Bundle

Miguel Ángel Treviño, David Pantoja-Uceda, Margarita Menendez, M. Victoria Gomez, Miguel Mompean, and Douglas V. Laurents

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.8b05261 • Publication Date (Web): 15 Nov 2018

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

The Singular NMR Fingerprint of a Polyproline II Helical Bundle

Miguel Ángel Treviño[1], David Pantoja-Uceda[1], Margarita Menéndez[1,2], M. Victoria Gomez[3], Miguel Mompeán[3,*], Douglas V. Laurents[1,*]

[1] "Rocasolano" Institute for Physical Chemistry, Spanish National Research Council, Serrano 119, 28006, Madrid, Spain

[2] Ciber of Respiratory Diseases (CIBERES)

[3] University of Castile-La Mancha, Instituto Regional de Investigación Científica Aplicada (IRICA), 13071, Ciudad Real, Spain.

*To whom correspondence may be addressed:
miguelangel.mompean@uclm.es or dlaurents@iqfr.csic.es

ABSTRACT

Polyproline II (PPII) helices play vital roles in biochemical recognition events and structures like collagen, and form part of the conformational landscapes of intrinsically disordered proteins (IDPs). Nevertheless, this structure is generally hard to detect and quantify. Here, we report the first thorough NMR characterization of a PPII helical bundle protein, the *Hypogastrura harveyi* "snow flea" AntiFreeze Protein (sfAFP). J-couplings and NOESY spectra confirm a natively folded structure consisting of six PPII helices. NMR spectral analyses reveal quite distinct Hα2 versus Hα3 chemical shifts for 28 Gly residues, as well as ^{13}Cα, ^{15}N and ^{1}HN conformational chemical shifts (Δδ) unique to PPII helical bundles. The ^{15}N Δδ and ^{1}HN Δδ values and small negative ^{1}HN temperature coefficients evince hydrogen bond formation. ^{1}H-^{15}N relaxation measurements reveal that the backbone structure is generally highly rigid on ps-ns time scales. NMR relaxation parameters and biophysical characterization reveal that sfAFP is chiefly a dimer. For it, a structural model featuring the packing of long, flat hydrophobic faces at the dimer interface is advanced. The conformational stability, measured by amide H/D exchange to be 6.24 +/− 0.2 kcal·mol^{-1}, is elevated. These are extraordinary findings considering the great entropic cost of fixing Gly residues and, together with the remarkable upfield chemical shifts of 28 Gly ^{1}Hα, evidence significant stabilizing contributions from CαH ||| O=C hydrogen bonds. These stabilizing interactions are corroborated by density functional theory (DFT) calculations and natural bonding orbital (NBO) analysis. The singular conformational chemical shifts, J-couplings, high hNOE ratios, small negative temperature coefficients and slowed H/D exchange constitute a unique set of fingerprints to identify PPII helical bundles, which may be formed by hundreds of Gly-rich motifs detected in sequence databases. These results should aid the quantification of PPII helices in IDPs, the development of improved antifreeze proteins, and the incorporation of PPII helices into novel designed proteins.

Introduction

Glycine and Proline are common residues in intrinsically disordered proteins (IDPs). For example, human CPEB3, a putative functional amyloid implicated in memory consolidation, contains over 25 prolines in the first hundred residues, and TDP-43 bears a Gly-rich disordered region[1] whose aggregation is implicated in Amyotrophic Lateral Sclerosis. Glycine's conformational freedom and proline's steric clashes disfavor their presence in α-helices and β-strands. Nevertheless, this odd couple collaborates to form the polyproline II triple helix of collagen, wherein every three-residue turn of helix contains an internal glycine residue that forms interhelical hydrogen bonds. Glycine and proline's unusual properties thwart efforts to predict their structural tendencies, leading to debate on the presence polyproline II helices in IDPs[2-5]. Despite progress[6-9], more work to identify polyproline II helices in IDPs and understand their biochemical contributions is needed.

In 2008, the glycine-rich *Hypogastrura harveyi* snow flea antifreeze protein's[10] (sfAFP) 3D structure was resolved by X-ray crystallography[11]; it consists of a bundle of six polyproline II helices composed of Gly-Gly-X repeats (**Fig 1A**), where X is Ala, Arg, Asn, Asp, His, Lys, Ser, Thr or Val. The glycine residues occupy internal positions and form a rich network of hydrogen bonds between multiple helices. To overcome the great free energy cost of fixing glycine residues in a folded protein[12], Sosnick and co-workers advanced[13], on the basis of folding experiments and computational analyses, that the structure would be stabilized by exceptional short and stable N-H ||| O=C hydrogen bonds and Cα-Hα ||| O=C hydrogen bonds. Motivated by these proposals, we have characterized the dynamics, stability and hydrogen bonding in sfAFP using NMR spectroscopy. This is the first thorough study of a polyproline II helical bundle by NMR. We find strong evidence for Cα-Hα ||| O=C hydrogen bonds, a high level of rigidity pervading the protein except for one loop segment and dimer formation via packing of the nonpolar faces. Thanks to the stably folded nature of sfAFP, this protein represents an ideal model system for obtaining NMR chemical shifts and coupling constants for polyproline II helical bundles. These findings allow us to propose NMR criteria for identifying polyproline II helical bundles, which may be rather common in humans. Polyproline helices play key roles in several of the most vigorous branches of modern Chemistry such as Computational Chemistry[14], Chemical Biology[15], Novel Materials[16], Protein Chemistry[17] and Mass Spectrometry Techniques[18] and NMR Spectroscopy Methods[19]. Therefore, these results should aid efforts to: 1) understand the influence of polyproline II helices on amyloid formation, 2) improve protein secondary structure prediction based on chemical shift data, 3) manufacture improved antifreeze proteins and 4) design novel proteins incorporating polyproline II helices.

Materials and Methods

Sample Preparation: The DNA coding sequence from the mature short isoform of sfAFP (residues 23 to 103) was synthesized (Genscript, New Jersey, NJ) and cloned in pET45b+. The translated product containing an N-terminal hexa-histidine tag, a cleavage site for enterokinase and sfAFP was expressed in BL21 (DE3) or SHuffle (NEB, Ipswich, MA). Briefly, the cells were grown in LB at 37 ºC until OD_{600} = 0.8, then the temperature was decreased to 20 ºC and 0.4 mM IPTG was added to induce the expression of the heterologous protein. The protein was purified by Ni^{++} affinity chromatography followed by cationic exchange chromatography and the eluted product was dissolved in 8 M urea and refolded by a rapid, dropwise dilution into aqueous buffer containing 50 mM potassium phosphate pH 7.8, 8 mM cysteine, 1 mM cystine, followed by incubation at 4 ºC during 24 h[20]. The correctly folded protein was purified from the soluble incorrectly-folded protein taking advantage of its capacity to bind ice by collecting the ice formed around a cold finger with internally circulating liquid between -0.5 and -4 ºC[21].

Samples were uniformly labeled with ^{13}C and ^{15}N using a modified protocol derived from Sivashanmugam *et al.* [22]. Briefly, the transformed cells were grown in rich LB media until OD_{600}=2-3; then they were centrifuged and the pellet resuspended in minimal medium containing ^{13}C-glucose and $^{15}NH_4Cl$. The volume of the minimal media was half that of the LB media. The media was incubated at 37 ºC for one hour and then the temperature was decreased to 20 ºC before induction with IPTG in the same conditions described above. For trial samples employing Enterokinase (New England Biolabs) cleavage to remove the His-tag, the unlabeled sample showed a MALDI-TOF mass spectrometry peak close to its calculated MW of 6483.80 Da and the ^{13}C,^{15}N-labeled sample yielded a peak at 6804 Da. Considering that 100% ^{13}C,^{15}N labeled sfAFP would have a mass of 6830.23 Da, the average incorporation of ^{13}C and ^{15}N in the labeled sample is calculated to be 92.4 %. Sample concentrations were estimated by UV absorbance on a Cary 210 dual beam UV-VIS spectrophotometer using an extinction coefficient of 303,000 $cm^{-1} \cdot M^{-1}$ at 205 nm calculated using the parameters of Anthis and Clore[23].

NMR Spectroscopy and Spectral Assignment: A series of 2D and 3D NMR spectra were acquired at pH 6.1, 15 ºC in 15% D_2O buffer containing 10 mM NaCl and 5 mM sodium phosphate buffer (consisting of 4.5 mM NaH_2PO_4 and 0.5 mM Na_2HPO_4) and analyzed to obtain resonance assignments including: 2D ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC; 3D HNCO, HNcaCO, HNCA, HNcoCA, CBCAcoNH, HNCACB, and ^{1}H-^{1}H-^{15}N HSQC-NOESY. Additional spectra; namely: 2D ^{1}H-^{15}N HSQC and 3D CBCANH, HcccNH, HncacoNH and hNcacoHN, were recorded at 5.0 ºC in the same buffer to obtain additional assignments and determine temperature coefficients. Two 2D ^{1}H-^{1}H NOESY spectra, with mixing times of 80 and 120 ms, were recorded to assess the native folding. To assign the resonances in the pH* 5.12 H/D exchange experiments, additional 2D ^{1}H-^{15}N HSQC spectra were recorded at 5.0 ºC, pH 5.42 and 5 ºC, pH 4.75. pH* is the pH reading in D_2O without correction for the deuterium isotope effect. The ratio of the $^{1}H\alpha$-^{1}HN crosspeak to the ^{1}HN-^{1}HN diagonal peak intensities in the 3D HNHA

spectrum was utilized to calculate the $^{3}J_{HNCH_{\alpha}}$ coupling constants following the procedure of Vuister and Bax[24]. The $^{1}J_{CH_{\alpha}}$ coupling constants were measured by recording a non-decoupled 2D ^{1}H-^{13}C HSQC spectrum in D_2O. All spectra were recorded on a Bruker 800 MHz (^{1}H) Avance spectrometer fitted with a triple resonance TCI cyroprobe and Z-gradients, except the relaxation experiments described below were repeated on a Bruker 600 MHz (^{1}H) Avance spectrometer fitted with a ^{1}H cryoprobe and Z-gradients.

Chemical Shifts (δ): To calculate the conformational chemical shifts (Δδ), that is, the difference between the experimental chemical shift values and the corresponding values expected for a statistical coil, it is necessary to have accurate δ values for the statistical coil. These values were calculated using the sfAFP sequence and the data compiled by Poulsen, Dyson and co-workers[25-27] at pH 6.13 and 15 ºC using the web server: https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/. Since that server assumes reduced Cys, but the four Cys form disulfide bonds in sfAFP, a correction for Cys based on the data of Wishart and co-workers was made for Cys for the $^{13}C\alpha$, $^{13}C\beta$, $^{1}H\alpha$, ^{1}HN and ^{15}N nuclei[28]. No correction was made for the ^{13}CO of the Cys and these nuclei were excluded from **Figure 2, Sup. Fig. 2** and the statistical analyses. To evaluate the significance of chemical shift deviations, the p-value was calculated using the Wilcoxon-Mann-Whitney test using the normal approximation as implemented in Kaleidagraph version 3.6.

