COMMUNICATIONS CHEMISTRY

ARTICLE

https://doi.org/10.1038/s42004-018-0103-2 OPEN

Unveiling the n→π* interactions in dipeptides

I. León [1], E.R. Alonso [1], C. Cabezas[1], S. Mata[1] & J.L. Alonso [1]

Numerous studies have suggested that the n→π* interactions between carbonyls could contribute significantly to the stability of proteins. Nevertheless, their evaluation is challenging because of the solvent environment or crystal packing forces in solids. Here we study the rotational spectrum of HGlyProOH dipeptide, a very common sequence found in collagen, the most abundant protein in vertebrates, in isolated conditions. Three different structures are unequivocally characterized in the gas phase. Interestingly, the most abundant structure is stabilized by an n→π* interaction and adopts the same conformation as is found in crystalline collagen. This observation serves to support the importance of the n→π* interactions between carbonyl groups.

[1] Grupo de Espectrocopía Molecular (GEM), Edificio Quifima, Laboratorios de Espectroscopia y Bioespectroscopia, Unidad Asociada CSIC, Parque Científico UVa, Universidad de Valladolid, 47011 Valladolid, Spain. Correspondence and requests for materials should be addressed to J.L.A. (email: jlalonso@qf.uva.es)

The structural features and folding mechanisms of polypeptides and proteins arise from a complex and subtle balance of different interactions, fundamental not only as a whole but also at the molecular level. Their architecture is sustained by a fragile combination of inter- and intramolecular noncovalent interactions in the polypeptide chains[1]. Many of these interactions are very well known and include hydrogen bonds, Coulombic interactions, van der Waals interactions and hydrophobic effects[2].

Another type of interaction in which two carbonyl groups form attractive interactions with each other has been termed an $n \rightarrow \pi^*$ interaction. This interaction was first discussed in early 1970 but only recently has attracted significant attention, and it is hypothesized to impart substantial stability to proteins, mainly due to abundance of carbonyl groups[3–6]. Concretely, the Raines group has suggested that numerous residues in folded proteins are oriented to take advantage of this energy release, inferring that $n \rightarrow \pi^*$ interactions between carbonyls could contribute significantly to the three-dimensional structure and conformational stability of proteins[7–13].

Recent crystallographic studies are in agreement with the presence of such interactions[14]. Their evaluation is nevertheless challenging because of the solvent environment, or crystal packing forces in solids, which obscures them[15–18]. Moreover, the current challenges in de novo structure prediction and protein design show that our understanding of these interactions is at least incomplete[19–21]. This is reflected in the few experimental studies that directly probe $n \rightarrow \pi^*$ interactions[22–24]. These typically employ either esters[22] or alkenes[22–24] as surrogates for the peptide bond, or rely on synthetizing residue mimics and altering one of their substituents to explore the *trans*/*cis* isomerization ratio[7,25]. As the $n \rightarrow \pi^*$ interaction is only possible in the *trans* isomer, the ratio of isomers ($K_{trans/cis}$) reports on the energy of the interaction.

Despite the important information obtained thus far, none of these methods give an exact evaluation of the interactions involved. Furthermore, all of these methods are approximate and necessarily result in the alteration of these interactions[26,27]. Usually, either they overestimate the strength of the $n \rightarrow \pi^*$ interactions or they are obscured by other interactions. The evaluation of $n \rightarrow \pi^*$ interactions in solution is difficult as there are attenuated effects in polar solvents, suggesting polar interactions rather than orbital interactions[28,29]. What if we could take a simple dipeptide with such competing intramolecular interactions that mimics structures found in crystals and isolate it so that there is no external influence? What if we use a powerful spectroscopic technique that provides accurate structural information? This combination should allow us to provide evidence for $n \rightarrow \pi^*$ interactions in an unperturbed medium and evaluate their intrinsic importance.

On this basis, we pursued the study of the HGlyProOH dipeptide, which is a very common sequence in proteins. For example, collagen, the most abundant protein in vertebrates, consists of three peptide chains forming a triple helix. Each helix is composed of about 1000 amino acid residues with more than 300 repeats of the –Gly-Pro– amino acid sequence[30,31]. Furthermore, in a collagen triple helix, all of the peptide bonds are in the *trans* configuration, suggesting that the strength of the $n \rightarrow \pi^*$ interaction that may stabilize the *trans* configuration correlates with the thermostability of collagen[7]. Therefore, any structural determination of the HGlyProOH sequence is of vital importance to understand the structure of proteins and to gain insight into the nature of the interactions. Under jet-cooled conditions, gas phase spectroscopy studies of peptides offer superior spectral resolution as compared to solvent-broadened condensed phase efforts. Gas phase studies examine the intrinsic conformational propensities of peptide residues without any external perturbation. Combining this ideal environment with Fourier transform microwave spectroscopy[32–35], we obtain accurate rotational constants and quadrupole coupling constants of the nitrogen atoms, allowing us to obtain accurate structural parameters. Only very recently, a solid sample of the simplest HGlyGlyOH dipeptide has been transferred into the gas phase by laser ablation[36], opening up the possibility of measuring a new range of molecules of biological importance using high-resolution rotational studies. In connection with the interest on the –Gly-Pro– sequence in proteins, we tackle the challenging problem of determining the exact nature of the HGlyProOH dipeptide.

Here, we present evidence in gas phase that unambiguously confirms the relevant role that the $n \rightarrow \pi^*$ interaction has between carbonyls, which manifests even in a small system such as the HGlyProOH dipeptide. The results not only show that the *trans* arrangement is significantly stabilized over the *cis* configuration by an $n \rightarrow \pi^*$ interaction, but also that this stabilization is meaningful, being even more stable than the hydrogen bonded form. Furthermore, there is a remarkably good agreement between the crystal structure and the experimentally determined gas phase structure, which further highlights the importance of the conclusions extracted.