Backbone Dynamics from ^{15}N Relaxation: The heteronuclear $^{15}N\{^{1}H\}$ NOE (hNOE) of backbone amide groups was recorded at 5.0 ºC as the ratio of spectra registered with and without saturation in an interleaved mode. Following the recommendations in Renner *et al.*[29], steady-state ^{15}N-^{1}H heteronuclear NOE measurements were performed with an overall recycling delay of 10 s or more. This allows solvent relaxation and ensures maximal development of NOEs before acquisition. Four sets of experiments, three at 800 MHz and the fourth at 600 MHz were recorded. Regarding the three experiments done at 800 MHz, the first used the Bruker default 120º high power pulse for ^{1}H saturation and the ^{13}C,^{15}N-labeled sample. This experiment was repeated using a ^{15}N-labeled sfAFP sample with a 180º ^{1}H saturation pulse, shown by Ferrage *et al.*[30] to afford superior ^{1}H saturation, and for comparison, pulses of 120º (Bruker default). Uncertainties in peak intensities were determined from the standard deviation of intensities from spectral regions devoid of signal that contain only noise. See **Sup. Infor.** for details on R_1 and R_2 measurements. The HYDRONMR program[31] was utilized to predict the correlation time of the protein using the X-ray crystal structures reported in PDB files: 2PNE and 3BOG[11]. The correlation time, t_c was calculated as $t_c \simeq 1 / (4 \pi * {}^{15}N \text{ frequency in Hz}) * ((6* R_2/ R_1) - 7))^{½}$ which was derived from eqn. 8 of Bax and coworkers[32], that takes into account J(o) and J(w) spectral densities and discounts terms from higher frequencies, which is a good approximation for smaller proteins (< 25 kDa). From the t_c, the molecular weight can be estimated using the empirical equation MW (in kDa) $\simeq 1.494 \cdot t_c + 1.1187$, based on the experimental data of Montelione and coworkers[33].

Analytical Gel Filtration: A Superdex G75 column (GE Healthcare) cooled to near 0 ºC with an ice jacket was calibrated with bovine serum albumin (66 kDa), egg white albumin (44 kDa), carbonic anhydrase (29 kDa), RNase A C-dimer (27.4 kDa), RNase A (13.7 kDa), 434 Cro (7.4 kDa) and aprotinin (6.5 kDa). All proteins were from Sigma, except RNase A C-dimer and 434 Cro, which were kind gifts of Prof. Giovanni Gotte (University of Verona) and Prof. S. Padmanabhan (IQFR, CSIC), respectively. These proteins were prepared as previously described[34-35]. A 0.75 mL/min flow rate and 0.10 mL injection volumes were used. A far UV Zn lamp was used to detect sfAFP, which lacks Tyr and Trp residues. The sfAFP concentration upon injection was approximately 30 μM. The running buffer contained 5 mM sodium phosphate, pH 6.1 plus 100 mM NaCl to minimize nonspecific interactions between proteins and the column matrix. Experiments were performed with an ÄKTA FPLC system (GE Healthcare) and data were collected and analyzed using the system's Unicorn software.

Analytical Ultracentrifugation: Sedimentation equilibrium experiments were carried out at 18,000 rpm as described previously[36], and weight-average molecular weights ($M_{w,app}$) were calculated using the HETEROANALYSIS program, version 1.1.2 (http://www.biotech.uconn.edu/auf/). Measurements were performed at 10 °C in 5 mM sodium phosphate (pH 6.2) with 15% D_2O, using protein concentrations ranging from 54.5 μM to 322 μM in an Optima XL-A analytical ultracentrifuge (Beckman Coulter) with 12-mm or 3-mm epon-charcoal double-sector centrepieces (AN50-Ti rotor). Radial scans were taken at 240 nm. The partial specific volume of sfAFP and the buffer's density were estimated from the amino acid sequence and the buffer composition, respectively, using the SEDNTERP program[37].

Mass Spectrometry: MALDI-TOF mass spectra were registered in the "Rocasolano" Institute for Physical Chemistry's Mass Spectrometry laboratory with a Voyager-DE Pro instrument (Applied Biosystems). Sinapinic acid was utilized as matrix.

Temperature Coefficients: The change in δ^1HN with temperature ($\partial\ \delta\ /\ \partial\ T$) was measured by comparing the ^{1}H δ values in 2D ^{1}H-^{15}N HSQC spectra at 5.0 ºC and 15.0 ºC. The linearity of this dependence was corroborated for several ^{1}HN signals by measurement of 1D ^{1}H NMR spectra at 5.0 ºC, 10.0 ºC and 15.0 ºC on an independent, unlabeled sample.

Hydrogen/Deuterium (H/D) Exchange: The residue level conformational stability (ΔG_{HX}) of sfAFP at concentrations ranging from 400 to 700 μM (as monomer) was studied by H/D exchange. Two experiments, one at pH* = 5.12 and one at pH* = 6.13, were performed by dissolving lyophilized sfAFP in buffer prepared with D_2O (99.96 % atom D) from Euriso-top (St. Aubin, France). Please see the **Supp. Info.** for additional experimental details.

Structural Modeling: Data-driven docking of two sfAFP monomers was performed using HADDOCK[38]. The calculation was based residues whose H/D exchange rate was unusually slow, considering that they appear solvent-exposed in the monomer structure. More precisely, the residues used as the "active residues" (*i.e.,* directly present in the monomer/monomer interface) are 25, 47, 50, 53, and 80. Regarding the calculation procedure itself, 200 solutions were picked after rigid body-energy minimization, based on their inter-molecular energies. These structures were then subjected to semiflexible docking and explicit water solvation. Finally, 159 water-refined models were obtained, clustered, and sorted based on their HADDOCK scores. Structures were analyzed and figures were prepared using KiNG (version 2.23) and PyMOL (Schrodinger LLC, version 2.0.6).

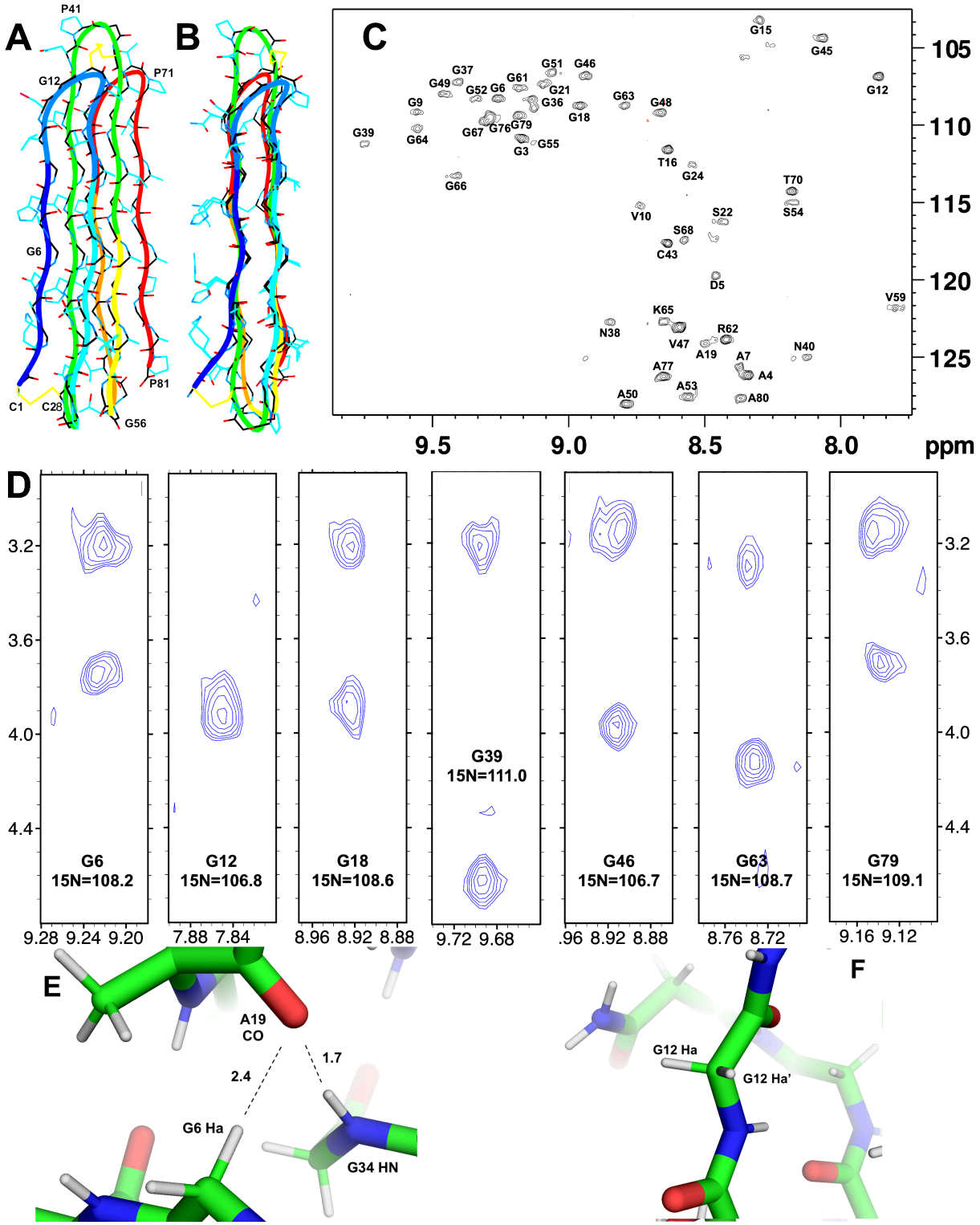
Density Functional Theory (DFT) and Natural Bonding Orbital (NBO) Computational Analyses: Hydrogen atoms were added to PDB structure 3BOG and their positions were optimized at the B3LYP/6-31g(d,p) level of theory. All DTF calculations were performed utilizing Gaussian09 software[39], using the B·LYP functional[40] and the 6-311+G(2d,p) basis set. Chemical shielding tensors were obtained using the GIAO method[41-42]. NBO analyses were carried out with the NBO 3.1 program[43] as implemented within the Gaussian09 software. The structures used to obtain both the chemical shielding tensors and NBOs were generated utilizing PyMOL to edit PDB entry 3BOG as explained in the Results section and all Gaussian input files were generated using JMOL. To visualize NBOs, we used the cubegen Gaussian09 tool to generate appropriate cubefiles containing the orbitals' coefficients that can be directly visualized with JMOL.

Database analysis: The search for human proteins containing sfAFP-like polyproline II helix sequences was performed using the MOTIF program on the GenomeNet webpage: http://www.genome.jp, developed by the laboratory of Prof. M. Kanehisa[44]. Sequence comparison of the nucleoplasmin C-terminal domain sequence was performed with SMART BLAST (http:blast.ncbi.nlm.nih.gov)[45].