Results

The rotational spectrum of HGlyProOH dipeptide. Briefly, neutral molecules of HGlyProOH (m.p. > 183 °C) were transfered to gas phase by laser ablation (LA) of finely powdered samples, using the third harmonic (355 nm) of a picosecond laser. The vaporized products were seeded in neon and supersonically expanded into a vacuum chamber. We can anticipate a rich conformational behavior arising from all plausible interactions between the polar groups, the torsional flexibility of its side chains and the endo or exo configurations that pyrrolic ring could present (Fig. 1a). Hence, up to 10 conformers have been predicted by *ab initio* calculations within 1000 cm^{-1} relative to the global minimum (see Fig. 1b, Supplementary Table 1 and Supplementary Data 1–10). Fortunately, the two-body collision with the carrier gas should cool the stable conformers to very low temperatures, thereby trapping them in their energy minima. Those with sufficient population can be probed in the supersonic expansion by Fourier transform microwave spectroscopy. It should also be highlighted that, while in solution phase the carboxyl group of proline is ionized, the molecules are in their neutral and protonated form in the gas phase.

More than 100,000 free induction decays were averaged in the time domain to obtain the broadband chirped pulse Fourier transform microwave (CP-FTMW) spectrum in the 3.0-8.0 GHz frequency range. We demonstrated[36] how by careful control of the experimental parameters the signal of moderately large species for rotational spectroscopy in gas phase could be obtained. Figure 2a displays a portion of the broadband spectrum, while the whole spectrum is included in the Supplementary Figure 1. The spectrum shows that there is a large number of rotational transitions pointing to the presence of more than one conformer. On a first inspection, after the lines of photofragment species were identified and removed from the spectrum, it was possible to recognize three sets of *a*-type *R*-branch transitions as belonging to three distinct rotamers of HGlyProOH dipeptide labeled as 1, 2 and 3. New predictions and observations allowed the assignment of other *b*- and/or *c*-type *R*-branch lines confirming rotational assignments (see all details in the Supplementary Note 1). The main difficulty of observing the rotational spectrum of HGlyProOH was the nuclear quadrupole coupling interactions produced by two quadrupole nuclei $^{14}N_p$ and $^{14}N_a$ that splits

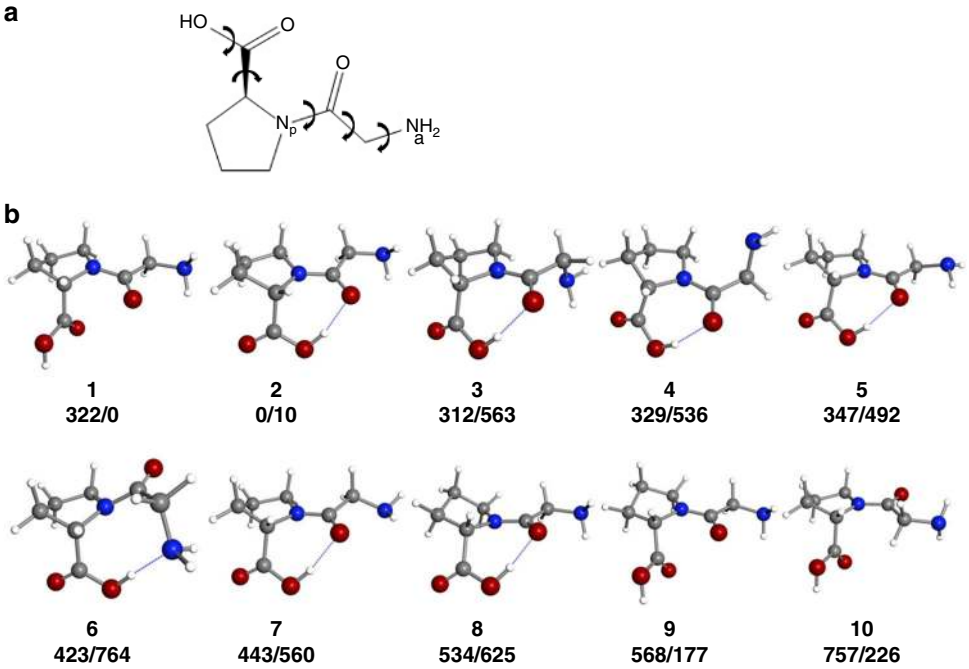


Fig. 1 Chemical structure and most stable conformers of HGlyProOH dipeptide. **a** The torsional flexibility of the side chain and the *endo/exo* configurations of the pyrrolic ring of HGlyProOH dipeptide gives rise to several conformers. The arrows of the sketch indicate the hindered single-bond rotations that govern conformational equilibrium. **b** The calculated low-lying conformers at MP2/6-311++G(d,p). The colors indicate the different atoms: red for oxygen, blue for nitrogen, gray for carbon and white for hydrogen. The values of the energetics and Gibbs free energies at 298 K in wavenumbers are also indicated ($\Delta E/\Delta G$)

each rotational level into several sublevels[37]. Consequently, the overall intensity of each rotational transition spread over many hyperfine components that appeared not well resolved with the attainable resolution of our broadband LA-CP-FTMW technique. This is exemplified for the $4_{1,4}\leftarrow3_{1,3}$ rotational transition in the insets shown in Fig. 2a and Supplementary Figure 2.