RESULTS

sfAFP's Chemical Shifts Show Unique Changes Due to Its Singular Conformation: The backbone ^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα, ^{1}Hα and ^{13}Cβ chemical shifts (δ) were assigned by analysis of a suite of 2D and 3D spectra, taking advantage of ^{13}CO, ^{13}Cα, ^{13}Cβ, and ^{1}Hα connectivities in order to overcome difficulties arising from the low sequence diversity. The assignments are nearly complete; the only missing resonances are for residues Cys 1 (all nuclei), Cys 13 (^{13}Cβ and ^{13}CO), Gly 34 (all nuclei), Asn 35 (^{1}HN & ^{15}N), Cys 43 ^{13}Cβ and Pro 81 (all nuclei). These data are listed in **Sup. Table 1**. They are deposited in the BMRB database (accession number **27473**) along with the hNOE ratios, $^{3}J_{HNHa}$ coupling constants and H/D exchange rates listed in **Sup. Table 2**. The backbone ^{1}H-^{15}N chemical shifts at 5.0 ºC have also been deposited in the BMRB. Despite its low sequence complexity, sfAFP's 2D ^{1}H-^{1}H NOESY spectra show excellent chemical shift dispersion, which is a hallmark of a well-folded protein (**Sup. Fig 1**). In fact, the downfield shifted HN groups show NOE crosspeaks revealing close contacts very similar to those observed in the crystal structure of sfAFP (**Sup. Fig 1, Fig. 1A, 1B**). The broad chemical shift (δ) dispersion characteristic of folded proteins is also observed in the 2D ^{1}H-^{15}N HSQC spectrum (**Fig 1C**). What is even more fascinating is that the Hα2 and Hα3 signals of Gly residues at internal positions show strikingly large chemical shifts differences, generally over 0.5 ppm and even up to 1.4 ppm in the case of Gly 39 (**Fig 1D**). This is ascribed to participation of one Gly ^{1}Hα -but not the other- in internal hydrogen bonds (**Fig. 1E**). In contrast, the chemical shifts of Gly Hα2 and Hα3 nuclei at non-hydrogen bonded loop positions are degenerate (**Fig. 1F**).

Figure 1. shAFP Glycines show NMR Peak Dispersion Related to H-bonding.



A. The structure of sfAFP displayed as sticks colored black for backbone C, light blue for side chain C, red for O and blue for N. Note the long extended shape of the protein. A thick line shows the Cα trace in a rainbow color gradient from N-terminus (indigo) to the C-terminus (red). The protein's two disulfide bonds, which connect C1··C28, and C13··C43, are shown as yellow lines. Some residues are labeled.

B. The shAFP structure as in **A**. but rotated ≈ 45º to highlight the polar (*left*) and hydrophobic (*right*) faces.

C. 2D ^{1}H-^{15}N HSQC spectrum of ^{13}C,^{15}N-labeled sfAFP freshly dissolved in D_2O at 5.0 ºC in 5 mM sodium phosphate buffer containing 10 mM sodium chloride, pH* 6.13. Most hydrogen-bonded ^{1}HN are shifted downfield. The multiplication factor between contours is √2.

D. Strips of the 3D ^{1}HN-^{15}N-^{1}Hα HNHA spectra showing the resonances of representative glycines. The multiplication factor between contours is 1.25. When in PPII helices, one of the two Gly ^{1}Hα peaks is shifted upfield; this is correlated with Cα-Hα |||| O=C hydrogen bond formation as observed in the X-ray crystal structure[11], as illustrated in the following panels. In contrast, glycines in loops, like Gly 12, show degenerate ^{1}Hα2, ^{1}Hα3 resonances.

E. Gly 6 has one Hα positioned to donate a weak H-bond to the carbonyl of A19. This would be the H which shows an upfield shift to 3.19 ppm as shown in panel **D**. Gly 6's other Hα is oriented out of the plane of the page and towards solvent and has a less perturbed chemical shift of 3.75 ppm. Note that the CO of Ala 19 accepts a second H-bond from Gly 34.

F. Both of Gly 12's Hα are not positioned to donate hydrogen bonds and both resonate at the same chemical shift value (3.93 ppm, see panel **D**) which close to the value expected for a statistical coil.

The conformational chemical shifts (Δδ) of sfAFP, that is the experimental chemical shifts relative to values expected for statistical coil, were scrutinized for tendencies indicative of the polyproline II conformation (**Table 1**, **Fig. 2**, **Sup. Table 1 & Sup. Fig. 2**). Comparing residues observed to be in PPII helices versus loops in the X-ray structure[11], significantly lower ^{13}Cα δ values are observed for the former (**Fig. 2A,** see **Sup. Fig 2** for histograms). In contrast, only a weak non-statistically significant trend toward higher $\Delta\delta^{13}$Cβ values is observed for the ^{13}Cβ nuclei in loop position (**Fig. 2B**). $\Delta\delta^{13}$Cβ values close to zero were reported previously for α-helices[46]. Two proline ^{13}Cβ (residues 14 and 71) show high δ values which are consistent with *cis* Cys13-Pro14 and *cis* Thr70-Pro71 peptide bonds in accordance with the crystal structure[11]. On average, the ^{13}CO δ values are close to coil values; however, there is a minor tendency of H-bonded ^{13}CO towards slightly lower ($\Delta\delta^{13}$CO ≈ -0.2 ppm) δ values (**Table 1**, **Fig. 2C**).

As noted above, strikingly disparate ^{1}Hα δ values are observed for 28 glycine residues in polyproline II helices (**Fig. 2D**). Typically, one Gly ^{1}Hα has a chemical shift close to the excepted coil value whereas the other is shifted significantly upfield. This ^{1}Hα putatively participates in Cα-Hα ||| O=C hydrogen bonds, even though the direction of its shift is opposite that observed prevously[47] and here for non-Gly residues whose ^{1}Hα is positioned to donate a hydrogen bond. The differences between these two sets of values are highly significant (**Table 1, Sup. Fig. 2D**). The putatively H-bonded Gly ^{1}Hα's Δδ values are also very different from those values of ^{1}Hα not positioned to donate an H-bond (**Table 1, Sup. Fig. 2D**). We propose these highly contrasting Gly ^{1}Hα2 and ^{1}Hα3 δ values as useful indicators of Gly-rich polyproline II helical bundles.

Regarding the amide groups, those donating an H-bond tend to have higher ^{1}HN δ values (**Fig. 2E, Sup. Fig. 2E**), low temperature coefficients and slowed H/D exchange (*vide infra*). By contrast, non H-bonded ^{1}HN nuclei have δ values close to the predicted coil value (**Fig. 2E**).

This trend was reported previously and the magnitude of the ^{1}HN Δδ inversely correlates with the cube of the acceptor O···H-N distance[47]. A related trend is observed for the ^{15}N δ values (**Fig. 2F**). When the amide group donates an H-bond to a main chain carbonyl, the ^{15}N δ shows a significant shift (**Table 1, Sup. Fig. 2F**) to lower δ values relative to non-H-bonded ^{15}N. This trend is much weaker when the acceptor is a side chain moiety such as a Ser Oγ. Earlier studies linked the decrease in ^{15}N δ to hydrogen bonding[48]; more modern investigations have reported that hydrogen bonding, backbone and side chain orientations, and neighboring residue identity can all perturb ^{15}N δ several ppm[49].

Table 1: sfAFP Shows Significant Conformational Chemical Shifts

Nuclei	Comparison Code[a]	Mean Δδ (ppm)	σ (ppm)[b]	n[c]	p-value
^{13}Cα (helix)		-0.59	0.65	59	0.003
^{13}Cα (loop)		-0.28	0.97	16	
^{13}Cβ (helix)		+0.25	0.74	29	0.16
^{13}Cβ (loop)		+1.18	0.39	9	
^{13}CO (H-bonded)		-0.20	1.12	35	0.34
^{13}CO (not H-bonded)		+0.02	1.31	42	
^{1}Hα (Gly, H-bonded)	**A**	-0.59	0.32	30	< 0.0001 **A/B**
^{1}Hα (non-Gly, H-bonded)	**B**	+0.26	0.31	20	0.0013 **B/C**
^{1}Hα (not H-bonded)	**C**	-0.03	0.20	24	< 0.0001 **C/A**
^{1}HN (H-bonded)		+0.32	0.49	48	0.0105
^{1}HN (not H-bonded)		+0.04	0.26	34	
^{15}N (H-bonded to mainchain OC)	**D**	-1.69	2.35	39	0.16 **D/E**
^{15}N (H-bonded to side chain acceptor)	**E**	-0.17	3.73	8	0.51 **E/F**
^{15}N (not H-bonded)	**F**	+0.74	2.64	24	0.00012 **F/D**

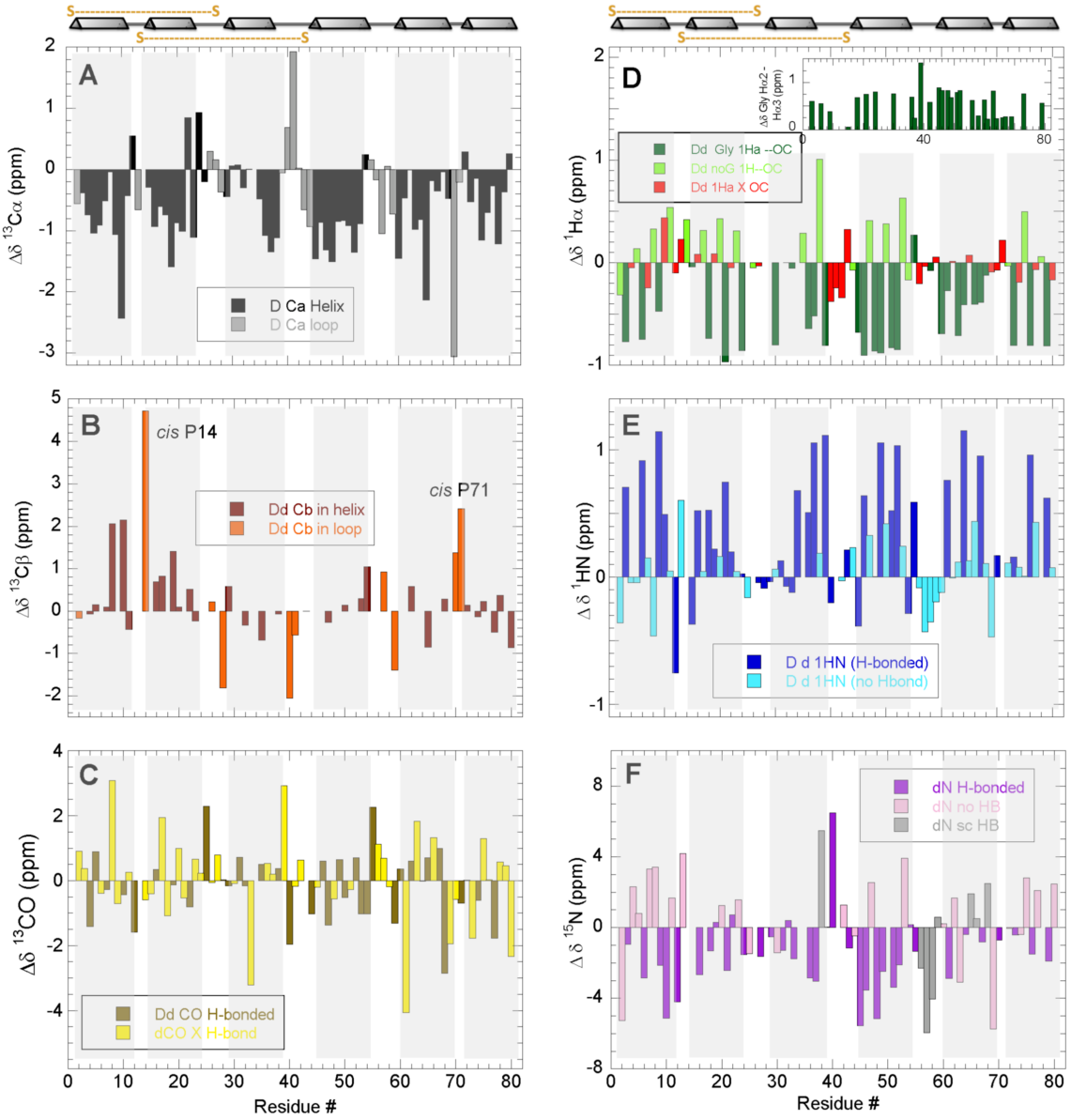
[a] The comparison code clarifies which data sets are being compared in the p-value analysis. For example, **A/B** means the p-value for comparing the data sets **A**: ^{1}Hα (Gly, H-bonded) and **B**: (non-Gly, H-bonded).

[b] σ is one standard deviation.

[c] n refers to the number of observations.

Figure 2

Conformational Chemical Shifts in the sfAFP Polyproline II Helix Bundle



Silver triangular prisms above the panels and light gray shading within the panels mark the positions of the six PPII helices. Golden "S" linked by dashed lines represent the two disulfide bonds.

A. The conformational chemical shift (Δ δ ^{13}Cα) between the observed chemical shift value, and the value predicted for a statistical coil for this nucleus in this sequence at this temperature is shown. Black bars show ^{13}Cα in helices; gray bars show those in loops.

B. Δ δ ^{13}Cβ, maroon = helix; salmon = loops.

C. Δ δ ^{13}CO, H-bonded ^{13}CO are shown in tawny and non H-bonded ^{13}CO are colored yellow.

D. Δ δ ^{1}Hα, putatively H-bonded Gly ^{1}Hα Δ δ are represented by dark green bars; non-Gly H-bonded by light green bars, and non-H-bonded as red bars. The *inset* shows the Δ δ between the two glycine ^{1}Hα (δHα2 – δHα3)

E. Δ δ ^{1}HN, dark blue and light blue bars show the values of H-bonded and non H-bonded amide hydrogen, respectively.

F. Δ δ ^{15}N, purple bars represent ^{15}N participating in H-bonds with main chain carbonyl acceptors, pink = ^{15}N with side chain acceptors and gray = non H-bonds.

J-couplings Are Consistent with the PPII region of the Ramachandran Map: The polyproline II helix occupies a singular region of the Ramachandran map, its ϕ angle is near -78º, akin to that of an α-helix, whereas it has positive ψ angle close to 146º, which is similar to β-strands [50-51]. J-couplings give information on the ϕ,ψ angles, in particular, $^{3}J_{HNH\alpha}$ of 4.8, 8.6 and 6.5 Hz are expected for the average ϕ angles in α-helices, β-strands, and PPII helices, respectively. $^{1}J_{CH\alpha}$ of 146.34, 140.9 and 142.6 Hz are characteristic of the average ψ angles in α-helices, β-strands, and PPII helices, respectively. Utilizing a 3D HNHA spectrum, $^{3}J_{HNH\alpha}$ coupling constants were measured for sfAFP residues (**listed in Sup. Table 2**) and the mean $^{3}J_{HNH\alpha}$ constants for residues in the PPII helices observed in the X-ray structure[11], plotted onto the Karplus equation (**Sup. Fig 3A**) show ϕ angles ranging from -57º to -71º. The $^{1}J_{CH\alpha}$ constants of fifteen residues were estimated from analysis of a non-decoupled ^{1}H-^{13}C HSQC spectrum. Despite some scatter, most $^{1}J_{CH\alpha}$ constants are < 143 Hz (**Sup. Fig. 3B**) and therefore indicate this conformer occupies the permitted region of the Ramachandran map with positive ψ values. Combining the ϕ values for each helix and the overall positive ψ value, the average position of each helix maps to the polyproline II region on the Ramachandran diagram (**Sup. Fig. 3C**).

^{15}N Relaxation Reveal Low Protein Dynamics on ps-ns Time Scales: The hNOE values are generally high for sfAFP and average 0.78 at 800 MHz (**Fig. 3A**) and 0.78 at 600 MHz (**Sup. Fig 4A**). This means that the structure is generally rigid. Similar results were obtained using two different samples and 180º or 120º pulses for ^{1}H saturation (**Sup. Fig. 4B**). In sfAFP, the connections linking the PPII helices are short except for somewhat longer loops between helices 2 and 3 (residues 25 – 28) and helices 4 and 5 (residues 55 - 58). Somewhat lower hNOE ratios, indicative of moderate flexibility, are seen for the short turn between helices 1 and 2. Signficantly lower hNOE values are seen for the former loop and the beginning of next helix (residues 27 - 31) and indicate increased ps – ns dynamics, whereas the second long loop has high hNOE values. It is curious that the loops have relatively higher B-factors, which are indicative of increased thermal motion, in the crystal state (see **Sup. Fig. 4B**). The R_1 and R_2 relaxation rates measured at 800 and 600 MHz (^{1}H) are shown in **Sup. Fig. 4C & 4D, respectively.** The R_1 and R_2 values are generally higher and lower, respectively, for residues 27 – 31 than the average values, which evinces elevated dynamic behavior. This coincides with lower hNOE values for these residues.

The correlation times calculated on the basis of R_1 and R_2 are 9.89 ns at 800 MHz and 9.94 ns at 600 MHz. These τ_c values are much higher than the expected correlation time of 4.0 ns for a 6.5 kDa *globular* protein, and may be due to the elongated form of sfAFP. To investigate this further, we used the program "HydroNMR" and the known X-ray crystal structure to calculate the expected τ_c value for sfAFP at 5.0 ºC. Once the shape is taken into account, the

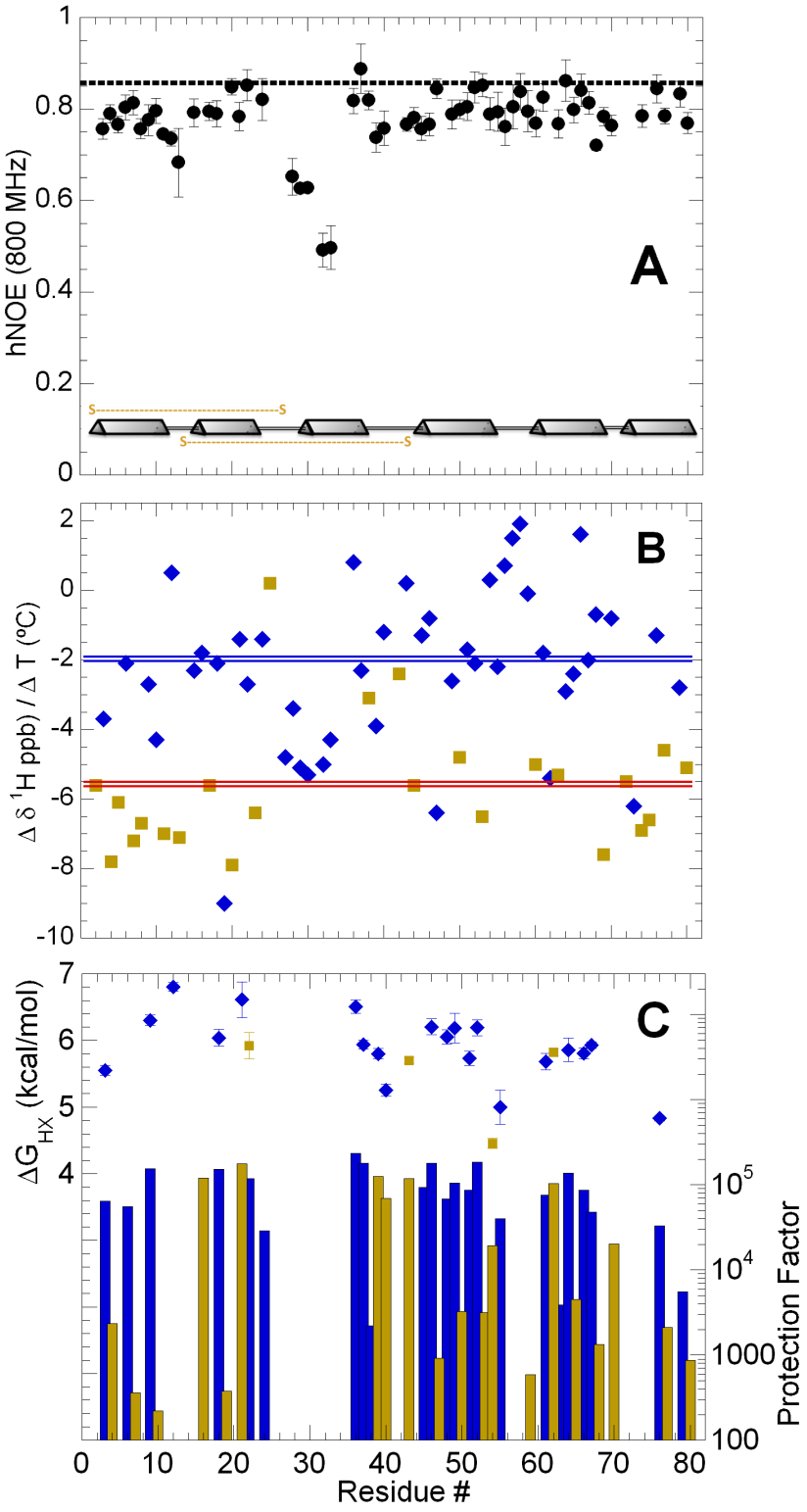
calculated τ_c increases modestly to 4.4 ns, but is still approximately half the experimental value. This suggests that sfAFP may exist as a dimer in solution under these conditions. In the crystal structure 3BOG, two sfAFP monomers are in close contact; taking this assembly as a model for the dimeric form in solution, a value of 8.0 ns for the correlation time was obtained via HydroNMR. This good agreement is evidence that sfAFP is a dimer in our solution conditions. As an additional test, we recorded additional 2D ^{1}H-^{15}N HSQC spectra and compared the chemical shift values for an sfAFP sample at the original concentration (≈ 0.35 mM as dimer, 0.70 mM as monomer) and diluted five fold to approximately 0.070 mM (as dimer; 0.140 mM as monomer). However, no significant ^{1}H chemical shift changes were observed, and very few were larger than 0.010 ppm (*data not shown*). This suggests that sfAFP may form a dimer at both concentrations. Additional ^{1}H-^{15}N longitudinal and transverse relaxation measurements at 65 μM concentration (as monomer) yielded slightly smaller τ_c values relative to 700 μM; this is consistent with a low μM K_D value (data not shown). Along these lines Sosnick and coworkers observed no change in the midpoint of thermal denaturation (T_M) of sfAFP over a similar range of concentration (85 – 300 μM), which they interpreted as evidence that shAFP is a monomer is solution[13]. Nonetheless, due to the dependence of T_M on the logarithm of protein concentration[52] the expected change is small (< 1.1 ºC).