At first glance, the frequencies of the rotational transitions were roughly measured at the center of the line clusters and fitted to a rigid rotor Hamiltonian leading to a preliminary set of rotational constants for the three rotamers (see all details in the Supplementary Note 1). These values were compared with those theoretically predicted for the most stable conformers in Table 1. Unfortunately, the difference in the values of the rotational constants is not large enough to allow complete discrimination of the observed rotameric species (see all details in the Supplementary Note 1). However, while the rotational constants are strongly related to mass distribution, the diagonal elements of the nuclear quadrupole coupling tensor, also included in the predictions of Table 1, depend critically on the electronic environment, position and orientation of the ^{14}N nuclei. Hence, the predicted values for the diagonal elements of the nuclear quadrupole coupling tensor (χ_{aa}, χ_{bb} and χ_{cc}) for the ^{14}N nuclei (see Table 1) could provide an independent approach to discriminate the observed species if the nuclear quadrupole hyperfine structure is well resolved and analyzed.

At this point, we took advantage of our narrowband LA-MB-FTMW technique[38], which provides the sufficient resolution to resolve the complicated nuclear hyperfine structure. An example is illustrated in Fig. 2c–e which shows the same $4_{1,4}\leftarrow3_{1,3}$ transitions identified in the broadband spectrum in Fig. 2a now fully resolved (see also Supplementary Figure 3). Thus, in a second stage of the investigation, a selected set of rotational transitions of the three rotamers were analyzed using this high-resolution technique. All measured hyperfine components and detailed explanation of the fitting procedure[39] are given in the Supplementary Tables 2–4.

Conformational characterization and assignment. The main goal of this paper is to obtain meaningful information about the interactions of the HGlyProOH dipeptide, more precisely to validate the $n\rightarrow\pi^*$ interactions in the absence of any external perturbation. If this is possible in these structures, then it serves as a probe of principle that these interactions are fundamental. Therefore, an accurate structural determination is mandatory. Table 2 lists the spectroscopic parameters obtained from the analysis. They can be directly compared with those of Table 1 from in vacuo *ab initio* predictions and clearly discriminate the different rotamers. Notice how these values are unique and can be considered as the fingerprint of each structure. Thus, rotamers 1, 2 and 3 are conclusively identified as conformers 1, 2 and 10, respectively, depicted in Fig. 3 and Supplementary Data 1–3. These assignments are further confirmed by the consistency of the observed selection rules and intensities with the predicted values for electric dipole moment components in Table 1. Finally, taking into account the predicted value of the dipole moment components and averaging the intensity of selected transitions using the LA-MB-FTMW spectrometer, a good qualitative estimation of the relative abundances can be done. With this approach, the relative abundances of the characterized structures are conformer $1\geq2\gg10$. The absence of the rest of the conformers can be easily explained by the conformational cooling as described in the Supplementary Note 2 and Supplementary Figure 4.

Noncovalent interactions. We unequivocally demonstrate the experimental characterization of conformers 1, 2 and 10. In fact, because of the excellent matching between the predicted and accurate experimental values of the rotational constants and quadrupole terms, the structures are very close to the calculated ones. Therefore, the intermolecular interactions in the HGlyProOH dipeptide can be analyzed through the structures adopted

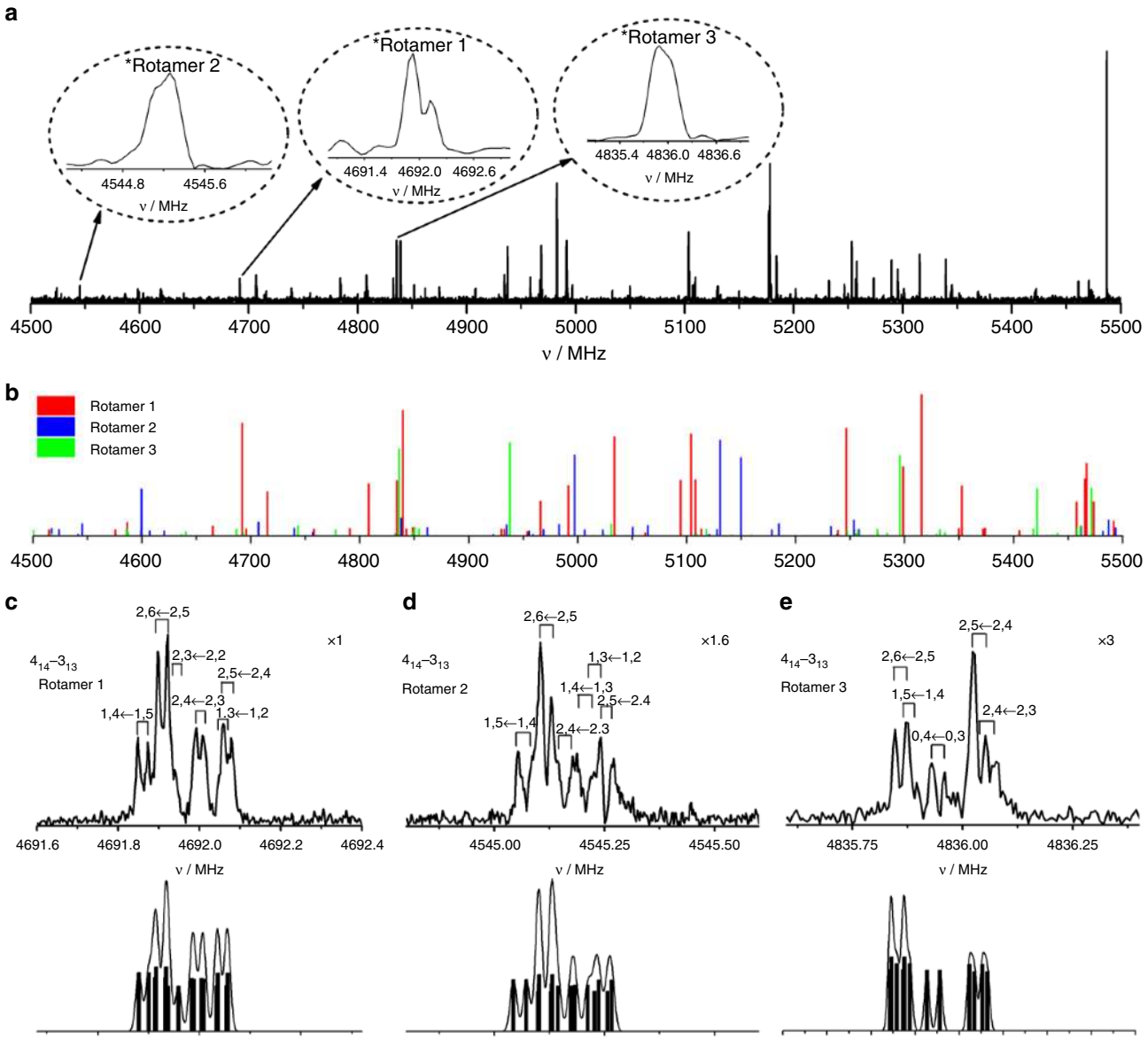
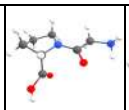
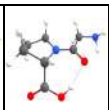
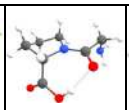
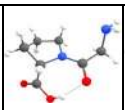
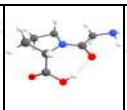
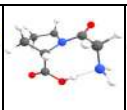
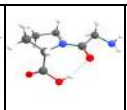
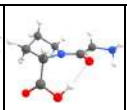
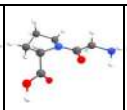
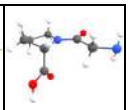