Corroboration of the sfAFP Dimer Using Analytical Gel Filtration and Ultracentrifugation: Upon analysis of sfAFP using gel filtration near 0 ºC, the elution volume of the major peak corresponds closely to the value expected for a dimer (**Sup. Figure 5A**). When subjected to analytical ultracentrifugation, at equilibrium, the weight-average molecular weight is 13044 +/- 1%, which corresponds approximately to a 46% monomer : 54% dimer ratio (**Sup. Figure. 5B**) at the slightly higher temperature of 10 ºC. MALDI-TOF mass spectra (not shown) allow us to conclude that these dimeric species are not stabilized by intermolecular disulfide bonds.

Backbone Amide HN Hydrogen Bonding and Stability: ^{1}HN chemical shift temperature coefficients ($\partial\ \delta^{1}\mathrm{HN}\ /\partial\ \mathrm{T}$ (ppb · K^{-1})) are a well-established, quick and reliable way to characterize HN H-bonding in proteins[53]. Values > -5 ppb·K^{-1} are characteristic of HN groups H-bonded to intramolecular acceptors in α-helices and β-sheets whereas more negative values are typical of HN groups H-bonded to water. In sfAFP, HN groups identified as being H-bonded to protein acceptors in the X-ray structure have less negative temperature coefficients (mean = -2.1 ppb·K^{-1}) whereas those seen to H-bond to water have more negative temperature coefficients (mean = -5.8 ppb·K^{-1}) (**Fig. 3B**). Similar values are observed for HN groups H-bonded to backbone OC moieties or side chain groups (**Sup. Table 2**). These results show that temperature coefficients for PPII helices are similar to those seen in α-helices and β-sheets and that temperature coefficients are a fast yet generally accurate approach to identify intramolecular H-bonding in PPII helical bundles. The residue-level conformational stability measured by H/D exchange at pH* 6.13, 5 ºC is consistently strong in polyproline II helices 1, 2, 3, 4 and 5 and is somewhat lower in the last helix (**Fig. 3C**). It is evident that Gly residues whose HN groups form

H-bonds at internal positions exchange significantly more slowly than non-Gly residues with solvent exposed HN groups. Some protection against H/D exchange is observed for V47, A50 and A53, yet their HN groups seem to lack H-bond acceptors. The HN moieties of several residues; namely V10, H32, N35, G56, T57, G58, V59, R62 and K64 appear to donate hydrogen bonds to side chains acceptor groups (as listed in **Sup. Table 2**). Whereas these H-bonds do correlate with low temperature coefficients (*vide supra*), they afford weak protection against H/D exchange. The global conformational stability, measured by averaging the highest values[54], is 6.24 +/- 0.20 kcal·mol^{-1}. In a second exchange experiment performed at pH* 5.12, similar trends across the helices are observed (**Sup. Figure 6**). However, the global stability is only 5.08 +/- 0.20 kcal·mol^{-1} at pH* 5.12 which is significantly lower than the value measured at pH* 6.13. This may be attributed to the loss of two H-bonds in which neutral His Nδ act as acceptors (listed in **Sup. Table 2**) as well as less favorable electrostatic interactions.

Figure 3: Residue Level NMR Experiments Reveal that sfAFP is Rigid and Has a High Conformational Stability.



A. hNOE ratio measured at 800 MHz using a 180º ^{1}H saturation pulse on the ^{15}N-labeled sample. The data and error bars shown are the average values from three experiments. The results from each experiment are shown in **Sup. Fig. 4B**. Silver triangular prisms mark the positions of the six PPII helices. Golden "S" linked by dashed lines represent the two disulfide bonds.

B. Per residue temperature coefficients for H-bonded (blue diamonds) and non H-bonded HN groups (brown squares). Blue and red double lines mark the mean values for H-bonded and non H-bonded HN groups, respectively. The experimental uncertainties are about ≈ ±14%, as calculated by propagation of the uncertainties of the ^{1}HN δ values.

C. Per residue conformational stability (left y-axis) for slow exchanging Glycine (blue diamonds) and non-Glycine HN groups (brown squares). The experimental uncertainties are shown as error bars and represent one σ as calculated by propagating the uncertainties in the fitted experimental H-> D exchange rates. Blue and brown bars (right y-axis) show hydrogen exchange protection factors for all HN groups with detectable protection.

Structural Modeling Suggests the Hydrophobic Effect and Hydrogen Bonding Drive sfAFP Dimerization: Some HN groups; namely G25, V47, A50, A53 and A80 seem to lack H-bond acceptors in the X-ray structures of sfAFP, yet they show significant protection against H/D exchange or ^{1}HN temperature coefficients > -4 ppB·K^{-1}. As this could result from burial at a monomer/monomer interface, these groups were employed as input to HADDOCK to model a dimeric structure for sfAFP. Three independent runs of the program yielded remarkably similar dimers in which the long, flat hydrophobic surfaces of PPII helices 2, 4 and 6 of one monomer contact helices 6, 4 and 2, respectively, in the other subunit. The tentative structural model depicted in **Sup. Figure 7** and the **Supporting Movie** corresponds to the first structure of the top cluster, which we consider a reliable solution based on the following parameters: HADDOCK score: -92.4 ± 0.3; RMSD from the overall lowest-energy structure: 0.4 ± 0.3 Å (for this cluster's 84 structures); Energies (kcal·mol^{-1}): van der Waals = -71.3 ± 2.9; electrostatic = -54.5 ± 3.6; desolvation = -11.2 ± 0.3; restraints violation = 9.4 ± 2.23; Z-score: -1.5. The HADDOCK score is given in arbitrary units, as it contains additional empirical terms (*e.g.* buried surface area). Whereas the structural model is also supported by the observation of dimer specific NOE signals between V23 Hγ – HN G79 and A19 Hβ – HN A80, its rigorous validation requires additional experiments and analysis. In this structural model, 1760 square angstroms of surface area, including approximately 20 methyl or methylene moieties, would become buried at the dimer interface. Six H-bonds form between the HN of V47, A50 and A53 from each subunit with backbone O=C acceptors in the opposite subunit, which accounts for these groups' slow H/D exchange and positive ^{1}HN Δδ values. Thus dimerization is driven by the hydrophobic effect and hydrogen bonding.

DFT Calculations Corroborate Gly Hα Hydrogen Bonding and Afford Stereospecific Assignments. As stated above, the two ^{1}Hα of each of the 28 Gly residues depicted in **Fig. 2D** possess very different chemical shifts (δ), with one of which is always shifted to δ values lower than expected for statistical coil. We attribute this observation to the involvement of such ^{1}Hα in a CαHα ||| O=C H-bonding interaction. In an effort to understand how the formation of such an attractive interaction could result in the observed shielding the ^{1}H nuclei, we resorted to DFT

calculations to evaluate the magnetic shielding tensors upon CαHα ||| O=C H-bond formation. To this end, we first considered a minimal system consisting of three segments containing two of the putative CαHα ||| O=C bonds between four glycine residues. This model system is based on the X-ray structure of sfAFP (PDB ID: 3BOG)[11] in which three helices are selected, and acetylated or amidated N- and C-terminal positions, respectively, to avoid electrostatic bias from charged termini (**Fig. 4A-D**). The central segments contained Hα ||| O=C H-bonded residues involving Gly33 Hα (PPII helix # 3) to Gly49 O=C (PPII helix # 4), and Gly64 Hα (PPII helix # 5) to Gly34 O=C (PPII helix # 3), as depicted in **Fig. 4A**.

From this model system, we obtained and examined the magnetic shielding tensors of Gly33 and Gly64 Hα nuclei in the presence and absence of O=C acceptor counterparts (Gly49 and Gly34, respectively). These results showed that of the four Hα nuclei in both Gly33 and Gly64, only the two participating in the H-bonding interaction show large variations in isotropic magnetic shielding and increased chemical shift anisotropy (CSA). This is reminiscent of changes in amide ^{1}HN CSA upon H-bonding formation[55]. Particularly, the ^{1}Hα CSA value of the ^{1}Hα nuclei involved in the H-bonding interaction increases in both Gly33 and Gly64.

These variations in ^{1}Hα CSA and nuclear magnetic shielding upon H-bond formation prompted us to further explore the "H-bond" nature of these interactions. To this end, we utilized Natural Bonding Orbital (NBO) analyses, which provide a more intuitive picture of the electron density surrounding the ^{1}Ha nuclei that forming H-bonds. Within this representation, an H-bond features an empty antibonding sigma orbital (σ*) onto which a lone pair electron (LP) from the O atom delocalizes. In other words, an H-bond involves an LP(O) → σ*(N–H) electron delocalization (orbital overlap). We argue that here, one should expect orbital overlap of the type LP(O) → σ*(Cα–Hα), in analogy to LP(O) → σ*(N–H). Using second order perturbation theory analyses as implemented in the NBO software, we found that this electron delocalization is akin to that of "conventional" H-bonds is present in both Gly33 Hα ||| O=C Gly49, and Gly64 Hα ||| O=C Gly34 (**Fig. 4E**). For the carbonyl oxygen in both cases studied here, the lone pair of electrons comes from the molecular orbital with significant s character (60% s, 40% p). Although individual Hα ||| O=C hydrogen bonds are weak in nature (**Sup. Figure 8)**, the concurrence of 28 bonds in sfAFP, as detected by NMR ^{1}H chemical shift displacements, significantly contributes to its stability.

Finally, we have extended the DFT calculations to obtain theoretical chemical shift values for the full length (residues 1 – 81) protein. The complete results, shown in **Sup. Table 1**, afford stereo specific assignments of all the Gly ^{1}Hα involved in H-bonds and corroborates the displacement of their chemical shifts to lower values.