Fig. 2 Rotational spectrum of HGlyProOH. **a** Portion of the broadband spectrum of HGlyProOH in the 4.5–5.5 GHz frequency region (see the Supplementary Figure 1 for the whole spectrum). The insets show the 4_{14}-3_{13} rotational transitions of the three detected conformers of HGlyProOH highlighting that their hyperfine structure cannot be resolved using LA-CP-FTMW spectrometer. **b** Predicted spectrum for each of the detected conformers. Due to multi-resonance excitations[52] the relative intensity of the experimental transitions is affected, and therefore only the position of the peaks must be taken into account. For a correct interpretation of the intensities, the high-resolution spectra using the LA-MB-FTMW spectrometer must be considered. **c**, **d**, **e** The 4_{14}-3_{13} rotational transitions of the three detected conformers of HGlyProOH highlighting their hyperfine structure completely resolved using LA-MB-FTMW spectrometer. Each hyperfine component labeled with the corresponding quantum numbers I', $F' \leftarrow I''$, and F'' is split by the Doppler effect. The predicted components assuming the Doppler effect (as bars) and the convoluted spectra by adjusting each line to a Gaussian function show an excellent agreement. The label in the right top indicates the scaling factor used to scale the intensity

by the detected conformers. Figure 3 shows that all the conformers are stabilized by a bifurcated N-H•••O = C hydrogen bond, i.e., a C5 hydrogen bond, similar to those observed in α-amino acids[40–42]. Conformer 2 possess an additional O-H•••O = C that forces a *trans*-COOH arrangement. This interaction is missed in conformer 1, which is also in a *trans* arrangement. Conformer 10 shows the bifurcated N-H•••O = C interaction, but the C = O group is rotated opposite to the COOH group in a *cis* arrangement. The first important observation comes from this fact. Why is conformer 1 more stable than conformer 10? In principle, they both adopt a similar structure and are stabilized solely by the same bifurcated N-H•••O = C hydrogen bond. Therefore, the explanation must lie somewhere "hidden".

In a quest to solve this puzzle, as well as to gain some insight into the nature of the interactions, we employed calculations at the MP2/6-311++G(d,p) level of theory with natural bond orbital (NBO) analysis.[43] The markedly energetic difference from conformer 1 is attributed to the existence of an $n \rightarrow \pi^*$ interaction between the non-bonding electron pair of the oxygen atom of the carbonyl group and the π^* orbital at the carbonyl group of the carboxylic group. This interaction is shown in Fig. 3 and is a direct indication of why the *trans* disposition is preferred in proteins. Indeed, the second-order perturbation theory analysis shows that there is surprising stabilization energy of 0.77 kcal/mol (269 cm^{-1}) due to this interaction. Furthermore, we recall that our results show that conformer 1 is more abundant than

Table 1 Calculated spectroscopic parameters for the low-lying conformers of HGlyProOH