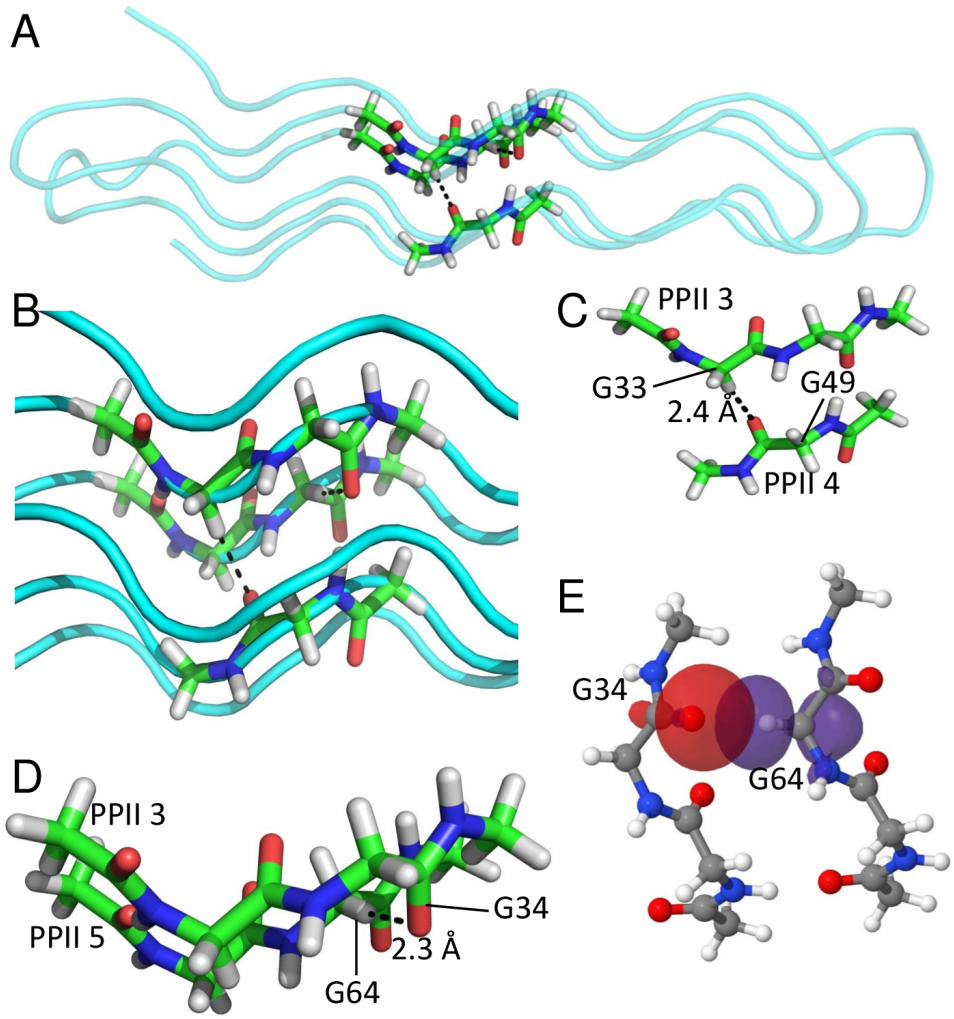


Figure 4: Computational Analysis of Glycine Cα-H ||| O=C Hydrogen Bonds.

A. View showing the complete sfAFP structure (cyan) and the model (green) used to evaluate the putative Cα-H ||| O=C H-bonds.

B. Zoom view of the structural model, highlighting the H-bonded Gly residues from three different helices.

C. G33 in PPII helix 3 donates a Hα to form a H-bond with the O=C group form G49 (in helix 4). The distance between hydrogen and oxygen is indicated (2.4 Å).

D. G64 in PPII helix 5 donates a Hα to form a H-bond with the O=C group form G34 (in helix 3). The distance between hydrogen and oxygen is indicated (2.3 Å).

E. NBO analysis confirms the H-bonding nature of the CαHα ||| O=C interactions. The previous interaction between G34 C=O and G64 Hα shown in **panel D** is represented here using NBOs: in red, the lone pair (LP) from the carbonyl oxygen of G34. In purple, the Cα - Hα σ antibonding. The orbital overlap indicates that electron density from the O lone pair (red) fills the Hα's σ* orbital (purple).

Numerous human proteins may contain glycine-rich PPII helical bundles: Using the MOTIF search algorithm, we found 1167 human proteins with a GGXGG segment (where G is Gly and X is any residue) which might form a short Gly-rich PPII helix and 265 human proteins with a longer segment, GGXGGXGG, which possibly could form a longer PPII helix. In almost all cases (257 of 265), both Xs are non-glycine residues so these are not simply long stretches of

Gly residues. Many of these proteins are putative structural proteins; others, such as MasterMind-like transcription coactivator 3, are involved in signaling, cell differentiation and viral infection. Nucleolin is a protein present in one of the distinct, non-aqueous liquid phases of the nucleolus[56]. Although the C-terminal domain Nucleolin is assumed to be intrinsically disordered; based on its content of GGXGGXGG motifs:

(E_{650}GGFGGRGGGRGGFGGRGGGRGGRGGFGGRGRGGFGGRGGFRGGRGGGGDH$_{700}$),

in residues 650 – 700, we identified a domain which are strictly conserved in mammals with sequence identities = 100% for the 98 top hits and E-scores < $1 \cdot 10^{-15}$. This strong evolutionary conservation is evidence that this domain performs an important structural/functional role and as a speculative hypothesis we propose that it may fold into a bundle of up to five PPII helices. Whereas Cys residues are lacking, these helices might be stabilized by cation – π interactions as many Arg and Phe residues are spaced to interact as there are exactly three residues per turn of the PPII helix.

DISCUSSION

The results presented here are the first thorough NMR study of a polyproline II helical bundle. An NMR study of the peptide YGRKKRRQRRRP advanced $\Delta\delta^{15}$N values of +1.1 ppm and intermediate values of $^{3}J_{HNHA}$ as criteria for identifying isolated PPII helices[8]. In contrast, we find that PPII helical bundles are characterized by $\Delta\delta^{13}$C = -1 ppm, highly distinct Gly Hα2 and Hα3 δ values, $\Delta\delta^{1}$HN = +1 ppm, temperature coefficients of -2 ppb·K^{-1}, high hNOE ratios and greatly slowed backbone amide H/D exchange (see **Table 1** and **Sup. Table 2**). Of these criteria, the temperature coefficients and $\Delta\delta$ ^{1}HN chemical shifts are particularly handy as they can be easily obtained from 2D ^{1}H-^{15}N HSQC spectra. For this reason, they could be used as quick yet accurate tools to verify PPII helical bundle formation in the hundreds of human proteins containing this conformation's hallmark GGXGGX motif.

Previous liquid state NMR investigations of polyproline II helices focused on a dodecamer peptide forming a partially populated, monomeric PPII helix in equilibrium with coil [8], a cleverly designed polypeptide consisting of a polyproline II helix packed against an α-helix[57], as well as collagen analogs consisting of a three helix bundle composed of three thirty residue (Pro-hydroxyPro-Gly)$_{10}$ peptides[6]. A solid state ^{13}C NMR study has investigated side chains conformational changes collagen fibers[58]. The first system is characterized by $\Delta\delta^{13}$Cα, $\Delta\delta^{1}$Hα, $\Delta\delta^{1}$HNα, $\Delta\delta^{13}$CO values close to zero and $\Delta\delta^{15}$N ≈ +1.1 ppm [8]. These values are consistent with the amide group being hydrogen-bonded to water and are quite different from the results reported here for a completely populated PPII helical bundle. This dodecapeptide (sequence: Y**G**RKKRRQRRRP) has one Gly whose two

Hα have degenerate chemical shift values = 3.97 ppm. Different Δδ trends, including distinctly positive $\Delta\delta^{13}C\alpha$ were also observed for the PPII helix in the PPII helix/α-helix miniprotein[57] (**Sup. Fig 9**). This system contains one Gly (G9) in the PPII conformation and the δ difference between its two Hα is 4.017 – 3.883 ppm = 0.134 ppm (**Sup. Fig 9**). The collagen analog study cited above reported Δδ Gly Hα2 – Hα3 values of 0.10 to 0.11 ppm. Additional studies on protein segments adopting relatively high populations of isolated PPII helix have shown the utility of residual dipolar couplings in detecting PPII helix[59] and ^{1}Hα δ values in the range expected for statistical coil[60] .

In these PPII systems[6, 8, 57, 60], the Δδ Gly Hα2 – Hα3 values are much smaller than those reported here for sfAFP's PPII helical bundle (**Fig. 1 D & Fig. 2D inset**). Previous work from this laboratory[61] and others[47, 62] has shown that ^{1}Hα conformational chemical shifts are chiefly determined by magnetic anisotropy and electric effects, *i.e.* hydrogen bonding. The main source of magnetic anisotropy in sfAFP, which has few aromatic groups, are carbonyl groups, whose contribution to ^{1}Hα's δ depends strongly on the ϕ dihedral angle and weakly on ψ (see Figure 8 in ref. [62]). Considering that ϕ and ψ are approximately the same in all PPII helices, the contribution of carbonyl anisotropy to conformational chemical shifts will be similar in simpler PPII systems (whose Δδ Gly Hα2 – Hα3 ≤ 0.134 ppm) as well as the PPII helical bundle considered here. Therefore, the striking difference in the Hα2 & Hα3 chemical shifts likely arises mainly from hydrogen bonding effects as corroborated here by DFT and NBO computational analyses (**Fig. 4**).

The Chiefly Rigid Nature of sfAFP Suggests Stabilization by Unanticipated Factors. It is remarkable that sfAFP folds despite the large unfavorable entropy change experienced by Gly residues during this process. The global conformational stability and folding kinetics of sfAFP have been recently studied by Sosnick's laboratory[13] and they attributed this protein's perplexingly high conformational stability and rigidity, which resemble those of globular proteins[63-64], to strong N-H ||| O=C H-bonds and some contributions from relatively weak Cα-H ||| O=C H-bonds, whose existence we corroborate here. Cα-H ||| O=C H-bonds have also been detected in globular proteins and collagen by Derewenda *et al.*[65] and Bella and Berman[66], respectively, who argued that these H-bonds contribute to conformational stability. Our HN relaxation measurements clearly show low dynamics for sfAFP on ps-ns time scales; confirming that it is as rigid as typical globular proteins[67] and antifreeze proteins rich in beta sheet[68]. In addition, the burial of hydrophobic surface area at the dimer interface and inter-subunit hydrogen bonds likely contribute to the conformational stability.

The H/D exchange measurements reported here show high stability for the first 5 helices with lower values for the last helix. This could be due to helix 6 making fewer interhelix contacts and H-bonds as it contains three proline residues and it lacks the disulfide bonds that stabilize helices 1, 2, 3 and 4. Since less stable elements of secondary structure generally exchange through partial unfolding events[69-70], these results suggest that the last helix may

unfold autonomously before the rest of the protein, which is in line with the results of the folding kinetics experiments previously reported[13].

The conformational stability of sfAFP has been studied by Sosnick and co-workers[13] at pH 8.0 in 10 mM phosphate buffer using thermal denaturation and guanidinium chloride equilibrium denaturation experiments and kinetic measurements, and they obtained values ranging from 5.92 ± 0.56 kcal·mol^{-1} (thermal denaturation extrapolated to very fast scan rates), 5.33 ± 0.1 kcal·mol^{-1} (kinetics fit to a two state model) and 6.43 ± 0.58 (kinetics analyzed with a three state model). Differences in the global conformational stability measurements by NMR-monitored hydrogen exchange (reported here to be 6.24 ± 0.20 kcal·mol^{-1}) and equilibrium measurements cited above can be attributed to four factors: 1) proline *cis/trans* isomerization, 2) different pH, 3) distinct salt concentrations, and 4) a deuterium isotope effect. In particular, Xaa-Pro *cis/trans* peptide bond isomerization, which occurs in equilibrium denaturation experiments but not H/D exchange conditions, destabilizes the former relative to the latter. The effect is especially important in proteins, like sfAFP which contain *cis* Xaa-Pro peptide bonds, and can be calculated[54, 71]. It equals 2.91 kcal·mol^{-1} in the case of sfAFP under our experimental conditions (**Sup. Table 3**). By extrapolating measurements to infinitely fast scan speeds, as Sosnick and co-workers have done, the effect of proline isomerization should be minimal.

Regarding the other factors, lowering the pH from 8.0 (Sosnick *et. al.*) to pH* 6.13 (this study) would tend to destabilize the protein by weakening two hydrogen bonds where His side chains where participate as acceptors[11] (**Sup. Table. 2**), but could reduce electrostatic repulsion within the folded protein. Whereas differences in ionic strength between the two studies are minimal, the presence of D_2O in our H/D exchange experiments could well be stabilizing, considering that D_2O has been reported to stabilize isolated PPII helices[72]. Despite these additional factors, the values measured by different approaches are in fairly good agreement and point to a conformational stability around 6 kcal·mol^{-1}. Such a high conformational stability is truly remarkable considering the highly unfavorable entropic cost of fixing Gly residues[12], and this protein's very high Gly content (>45%)[10].

The formation of a dimer, evidenced by NMR relaxation, gel filtration, ultracentrifugation data and computational modeling features contacts between the nonpolar faces of two sfAFP monomers. The burial of about twenty methyl/methylene groups and the formation of six backbone H-bonds favor dimerization. However, monomeric sfAFP is stable as it successfully refolds at high dilution where the sfAFP dimer population would be low. Moreover, if the sfAFP monomer were unfolded, the observed protection factors from the H/D exchange experiments could not exceed the [monomer]:[dimer] ratio, which is < 20. The observation of protection factors much higher than 20 proves that the sfAFP monomer is stable on its own. The predominance of dimer at sfAFP concentration of 30 μM is consistent with a K_D in the low micromolar range. In biological context, this medium strength dimerization would be strong enough to help avoid interaction and aggregation with other hydrophobic components in the cell yet weak enough to allow the subunits to separate and interact with ice.

Broader Implications: Polyproline II helices are important in numerous biological systems[73]. The sequence analysis reported here suggests that there could be hundreds of glycine-rich PPII helices and even bundles in the human proteome. This is the first report, to our knowledge, describing NMR parameters for a polyproline helical bundle, which, together with previous investigations of isolated PPII helices or PPII helices combined with other types of secondary structure, will aid future progress on several important problems. One such problem is understanding how polyproline segments inhibit amyloid formation by polyQ tracts in Huntingtin protein[4, 74-75], the Androgen receptor[76], Tau[77] and potentially CPEB3, a functional amyloid very rich in proline whose fibrillation is essential for human memory consolidation[78]. In addition, these results could have implications for the role of glycine rich regions in TDP-43[79], FUS/TLS and nucleolin to promote liquid/liquid phase separation.

Improving programs for detecting and quantifying the amount of PPII conformation in peptides and proteins based on conformational chemical shifts is another goal. Currently only the program δ2D from the laboratory of Vendruscolo[80], predicts the population of PPII conformation from chemical shifts. When applied to the chemical shifts for shAFP's PPII helix 2 reported here, this program predicted a population of < 35 % PPII conformation. Future advances could also help characterize new polyproline II helical bundles, such as the fascinating core of acetophonene carboxylase, which is predicted to consist of six PPII helices surrounding a seventh central helix composed almost exclusively of Gly residues[81].

sfAFP is a much more potent antifreeze than single α-helix antifreeze peptides and a longer, even more potent version of sfAFP with additional PPII helices has been described[82]. Antifreeze proteins have important practical applications in farming cold-resistant salmon, preserving food and organs for transplant, preventing frost damage to crops, improving the texture of ice cream, *etc* [83]. The improved understanding of sfAFP structure and stabilizing interactions afforded by the results reported here could aid the design of more stable and active anti-freeze proteins.

Finally, following the pioneering work from Woolfson's, Raines' and Zondlo's laboratories, these results could facilitate the design and production of novel proteins containing PPII helices[57] whose conformation can by controlled by steric[84] and electronic[85] effects or proline editing[86].

Associated Content

Supporting Information

This material is available free of charge on the ACS Publications website.

Tables listing sfAFP H-bonds, sfAFP NMR parameters and proline contributions to stability; figures reporting additional NMR spectra, relaxation data, J-couplings, statistics and a dimeric structural model, additional methodology; and supporting references (PDF).

Acknowledgements. We are grateful to Dr. MªÁngeles Jiménez, Dr. Mariano Carrión-Vázquez and Prof. Robert L. Baldwin for critically reading this manuscript and encouragement. This work was supported by grant SAF-2016-76678-C2-2-R (DVL), CTQ2017-84371-P (MªÁJ),

BFU2015-70052-R (MM), CTQ2014-54987-P (MVG), CTQ2017-84825R (MVG) and the Ramón y Cajal Fellowship: RYC2007-14768 (MVG). Additional funding was provided by the CIBER de Enfermedades Respiratorias (CIBERES), an initiative of the Instituto de Salud Carlos III (ISCIII). We are grateful to Prof. Giovanni Gotte (University of Verona) and Prof. S. Padmanabhan (IQFR, CSIC) for generous gifts of RNase A C-dimer and 434 Cro, respectively. We thank Mª Victoria López Moyano, Sergio Camero Gigante and Plácido Galindo Iranzo for expect technical assistance with analytical ultracentrifugation, NMR spectroscopy and mass spectrometry, respectively. We would like to dedicate this article to our mentor Prof. Marta Bruix Bayés, head for many years of the "Manuel Rico" NMR Spectroscopy Laboratory, on the occasion of her retirement.

Conflict of interest. The authors declare that they have no conflict of interest.

References

(1) Buratti, E.; Baralle, F. E., TDP-43: Gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem Sci.* **2012,** *37*, 237-247.

(2) Shi, Z.; Anders Olson, C.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R., Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **2002,** *99* (14), 9190-9195.

(3) Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A., Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (6), 1744-1749.

(4) Darnell, G. D.; Derryberry, J.; Kurutz, J. W.; Meredith, S. C., Mechanism of *cis*-inhibition of polyQ fibrillation by polyP: PPII oligomers and the hydrophobic effect. *Biophys. J.* **2009,** *97*, 2295-2305.

(5) Chellgren, B. W.; Creamer, T. P., Short sequences of non-proline residues can adopt the polyproline II helical conformation. *Biochemistry* **2004,** *43* (19), 5864-5869.

(6) Acevedo-Jake, A. M.; Jalan, A. A.; Hartgerink, J. D., Comparative NMR analysis of collagen triple helix organization from N- to C-termini. *Biomacromolecules* **2015,** *16*, 145-155.

(7) Ma, K.; Kan, L.; Wang, K., Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. *Biochemistry* **2001,** *40*, 3427-2438.

(8) Lam, S. L.; Hsu, V. L., NMR identification of left-handed polyproline type II helices. *Biopolymers* **2003,** *69*, 270-281.

(9) Sreerama, N.; Woody, R. W., PolyPro(II) helices in globular proteins: Identification and circular dichroic analysis. *Biochemistry* **1994,** *33*, 10022-10025.

(10) Graham, L. A.; Davies, P. L., Glycine-rich antifreeze proteins from snow fleas. *Science* **2005,** *310*, 461.

(11) Pentelute, B. L.; Gates, Z. P.; Tereshko, V.; Dashnau, J. L.; Vanderkooi, J. M.; Kossiakoff, A. A.; Kent, S. B. H., X-ray structure of snow flea antifreeze protein determined by

racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc.* **2008,** *130*, 9695-9701.

(12) D'Aquino, J. A.; Gómez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E., The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **1996,** *25* (2), 142-156.

(13) Gates, Z. P.; Baxa, M. C.; Yu, W.; BRiback, J. A.; Li, H.; Roux , B.; Kent, S. B. H.; Sosnick, T. R., Perplexing cooperative folding and stability of a low-complexity, polyproline 2 protein lacking a hydrophobic core. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (9), 2241-2246.

(14) Kang H; Vázquez FX; Zhang L; Das P; Toledo-Sherman L; Luan B; Levitt M; Zhou, R., Emerging β-sheet rich conformations in supercompact Huntingtin exon-1 mutant structures. . *J Am Chem Soc.* **2017,** *139* (26), 8820-8827.

(15) Mangubat-Medina, A.; Martin, S.; Hanaya, K.; Ball, Z., A Vinylogous photocleavage trategy allows direct photocaging of backbone amide structure. *J Am Chem Soc.* **2018,** *140*(27)*, 8401-8404*

(16) Merg, A. D.; Boatz, J. C.; Mandal, A.; Zhao, G.; Mokashi-Punekar, S.; Liu, C.; Wang, X.; Zhang, P.; Van der Wel, P. C; Rosi, N. L., Peptide-directed assembly of single-helical gold nanoparticle superstructures exhibiting intense chiroptical activity. *J Am Chem Soc.* **2016,** *138* (41), 13655-13663.

(17) Martin, E,W; Holehouse AS; Grace CR; Hughes A; Pappu RV; Mittag, T., Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J Am Chem Soc.* **2016,** *138* (47), 15323-15335.

(18) Shi L; Holliday AE; Shi H; Zhu F; Ewing MA; Russell DH; Clemmer, D. E., Characterizing intermediates along the transition from polyproline I to polyproline II using ion mobility spectrometry-mass spectrometry. *J Am Chem Soc.* **2014,** *136* (36), 12702-11.

(19) Delaforge, E.; Kragelj, J.; Tengo, L.; Palencia, A.; Milles S; Bouvignies, G.; Salvi, N.; Blackledge, M.; Jensen, M. R., Deciphering the dynamic interaction profile of an intrinsically disordered protein by NMR exchange spectroscopy. *J Am Chem Soc.* **2018,** *140* (3), 1148-1158.

(20) Pentelute, B. L.; Gates, Z. P.; Dashnau, J. L.; Vanderkooi, J. M.; Kent, S. B. H., Mirror image forms of snow flea antifreeze protein prepared by total chemical synthesis have identical antifreeze activities. *J. Am. Chem. Soc.* **2008,** *130* (30), 9702-9707.

(21) Kuiper, M. J.; Lankin, C.; Davies, P. L., Purification of antifreeze proteins by adsorption to ice. *Biochem. Biophys. Res. Commun.* **2003,** *300*, 645-648.

(22) Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q., Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **2009,** *18* (5), 936-948.

(23) Anthis NJ; Clore, G. M., Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **2013,** *22* (6), 851-858.

(24) Vuister, G. W.; Bax, A., Quantitative J correlation: a new approach for measuring homonuclear J(HN-Ha) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **1993,** *115*, 7772-7777.

(25) Kjaergaard, M.; Poulsen, F. M., Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR.* **2011,** *50* (2).

(26) Kjaergaard, M.; Brander, S.; Poulsen, F. M., Random coil chemical shifts for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* **2011,** *49* (2), 139-149.

(27) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; J., C.; Wright, P. E.; Dyson, H. J., Sequence-dependent correction of random coil NMR chemical shifts. *JACS* **2001,** *123* (12), 2970-2978.

(28) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D., ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of common amino acids. I Investigations of nearest neighbor effects. *J. Biomol. NMR* **1995,** *5*, 67-81.

(29) Renner, C.; Schleicher, M.; Moroder, L.; Holak, T. A., Practical aspects of the 2D ^{15}N-{^{1}H}-NOE experiment. *J. Biomol. NMR* **2002,** *23* (1), 23-33.