	 Structure 1	 Structure 2	 Structure 3	 Structure 4	 Structure 5	 Structure 6	 Structure 7	 Structure 8	 Structure 9	 Structure 10
$A^{[a]}$	1416	1482	1334	1456	1494	1251	1471	1350	1274	1270
B	725	709	795	766	713	889	720	736	773	772
C	559	529	606	559	533	601	537	522	548	566
$\lvert\mu_a\rvert$	2.9	1.3	4.1	5.3	2.2	4.4	2.5	1.5	2.8	4.4
$\lvert\mu_b\rvert$	2.9	4.7	4.5	3.7	4.8	2.9	6.0	5.1	3.1	0.2
$\lvert\mu_c\rvert$	0.6	1.4	1.6	0.4	0.3	0.3	2.2	1.0	0.2	1.6
N_p/χ_{aa}	1.90	1.67	1.99	0.98	1.54	2.11	1.75	1.85	2.18	1.90
N_p/χ_{bb}	1.58	1.18	1.26	1.23	1.21	0.01	1.02	0.78	1.29	1.54
N_p/χ_{cc}	-3.48	-2.86	-3.25	-2.21	-2.75	-2.12	-2.77	-2.63	-3.47	-3.44
N_a/χ_{aa}	-1.39	-1.19	-1.63	0.10	2.75	-4.16	2.57	-1.86	-2.23	2.46
N_a/χ_{bb}	0.96	0.11	-0.45	-2.19	0.73	2.44	1.62	0.64	0.97	-2.98
N_a/χ_{cc}	0.44	1.08	2.09	2.09	-3.48	1.72	-4.19	1.21	1.26	0.52
$\Delta E^{[b]}$	322	0	312	329	347	423	443	534	568	757
$\Delta G^{[c]}$	0	10	563	536	492	764	560	625	177	226

[a]*A*, *B* and *C* represent the rotational constants (in MHz); μ_a, μ_b and μ_c are the electric dipole moment components (in D); χ_{aa}, χ_{bb} and χ_{cc} are the diagonal elements of the ^{14}N nuclear quadrupole coupling tensor (in MHz); N_p and N_a correspond to the pyrrolic and amine ^{14}N nuclei, respectively
[b]Relative electronic energies (in cm^{-1}) with respect to the global minimum calculated at the MP2/6-311++G(d,p) level of theory
[c]Relative energies taking into account the total electronic energy plus the thermal correction to Gibbs free energy (in cm^{-1}) calculated at 298 K at the MP2/6-311++G(d,p) level of theory

Table 2 Experimental spectroscopic parameters for the detected conformers of HGlyProOH

	Rotamer 1	Rotamer 2	Rotamer 3
A^a	1418.7739 (28)[b]	1479.8441 (20)	1272.0047 (54)
B	711.20486 (38)	710.21800 (42)	771.32541 (41)
C	551.90421 (18)	529.00341 (19)	562.95309 (25)
$\lvert\mu_a\rvert$	Observed	Observed	Observed
$\lvert\mu_b\rvert$	Observed	Observed	Not observed
$\lvert\mu_c\rvert$	Observed	Observed	Observed
N_p/χ_{aa}	1.779 (10)	1.637 (18)	1.842 (19)
N_p/χ_{bb}	1.530 (19)	1.136 (17)	1.541 (21)
N_p/χ_{cc}	−3.309 (19)	−2.774 (17)	−3.383 (21)
N_a/χ_{aa}	−1.173 (13)	−1.1517 (76)	2.428 (19)
N_a/χ_{bb}	0.736 (14)	0.2766 (96)	−2.851 (20)
N_a/χ_{cc}	0.437 (14)	0.8750 (96)	0.423 (20)
σ^c	3.3	3.2	3.2
N^d	46	52	45

[a]*A*, *B* and *C* represent the rotational constants (in MHz); μ_a, μ_b and μ_c are the electric dipole moment components (in D); χ_{aa}, χ_{bb} and χ_{cc} are the diagonal elements of the ^{14}N nuclear quadrupole coupling tensor (in MHz); N_p and N_a correspond to the pyrrolic and amine ^{14}N nuclei, respectively
[b]Standard error in parentheses in units of the last digit
[c]RMS deviation of the fit (in kHz)
[d]Number of measured transitions

conformer 2, even when the latter has an extra O-H•••O interaction, highlighting once again the importance of this interaction. It is surprising that, in HGlyProOH, the C5-membered conformer 1 showing the $n \rightarrow \pi^*$ interaction is more energetic than the C7-membered hydrogen bonded equivalent conformer 2, analogous to a "ϒ-turn". The orbital overlap between the lone pair (n) of the donor carbonyl oxygen with the π^* antibonding orbital of the second carbonyl group is possible when the putative donor forms a sub-van der Waals' contact with the acceptor ($d < 3.22$ Å) along the Bürgi–Dunitz trajectory for nucleophilic addition ($95° < \theta < 125°$)[44]. The distance found in this work is 2.93 Å, and the angle of approach of the donor oxygen to the acceptor carbonyl is 93.3°, consistent with an $n \rightarrow \pi^*$ interaction. It is also interesting to note that the bifurcated C5 hydrogen bond and the $n \rightarrow \pi^*$ interaction share the same p orbital of the carbonyl's oxygen, which could weaken the strength of the $n \rightarrow \pi^*$ interaction. The same is true for conformer 2 and the C7 hydrogen bond.

Discussion

A striking observation comes when comparing the observed conformers, particularly the most abundant species of HGlyProOH conformer 1, with that found in collagen peptide crystals. Figure 4 shows a comparison between the crystal and molecular structure of a collagen-like peptide at 1.9 Ångstrom resolution[45] and the dominant structure of HGlyProOH obtained in this work. The results speak by themselves: the resemblance between the crystal structure and that got in the gas phase is almost identical. We note that in Fig. 4 no manipulation of the atoms has been done and only a reorientation of the views has been carried out for easier visualization. The reason for such a replica between the crystal and conformer 1 is due to the $n \rightarrow \pi^*$ interaction, confirming its importance even at a molecular level. Usually, these interactions that are present in protein structures, especially helices, are likely to contribute with a 0.27 kcal/mol of stabilization energy per interaction for the amides[25]. The results of this work show that the stabilization energy of this interaction is three times larger in HGlyProOH, highlighting the importance of this interaction. Its importance is such that the structure of the isolated dipeptide is maintained identically in the crystal, where the interactions between different peptide chains and the surrounding solvent are expected. This is very well illustrated in Fig. 4b.

The accurate structural determination of HGlyProOH, which is maintained in the crystal form, together with the observation of the $n \rightarrow \pi^*$ interaction in the most abundant conformer in this simple dipeptide, confirms the importance of this interaction

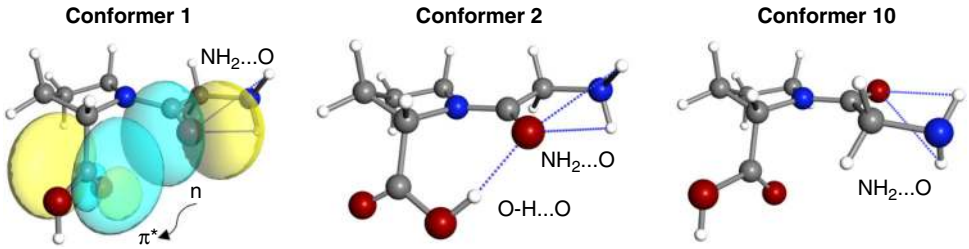


Fig. 3 Determined structures of HGlyProOH. The three determined structures of HGlyProOH dipeptide. The intramolecular bonds are highlighted

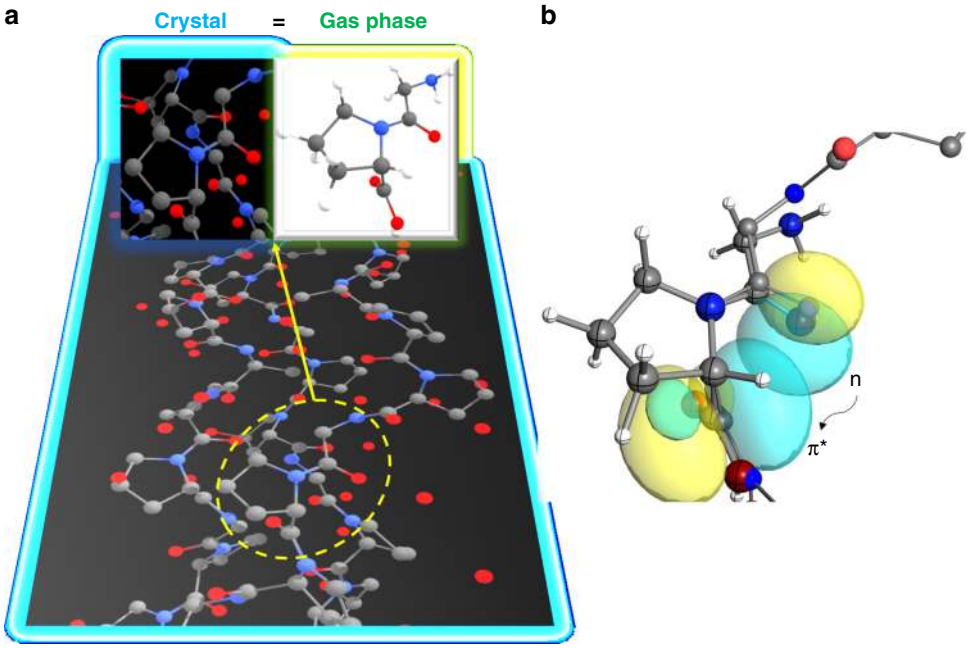


Fig. 4 Comparison between crystal and gas phase. **a** Comparison between the crystal and molecular structure of a collagen-like peptide at 1.9 Ångstrom resolution (pdb:1CAG) (http://www.rcsb.org/pdb/explore/explore.do?structureId=1cag) and the most relevant structure of the HGlyProOH dipeptide in a reoriented view. **b** The comparison of a subtracted piece from the crystal and that of the experimental conformer 1 in this work (with circled spheres and showing the $n \rightarrow \pi^*$ interaction) by overlapping the pyrrolic rings highlights the close resemblance between the two experimental structures. This consolidates the $n \rightarrow \pi^*$ interactions

which could have considerable implications for protein structure. For example, it is known that formation of the triple helix conformation in collagen requires the presence of a repeated –Gly-X-Y– sequence, the most common sequence being –Gly-Pro-Hyp–, because of being the most stabilizing tripeptide unit for the triple helix conformation[46–49]. A consequence of this folding pattern is that only Gly is small enough to fit as every third residue in each polypeptide chain where the three chains pack nearby. The hydrogen bonds in the triple helix occur between –NH group of glycine from a polypeptide α-chain and the carbonyl (C = O) group of proline residues from another chain (N-H•••O = C). The results in this work show that in HGlyProOH the glycine and proline skeletons are maintained in the same plane, like in the crystal structure, and that the COOH group whose OH will be replaced by the next amino acid is stabilized by an $n \rightarrow \pi^*$ interaction. This interaction not only could help to stabilize the peptide/protein, but it could also shape each peptide in a helicoidal arrangement due to the perpendicular and fixed disposition of the carboxylic group. Therefore, the presence of the carbonyl groups in these amino acids could serve not only to template the ideal backbone dihedral angles of the collagen triple helix, but also to rely on the effect attraction between adjacent backbone carbonyl groups through an $n \rightarrow \pi^*$ interaction. Clearly, this structural adaptation allows a close association of the collagen fibers within the molecule, facilitating hydrogen bonding and the formation of intermolecular cross-links. It is interesting how the stabilization energy is, in fact, 0.77 kcal/mol. This interaction is so significant that, in a simple dipeptide, the structure with such an interaction is as stable as that structure with a hydrogen bond. Therefore, due to the isolated conditions of the gas phase, the results presented in this work are probably the most accurate ones to evaluate the importance of the stabilizing interactions, and are in line with the hypothesis and findings regarding the importance of the $n \rightarrow \pi^*$ interactions[4,7,50].

Methods

Experimental. A commercial sample of HGlyProOH was used without any further purification. A solid rod was prepared by pressing the compound's fine powder mixed with a small amount of commercial binder and was placed in the ablation nozzle. A picosecond Nd:YAG laser (355 nm, 20 mJ per pulse, 20 ps pulse width) was used as a vaporization tool. Products of the laser ablation were supersonically expanded using the flow of carrier gas (Ne, 8 bar) and characterized by chirped pulse and molecular beam Fourier transform microwave spectroscopies (LA-CP-FTMW, LA-MB-FTMW), using a recently constructed instrument[38] dedicated to maximizing its performance from 2 to 8 GHz. It is ideal to record the rotational

spectrum of large molecules such as HGlyProOH providing the high resolution necessary to analyze the hyperfine structure due to the presence of several ^{14}N nuclei in the molecule.

Simulations. Geometry optimizations of HGlyProOH were done using Gaussian suite programs[51]. The model of choice was the Møller–Plesset (MP2) perturbation theory in the frozen core approximation, with the Pople's 6-311++G(d,p) basis set. Frequency calculations were also computed to ensure that the optimized geometries are true minima and to calculate the Gibbs free energies. The NBO analysis was also done using the same program and the value for the n→π* interaction is taken from the second-order perturbation theory for the donor–acceptor interaction. See ref. [44] and references therein for more details.

Data availability

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information file. Other relevant data are available from the corresponding author upon reasonable request.

Received: 18 October 2018 Accepted: 30 November 2018
Published online: 04 January 2019

References

1. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
2. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
3. Newberry, R. W., Bartlett, G. J., VanVeller, B., Woolfson, D. N. & Raines, R. T. Signatures of n→π* interactions in proteins. *Protein Sci.* **23**, 284–288 (2014).
4. Newberry, R. W. & Raines, R. T. The n→π* interaction. *ACC Chem. Res.* **50**, 1838–1846 (2017).
5. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205–211 (1951).
6. Pauling, L. & Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. USA* **37**, 729–740 (1951).
7. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. & Raines, R. T. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123**, 777–778 (2001).
8. DeRider, M. L. et al. Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J. Am. Chem. Soc.* **124**, 2497–2505 (2002).
9. Hinderaker, M. P. & Raines, R. T. An electronic effect on protein structure. *Protein Sci.* **12**, 1188–1194 (2003).
10. Hodges, J. A. & Raines, R. T. Energetics of an n→π* interaction that impacts protein structure. *Org. Lett.* **8**, 4695–4697 (2006).
11. Choudhary, A., Gandla, D., Krow, G. R. & Raines, R. T. Nature of amide carbonyl-carbonyl interactions in proteins. *J. Am. Chem. Soc.* **131**, 7244–7246 (2009).
12. Jakobsche, C. E., Choudhary, A., Miller, S. J. & Raines, R. T. n→π* Interaction and n)(π pauli repulsion are antagonistic for protein stability. *J. Am. Chem. Soc.* **132**, 6651–6653 (2010).
13. Bartlett, G. J., Choudhary, A., Raines, R. T. & Woolfson, D. N. n→π* interactions in proteins. *Nat. Chem. Biol.* **6**, 615–620 (2010).
14. Wilhelm, P., Lewandowski, B., Trapp, N. & Wennemers, H. A crystal structure of an oligoproline PPII-Helix, at last. *J. Am. Chem. Soc.* **136**, 15829–15832 (2014).
15. Robertson, E. G. & Simons, J. P. Getting into shape: conformational and supramolecular landscapes in small biomolecules and their hydrated clusters. *Phys. Chem. Chem. Phys.* **3**, 1–18 (2001).
16. De Vries, M. S. & Hobza, P. Gas-phase spectroscopy of biomolecular building blocks DFT: density-functional theory. *Annu. Rev. Phys. Chem.* **58**, 585–612 (2007).
17. Oomens, J., Steill, J. D. & Redlich, B. Gas-phase IR of deprotonated amino acids. *J. Am. Chem. Soc.* **131**, 4310–4319 (2009).
18. Caminati, W. Nucleic acid bases in the gas phase. *Angew. Chem. Int. Ed.* **48**, 9030–9033 (2009).
19. Fleishman, S. J. et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol.* **414**, 289–302 (2011).
20. Kryshtafovych, A., Fidelis, K. & Moult, J. CASP10 results compared to those of previous CASP experiments. *Proteins* **79**, 196–207 (2011).
21. Benjamin Stranges, P. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22**, 74–82 (2013).
22. Jenkins, C. L., Vasbinder, M. M., Miller, S. J. & Raines, R. T. Peptide bond isosteres: ester or (E)-alkene in the backbone of the collagen triple helix. *Org. Lett.* **7**, 2619–2622 (2005).
23. Dai, N., Wang, X. J. & Etzkorn, F. A. The effect of a trans-locked Gly-Pro alkene isostere on collagen triple helix stability. *J. Am. Chem. Soc.* **130**, 5396–5397 (2008).
24. Dai, N. & Etzkorn, F. A. Cis-trans proline isomerization effects on collagen triple-helix stability are limited. *J. Am. Chem. Soc.* **131**, 13728–13732 (2009).
25. Newberry, R. W., Vanveller, B., Guzei, I. A. & Raines, R. T. n→π* interactions of amides and thioamides: Implications for protein stability. *J. Am. Chem. Soc.* **135**, 7843–7846 (2013).
26. Choudhary, A. & Raines, R. T. An evaluation of peptide-bond isosteres. *Chembiochem* **12**, 1801–1807 (2011).
27. Newberry, R. W., VanVeller, B. & Raines, R. T. Thioamides in the collagen triple helix. *Chem. Commun.* **51**, 9624–9627 (2015).
28. Kubyshkin, V. & Budisa, N. Hydrolysis, polarity, and conformational impact of C-terminal partially fluorinated ethyl esters in peptide models. *Beilstein. J. Org. Chem.* **13**, 2442–2457 (2017).
29. Worley, B., Richard, G., Harbison, G. S. & Powers, R. 13C NMR reveals no evidence of n→π* interactions in proteins. *PLoS One* **7**, e42075 (2012).
30. Improta, R., Benzi, C. & Barone, V. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **123**, 12568–12577 (2001).
31. Lam, J. S. W. et al. Predicting the conformational preferences of N-acetyl-4-hydroxy-L-proline- N′-methylamide from the proline residue. *J. Mol. Struct. THEOCHEM* **666–667**, 285–289 (2003).
32. Alonso, J. L. & López, J. C. in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules* (eds Rijs, A. M. & Oomens, J.) 335–402 (Springer Publishing, Cham, 2015).
33. Alonso, J. L., Pérez, C., Sanz, M. E., López, J. C. & Blanco, S. Seven conformers of L-threonine in the gas phase: a LA-MB-FTMW study. *Phys. Chem. Chem. Phys.* **11**, 617–627 (2009).
34. Peña, I., Sanz, M. E., López, J. C. & Alonso, J. L. Preferred conformers of proteinogenic glutamic acid. *J. Am. Chem. Soc.* **134**, 2305–2312 (2012).
35. Cabezas, C. et al. The conformational locking of asparagine. *Chem. Commun.* **48**, 5934–5936 (2012).
36. Cabezas, C., Varela, M. & Alonso, J. L. The structure of the elusive simplest dipeptide Gly-Gly. *Angew. Chem. Int. Ed.* **129**, 6520–6525 (2017).
37. Gordy, W. & Cook, R. L. *Microwave Molecular Spectra* (Wiley, New York, 1984).
38. Bermúdez, C., Mata, S., Cabezas, C. & Alonso, J. L. Tautomerism in neutral histidine. *Angew. Chem. Int. Ed.* **53**, 11015–11018 (2014).
39. Pickett, H. M. The fitting and prediction of vibration-rotation spectra with spin interactions. *J. Mol. Spectrosc.* **148**, 371–377 (1991).
40. Godfrey, P. D., Firth, S., Hatherley, L. D., Brown, R. D. & Pierlot, A. P. Millimeter-wave spectroscopy of biomolecules: alanine. *J. Am. Chem. Soc.* **115**, 9687–9969 (1993).
41. Mcglone, S. J. & Godfrey, P. D. Rotational spectrum of a neurohormone: P-alanine. *J. Am. Chem. Soc.* **117**, 1043–1048 (1995).
42. Blanco, S., Lesarri, A., López, J. C. & Alonso, J. L. The gas-phase structure of alanine. *J. Am. Chem. Soc.* **126**, 11675–11683 (2004).
43. Glendening, E. D., Landis, C. R. & Weinhold, F. Natural bond orbital methods. *WIREs Comput. Mol. Sci.* **2**, 1–42 (2012).
44. Bürgi, H. B., Dunitz, J. D. & Shefter, E. Chemical reaction paths. IV. Aspects of O.C=O interactions in crystals. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **30**, 1517–1527 (1974).
45. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266**, 75–81 (1994).
46. Rich, A. & Crick, F. H. C. The molecular structure of collagen. *J. Mol. Biol.* **3**, 483–506 (1961).
47. Fraser, R. D. B., MacRae, T. P. & Suzuki, E. Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463–481 (1979).
48. Okuyama, K., Xu, X., Iguchi, M. & Noguchi, K. Revision of collagen molecular structure. *Biopolym. Pept. Sci. Sect.* **84**, 181–191 (2006).
49. Wedemeyer, W. J., Welker, E. & Scheraga, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry* **41**, 14637–14644 (2002).
50. 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).

51. Frisch, M. J. et al. *Gaussian. Gaussian 09 Rev. D01* (Gaussian, Inc., Wallingford, 2012).
52. Schmitz, D., Alvin Shubert, V., Betz, T. & Schnell, M. Multi-resonance effects within a single chirp in broadband rotational spectroscopy: the rapid adiabatic passage regime for benzonitrile. *J. Mol. Spectrosc.* **280**, 77–84 (2012).

Acknowledgements

The financial fundings from Ministerio de Ciencia e Innovación (Consolider-Ingenio 2010 CSD2009-00038 program "ASTROMOL", CTQ2013-40717-P and CTQ2016-76393-P), Junta de Castilla y León (Grant VA077U16) and European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC-2013-SyG, Grant Agreement no. 610256 NANOCOSMOS, are gratefully acknowledged. E.R.A. thanks Ministerio de Ciencia e Innovación for FPI grant (BES-2014-067776) and I.L thanks Junta de Castilla y León and Universidad de Valladolid for a postdoctoral contract.

Author contributions

J.L.A. conceived the experiment. J.L.A. and S.M. designed the experimental setup. I.L., E.R.A. and C.C. performed the experiment and analyzed the data. I.L., E.R.A. and C.C. contributed with the calculations. I.L. and J.L.A. wrote the paper. All authors discussed the results and contributed to the final manuscript.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s42004-018-0103-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019