(30) Ferrage, F.; Cowburn, D.; Ghose, R., Accurate sampling of high-frequency motions in proteins by steady state ^{15}N-{^{1}H} nuclear Overhauser effect measurement in the presence of cross-correlated relaxation. *J. Am. Chem. Soc.* **2009,** *131* (17), 6048-6049.

(31) Garcia de la Torre, J.; Huertas, M. L.; Carrasco, B., HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **2000,** *147*, 138-146.

(32) Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **1989,** *28* (23), 8972-8979.

(33) Rossi P; Swapna, G. V.; Huang, Y. J.; Aramini, J. M.; Anklin, C.; Conover, K.; Hamilton, K.; Xiao, R.; Acton, T. B.; Ertekin, A.; Everett, J. K.; Montelione, G. T., A microscale protein NMR sample screening pipeline. *J Biomol NMR.* **2010,** *46* (1), 11-22.

(34) Gotte, G.; Bertoldi, M.; Libonati, M., Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. *Eur J. Biochem.* **1999,** *265* (2), 680-687.

(35) Padmanabhan S; Jiménez MA; Gonzalez C; Sanz JM; Giménez-Gallego G; Rico, M., Three-dimensional solution structure and stability of phage 434 Cro protein. *Biochemistry* **1997,** *36* (21), 6424-6436.

(36) Varea, J.; Monterroso, B.; Sáiz, J. L.; López-Zumel, C.; García, J. L.; Laynez, J. L.; García, P.; Menéndez, M., Structural and thermodynamic characterization of Pal, a phage natural chimeric lysin active against pneumococci. *J. Biol. Chem.* **2004,** *279*, 43697–43707.

(37) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L., In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, Harding, S. E.; Rowe, A. J.; Horton, J. C., Eds. Royal Soc. Chem.: Cambridge, UK, 1992; pp 90 - 125.

(38) Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc.* **2003,** *125*, 1731-1737.

(39) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.;

Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas, O.; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. *Gaussian 09*, Wallingford, CT, 2009.

(40) Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

(41) Ditchfield, R., Self consistent perturbation theory of diamagnetism. *Mol. Phys.* **1974,** *27* (4), 789-807.

(42) Wolinsky, K.; Hinton, J.; Pulay, P., Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **1990,** *112*, 8251-8266.

(43) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO*, 3.1; Gaussian Inc.: Pittsburgh, USA, 2003.

(44) Ogiwara, A.; Uchiyama, I.; Takagi, T.; Kaneshisa, M., Construction and analysis of a profile library characterizing groups of structurally known proteins. *Prot. Sci.* **1996,** *5* (10), 1991-1999.

(45) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990,** *215* (3), 403-410.

(46) Wang, Y.; Jardetzky, O., Probability-based protein secondary structure identification using combined NMR chemical shift data. *Prot. Sci.* **2002,** *11* (4), 852-861.

(47) Wagner, G.; Pardi, A.; Wüthrich, K., Hydrogen bond length and ^{1}H NMR chemical shifts in proteins *J. Am. Chem. Soc.* **1983,** *105* (18), 5948-5949.

(48) Bachovchin, W. W., ^{15}N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: Evidence for a moving histidine mechanism. *Biochemistry* **1986,** *25*, 7751-7759.

(49) Xu, X. P.; Case, D. A., Probing multiple effects of ^{15}N, ^{13}Cα, ^{13}Cβ and 13C' chemical shifts in peptides using density functional theory. . *65* **2002**, 408-423.

(50) Cohen, P. M.; McGavin, S., Structure of poly-L-proline. *Nature* **1955,** *176* (4480), 501-503.

(51) Stapley, B. J.; Creamer, T. P., A survey of left-handed polyproline II helices. *Prot. Sci.* **1999,** *8*, 587-595.

(52) Pace, C. N.; McGrath, T., Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.* **1980,** *255*, 3862-3865.

(53) Cierpicki, T.; Otlewski, J., Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* **2001,** *21*, 249-261.

(54) Huygues-Despointes, B. M. P.; Pace, C. N.; Englander, S. W.; Scholtz, J. M., Measuring the conformational stability of a protein by hydrogen exchange. *Meth. Mol. Biol.* **2001,** *168*, 69-72.

(55) Yao, L.; Grishaev , A.; Cornilescu, G.; Bax, A., The impact of hydrogen bonding on amide ^{1}H chemical shift anisotropy studied by cross-correlated relaxation and liquid crystal NMR spectroscopy *J. Am. Chem. Soc.* **2010,** *132*, 10866-10875.

(56) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P., Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **2016,** *165*, 1686-1697.

(57) Baker, E. G.; Williams, C.; Hudson, K. L.; Bartlett, G. J.; Heal, J. W.; Porter Goff, K. L.; Sessions, R. B.; Crump, M. P.; Woolfson, D. N., Engineering protein stability with atomic precision in a monomeric miniprotein. *Nature Chemical Biology* **2017,** *13*, 764-770.

(58) De Sa Peixoto, P.; Laurent, G.; Azaïs, T.; Mosser, G., Solid state NMR study reveals collagen I structural modifications of amino acid side chains upon fibrillogenesis. *J. Biol. Chem.* **2012,** *288* (11), 7528-7535.

(59) Morgan, J. L.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Barbar, E., The LC8 recognition motif preferentially samples polyproline II structure in its free state. *Biochemistry* **2017,** *56* (35), 4656-4666.

(60) Krieger, J. M.; Fusco, G.; Lewitzky, M.; Simister, P. C.; Marchant, J.; Camilloni, C.; Feller, S. M.; De Simone, A., Conformational recognition of an intrinsically disordered protein. *Biophys J.* **2014,** *106* (8), 1771-1778.

(61) Blanco, F. J.; Herranz, J.; González, C.; Jiménez, M. A.; Rico, M.; Santoro, J.; Nieto, J. L., NMR chemical shifts: A tool to characterize distortion of peptides and protein helices. *J. Am. Chem. Soc.* **1992,** *114* (9676 - 9677.).

(62) Ösapay, K.; Case, D. A., A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* **1991,** *113*, 9436-3444.

(63) Privalov, P. L., Stability of proteins: small globular proteins. *Adv. Prot. Chem.* **1979,** *33*, 167-241.

(64) Pace, C. N., The stability of globular proteins. *CRC Crit Rev. Biochem.* **1975,** *3* (1), 1-43.

(65) Derewenda, Z. S.; Lee, L.; Derewenda, U., The occurrence of C-H···O=C hydrogen bonds in proteins. *J. Mol. Biol.* **1995,** *252*, 248-262.

(66) Bella, J.; Berman, H. M., Crystallographic evidence for Cα-H ··· O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **1996,** *264*, 734-742.

(67) Mompeán, M.; Romano, V.; Pantoja-Uceda, D.; Stuani, C.; Baralle, F. E.; Buratti, E.; Laurents, D. V., The TDP-43 N-terminal domain structure at high resolution. *FEBS J.* **2016,** *283* (7), 1242-1260.

(68) Daley, M. E.; Spyracopoulos, L.; Jia, Z.; Davies, P. L.; Sykes, B. D., Structure and dynamics of a beta-helical antifreeze protein. *Biochemistry* **2002,** *41* (17), 5515-5525.

(69) Li, R.; Woodward, C., The hydrogen exchange core and protein folding. *Prot. Sci.* **1999,** *8*, 1571-91.

(70) Englander, S. W.; Mayne, L. E.; Kan, Z. Y.; Hu, W., Protein Folding - How and Why: By hydrogen exchange, fragment separation and mass spectrometry. *Annu. Rev. Biophys.* **2016,** *45*, 135-152.

(71) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G., Side-chain effects of peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **1998,** *279*, 449-460.

(72) Chellgren, B. W.; Creamer, T. P., Effects of H_2O and D_2O on Polyproline II Helical Structure. *J. Am. Chem. Soc.* **2004,** *126*, 14734-14735.

(73) Adzhubei, A. A.; Sternberg, M. J.; Makarov, A. A., Polyproline-II helix in proteins: structure and function. *J Mol Biol.* **2013,** *425* (12), 2100-2132.

(74) Baias, M.; Smith, P. E. S.; Shen, K.; Joachimiak, L. A.; Zerko, S.; Kozminski, W.; Frydman, J.; Frydman, L., Structure and dynamics of the huntingtin exon-1 N-terminus: A solution NMR perspective. *J. Am. Chem. Soc.* **2017,** *139*, 1168-1176.

(75) Posey, A. E.; Ruff, K. M.; Harmon, T. S.; Crick, S. L.; Li, A.; Diamond, M. I.; Pappu, R. V., Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers.. *J Biol Chem.* **2018,** *293* (10), 3734-3746.

(76) Eftekharzadeh, B.; Piai, A.; Chiesa, G.; Mungianu, D.; García, J.; Pierattelli, R.; Felli, I. C.; Salvatella, X., Sequence context influences the structure and aggregation behavior of a polyQ tract. *Biophys J.* **2016,** *110* (11), 2361-2366.

(77) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J., OGlcNAcylation and phosphorylation have opposing structural effects in Tau: phosphothreonine induces particular conformational order. *J. Am. Chem. Soc.* **2014,** *136* (10), 3803-3816.

(78) Si, K.; Kandel, E. R., The Role of Functional Prion-Like Proteins in the Persistence of Memory. *Cold Spring Harb Perspect Biol.* **2016,** *8(4)* (4), a021774.

(79) Conicella, A. E.; Zerze, G. H.; Mittal, J.; Fawzi, N. L., ALS mutations disrupt phase separation mediated by alpha-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **2016,** *24*, 1-13.

(80) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, V., Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **2012,** *51* (11), 2224-2231.

(81) Warkentin, E.; Weidenweber, S.; Schühle, K.; Demmer, U.; Heider, J.; Ermler, U., A rare polyglycine type II-like helix motif in naturally occurring proteins. *Proteins* **2017,** *85* (11), 2017-2023.

(82) Mok, Y. F.; H., L. F.; Graham, L. A.; Celik, Y.; Braslavsky, I.; Davies, P. L., Structural basis for the superior activity of the large isoform of snow flea antifreeze protein. *Biochemistry* **2010,** *49*, 2593-2603.

(83) Cid, F. P.; Rilling, J. I.; Graether, S. P.; Bravo, L. A.; Mora MdeL.; Jorquera, M. A., Properties and biotechnological applications of ice-binding proteins in bacteria. *FEMS Microbiol Lett.* **2016,** *363* (11), fnw099.

(84) Shoulders, M. D.; Hodges, J. A.; Raines, R. T., Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J. Am. Chem. Soc.* **2006,** *128* (25), 8112-8113.

(85) Pandey, A. K.; Thomas, K. M.; Forbes, C. R.; Zondlo, N. J., Tunable control of polyproline hleix (PPII) structure via aromatic electronic effects: An electronic switch of polyproline helix. *Biochemistry* **2014,** *53*, 5307-5314.

(86) Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J., Proline editing: A general and practical approach to the synthesis of functionally and structurally diverse peptides. Analysis of steric versus steroelectronic effects of 4-substituted prolines. *J. Am. Chem. Soc.* **2013,** *135* (11), 4333-4363.

Graphical Abstract for Table of Contents: