Proline C–H bonds as loci for proline assembly via C–H/O interactions

Noah J. Daniecki, Megh R. Bhatt, Prof. Dr. Glenn P. A. Yap, and Prof. Dr. Neal J. Zondlo*

Department of Chemistry and Biochemistry

University of Delaware

Newark, DE 19716

United States

* To whom correspondence should be addressed. email: *zondlo@udel.edu*, phone: +1-302-831-0197.

Twitter @ZondloLab @UDelaware

http://www.udel.edu/chem/zondlo

Abstract

Proline residues within proteins lack a traditional hydrogen bond donor. However, the hydrogens of the proline ring are all sterically accessible, with polarized C–H bonds at Hα and Hδ that exhibit greater partial positive character and can be utilized as alternative sites for molecular recognition. C–H/O interactions, between proline C–H bonds and oxygen lone pairs, have been previously identified as modes of recognition within protein structures and for higher-order assembly of protein structures. In order to better understand intermolecular recognition of proline residues, a series of proline derivatives was synthesized, including 4*R*-hydroxyproline nitrobenzoate methyl ester, acylated on the proline nitrogen with bromoacetyl and glycolyl groups, and Boc-4*S*-(4-iodophenyl)hydroxyproline methyl amide. All three derivatives exhibited multiple close intermolecular C–H/O interactions in the crystallographic state, with H•••O distances as close as 2.3 Å. These observed distances are well below the 2.72 Å sum of the van der Waals radii of H and O, and suggest that these interactions are particularly favorable. In order to generalize these results, we further analyzed the role of C–H/O interactions in all previously crystallized derivatives of these amino acids, and found that all 26 structures exhibited close intermolecular C–H/O interactions. Finally, we analyzed all proline residues in the Cambridge Structural Database of small-molecule crystal structures. We found that the majority of these structures exhibited intermolecular C–H/O interactions at proline C–H bonds, suggesting that C–H/O interactions are an inherent and important mode for recognition of and higher-order assembly at proline residues. Due to steric accessibility and multiple polarized C–H bonds, proline residues are uniquely positioned as sites for binding and recognition via C–H/O interactions.

Introduction

Proline lacks a hydrogen-bond donor group when incorporated within peptides and proteins. Thus, proline strongly disrupts α-helix and β-sheet structures.[1] Proline is most commonly observed in loop and turn structures within proteins because of this disruption of the major hydrogen-bonded secondary structures of proteins.[2] Proline strongly promotes polyproline II helix (PPII), which does not involve hydrogen bonds to stabilize structure.[3] Alternatively, proline can nucleate the first turn of α-helices, where the amide nitrogens interact with solvent and the conformational restriction of proline is favorable.[4]

However, the C–H bonds of proline are sterically accessible, independent of conformation, due to the pyrrolidine ring structure. In contrast, in the canonical amino acids with side chains, accessibility to backbone C–Hα is partially sterically occluded as a function of side-chain structure and bond rotations, imposing an entropic cost for interaction.[5] In proline, the C–Hα and C–Hδ bonds are both solvent accessible and polarized due to the adjacent electron-withdrawing amide groups. This polarization is observable in their more downfield hydrogen chemical shifts (Hα ~4–5 ppm, Hδ ~ 3 ppm) compared to simple aliphatic groups (1–2 ppm).[6] The bond polarization results in specific interactions of these electron-deficient positions in proteins with electron-rich groups, including with the faces of aromatic rings (C–H/π interactions) and with oxygen lone pairs (C–H/O interactions).[5, 7] C–H/O interactions (Figure 1)[8] are particularly important in the structure and recognition of α-helices in membrane proteins.[5, 9] Here, the proline C–H bonds can function to replace traditional N–H hydrogen-bond donors to stabilize α-helices, without requiring the desolvation energy cost of inserting amide N–H groups into a membrane.[10] These proline-mediated C–H/O interactions either can be present near the N-termini of α-helices (Figure 1b), or can be present as "kinks" in the middle of α-

helices (Figure 1c).[2a, 5, 9a, 11] C–H/O interactions are also important in the recognition and structure of α-helical GXXXG motifs in membrane proteins.[12] In addition, the ζ conformation [(ϕ,ψ) ~ (–130°, +70°)] that is common at residues prior to proline is stabilized by a C_i=O•••Hδ–C_{i+2} C–H/O interaction (Figure 1d).[13]

Consistent with the structural roles of C–H/O interactions at proline residues in proteins, C–H/O interactions are also observed as mediators of proline assembly in small-molecule crystal structures. For example, in the crystal structure of acetylated 4*R*-hydroxyproline methyl ester (Ac-Hyp-OMe),[8c] a C–H/O interaction between the ester carbonyl of one molecule and the hydroxyproline C–Hδ of an adjacent molecule contributes to crystal assembly (Figure 1e). The H•••O interaction distance is 2.43 Å, which is substantially below the 2.72 Å sum of the van der Waals radii of H and O. Notably, a C–H/O interaction at hydroxyproline was identified in the earliest work that described C–H/O interactions in crystal structures.[14] Intermolecular C–H/O interactions are also observed in nucleic acids, in the recognition of biotin by streptavidin, and in protein complexes with small-molecule inhibitors, suggesting their potential applications in medicinal chemistry.[7e, 14-15]

The C–H/O interaction is promoted by favorable electrostatics between the δ^+ on the Pro Hδ and the δ^- on the carbonyl O.[15a] In addition, an O•••H distance that is well below the sum of the van der Waals radii of O and H suggests a role for stereoelectronic (molecular orbital) effects and electron delocalization (partially covalent bonding) in the C–H/O interaction.[7h, 16] Consistent with this interpretation, natural bond orbital (NBO)[17] analysis indicates substantial electron delocalization between the p-like oxygen lone pair (n_p) and the C–Hδ σ* molecular orbitals (an $n \rightarrow \sigma^*$ interaction) in the structure of the Ac-Hyp-OMe dimer (Figure 1f).

We have previously reported a significant number of small-molecule crystal structures of derivatives of 4*R*-hydroxyproline nitrobenzoate (Hnb) and 4*S*-hydroxyproline nitrobenzoate (hnb).[18] In that work, we demonstrated the ability of electron-rich acyl capping groups to promote n→π* interactions between consecutive carbonyls, and thus favor the α-helix and PPII conformations. In contrast, more electron-poor acyl capping groups exhibited weaker or no n→π* interactions and more extended conformations. We also have structurally examined 4*R*- and 4*S*-iodophenyl hydroxyprolines as conformationally biased proline derivatives that can be employed in bioorthogonal reactions (Suzuki and Sonogashira reactions) on peptides in water.[19] Our prior preliminary analysis[18b] of a small number of these structures suggested that C–H/O interactions might be important in their crystal assemblies. Herein, we broadly examine the role of C–H/O interactions as loci of intermolecular interactions with proline residues.

Results

Synthesis of 4-substituted proline derivatives. We synthesized three new molecules that are derivatives (Figure 2) of 4*S*-iodophenyl hydroxyproline or of 4*R*-hydroxyproline nitrobenzoate methyl ester. These amino acids were chosen due to the ability of iodophenyl and nitrobenzoate groups to promote crystal assembly,[18, 19b] via combinations of π stacking and/or halogen bonding interactions.[20] The derivative of Boc-4*S*-iodophenyl hydroxyproline (**2**) specifically included a methyl amide C-terminal group, analogous to that present in proteins. The nitrobenzoate derivatives included glycolyl (**4**) and bromoacetyl (**5**) acyl capping groups, to further explore the role of acyl capping group electronic effects on conformation.

Effects of acyl capping groups on conformation of 4R-hydroxyproline nitrobenzoates. Acyl capping groups can impact local conformation via their ability to modulate electron density

at an electron-donor carbonyl. For example, intercarbonyl n→π* interactions promote the *trans*-proline amide bond.[21] Thus, in acylated proline derivatives, the electronic properties of the acyl group (the electron donor in an n→π* interaction) impact the relative populations of the *trans*-proline and *cis*-proline rotamers. These electronic effects on conformation are independent of any effects on intermolecular C–H/O interactions in crystal assembly, which will be discussed separately later. Capping group electronic effects on conformation may be quantified via the ratio of *trans*-proline to *cis*-proline rotamer present in solution ($K_{trans/cis}$), with a higher value consistent with a more electron-donating acyl cap. Solution NMR data in $CDCl_3$ indicated a $K_{trans/cis}$ of 5.5 for the bromoacetyl derivative **5**, similar to the $K_{trans/cis}$ of the chloroacetyl (6.0) derivative (Figure S4, Table S13).[18c] The $K_{trans/cis}$ of the glycolyl derivative **4** was significantly larger than that of the methoxyacetyl derivative in $CDCl_3$ (5.3 versus 3.2), while the values were similar in CD_3OD (3.9 and 4.3, respectively), suggesting a role for hydrogen bonding in the acyl capping effects of the glycolyl group beyond the inherent electron-withdrawing effect of the oxygen. These electronic effects were also observed in geometry optimization calculations on the simpler derivatives Glyc-Flp-OMe and BrAc-Flp-OMe (Flp = 4*R*-fluoroproline), which exhibited n→π* interaction distances similar[18c] to those of the methoxyacetyl and chloroacetyl derivatives, respectively (Tables S8-S12).

In the chloroacetyl derivative, based on the crystallographic data and DFT calculations we previously hypothesized that the chlorine enhanced the carbonyl-carbonyl n→π* interaction via a halogen lone pair-carbonyl n→π* interaction. Computational data on the bromoacetyl derivative suggest a similar effect here as is observed in the chloroacetyl amide (Figure S6, Table S8). In contrast, the fluoroacetyl derivative more substantially reduced the driving force for a *trans* amide bond ($K_{trans/cis}$ = 2.6), consistent with the greater electronegativity of fluorine

reducing the electron density on the carbonyl and weakening the n→π* interaction. In our prior work, we observed these effects both on the equilibrium constant (ΔG) and on the enthalpy (ΔH) of the proline *cis-trans* isomerization equilibrium.[18c]

Crystal structures of 4-substituted proline derivatives. All compounds readily crystallized, as expected. The structure of **2** exhibited an unusual conformation for a 4*S*-hydroxyproline derivative, adopting an *exo* proline ring pucker (Figure 3). Surprisingly, the *exo* ring pucker was observed in combination with a *cis* Boc-proline carbamate conformation, despite the usual correlation of a *cis*-proline with an *endo* ring pucker.[22] The proline residue was in a PPII conformation, using standard conventions to describe conformations of amino acids within regions of the Ramachandran plot.[13b] Prior NMR data on 4*S*-iodophenyl hydroxyproline in an Ac-TY*Prox*N-NH_2 peptide context, *Prox* = 4-substituted proline,[19a] as well as computational data on **2** as a function of ring pucker and conformation (Table S5), suggest that the iodophenyloxy (or more generally aryloxy) substitution on proline may somewhat reduce the inherent stereoelectronic preference of an aryloxy group for a pseudo-axial position (i.e. an *endo* ring pucker with 4*S*-substitution) due to a modest steric cost of the aryloxy group when pseudo-axial, which relatively increases the likelihood of the group being pseudo-equatorial (as observed here, with an *exo* ring pucker).

Crystal assembly was mediated by a series of noncovalent interactions, including a dual hydrogen bond plus C–H/O interaction in which one carbamate carbonyl was simultaneously in close contact with both the amide hydrogen (H•••O distance 1.91 Å) and the Pro C–Hα (Hα•••O distance 2.32 Å) of an adjacent molecule (Figure 3c). This close Hα•••O distance is consistent with a particularly favorable C–H/O interaction. A similar geometry is observed in the assembly

of the collagen triple helix, with the Pro carbonyl of one strand interacting simultaneously with the C–Hα of Hyp and the amide NH of Gly in a Hyp-Gly sequence of a different strand.[23]

An intermolecular C–H/O interaction was also observed between the Pro carbonyl and a Pro C–Hδ (Hδ•••O distance 2.56 Å) (Figure 3c). In all cases, these distances were determined by optimization of the positions of the hydrogens from the crystal structures using DFT calculations, while keeping the crystallographically determined heavy atom positions fixed, in order to achieve the highest accuracy in hydrogen atom positions. Notably, both C–H/O interaction distances were substantially below the 2.72 Å sum of the van der Waals radii of H and O. Crystal assembly was also mediated by a halogen/π interaction, in which the aromatic π electrons interact favorably with the electron-deficient δ^+ σ hole and σ^*_{C-I} of the C–I bond.[20b]

The role of C–H/O interactions in crystal assembly was also explored via full geometry optimization using the dimeric assembly with a dual hydrogen bond and C–H/O interaction as an initial geometry (Figure 3d). The final geometry-optimized structure included two close C–H/O interactions, at C–Hγ (the site of the electron-withdrawing iodophenol; Hγ•••O 2.27 Å) and at C–Hδ (Hδ•••O 2.39 Å), as well as the carbonyl-amide hydrogen bond. These computational results suggest that the close C–H/O interactions observed crystallographically are not an artefact of crystal packing, but are instead an inherent feature of the intermolecular assembly of proline residues.

The crystal structure of **4** exhibited the proline in a PPII conformation (ϕ,ψ = –51°, +150°) with a close n→π* interaction (O_i•••C_{i+1} distance 2.84 Å), as well as the expected *exo* proline ring pucker and a *trans* amide bond (Figure 4). Interestingly, the glycolyl group exhibited an intramolecular hydrogen bond between the glycolyl alcohol and its carbonyl, which would be expected to modestly weaken the n→π* interaction.[24] The glycolyl n→π* interaction was

similar to that observed previously in the electronically similar chloroacetyl derivative, and more distant than that seen in the more electron-rich pivaloyl (2.68 Å) or *iso*-butyryl (2.69 Å, 2.74 Å) derivatives (Table S8), consistent with a role for electronic effects of the electron-donor carbonyl to impact n→π* interactions.[18c]

Crystal assembly via intermolecular π stacking and intermolecular hydrogen bonding was observed. Crystal assembly of **4** was also mediated in part by three C–H/O interactions (Figure 4bc): between the glycolyl alcohol O and Pro C–Hβ (O•••H 2.41 Å); between one nitro O and the other Pro C–Hβ (O•••H 2.46 Å); and between the glycolyl carbonyl and one Pro C–Hδ (O•••H 2.49 Å). Again, all three observed C–H/O interactions were substantially below the 2.72 Å sum of the van der Waals radii of H and O, consistent with favorable interactions that have a substantial stereoelectronic component. Interestingly, for both interactions at Pro Hβ, which is less partially positive than either Hα or Hδ, the interacting carbon-hydrogen bonds are *anti*-periplanar to an electron-withdrawing group (the pro-*R* hydrogen *anti* to the backbone C–N, the pro-*S* hydrogen *anti* to C*γ*–O). These C–H/O interactions would both be expected to further stabilize the *exo* ring pucker observed in the crystallographic form, due to their stabilization of these $\sigma_{C-H} \rightarrow \sigma^*_{C-EWG}$ stereoelectronic effects.

The crystal structure of **5** included two molecules in the unit cell (Figure 5). Both molecules were in the α-helix conformation, stabilized by n→π* interactions (O•••C distances 2.85 Å, 2.89 Å). These molecules provide further evidence that the α-helix conformation can be stabilized by *i*/*i*+1 n→π* interactions, even in the absence of the *i*/*i*+3 or *i*/*i*+4 hydrogen bonds that stabilize 3_{10}-helix and α-helix secondary structures, respectively.[18c, 21, 25] Both molecules also exhibited the expected *exo* proline ring pucker of a 4*R*-hydroxyproline nitrobenzoate.

Crystal assembly was mediated by π stacking of the nitrobenzoate groups,[20a] as well as by halogen bonds between carbonyl oxygens and the C–Br bonds (O•••Br distances 3.19 Å and 3.25 Å, which are below the 3.37 Å sum of the van der Waals radii of O and Br) (Figure S1).[20b] In addition, intermolecular noncovalent interactions were observed that were centered on the proline C–H bonds (Figure 5c). At the C–Hβ and C–Hδ bonds, a series of close C–H/O interactions was observed, mediated by the nitro oxygens or the nitrobenozate carbonyl (O•••Hβ distances 2.31, 2.57, 2.62, 2.66 Å; O•••Hδ distances 2.45, 2.47, 2.47 Å).

In addition, C–H/Br interactions at C–Hα were observed with a geometry similar to those of C–H/O interactions, in which Br lone pairs were aligned for orbital overlap with the C–Hα σ*. The Br•••H distances (2.78 Å, 2.82 Å) were significantly below the 3.05 Å sum of the van der Waals radii of Br and H, consistent with a role for stereoelectronic effects/electron delocalization in this interaction. Notably, at both Br atoms in the unit cell, both modes of halogen bonding were observed, with the δ^+ σ hole/σ^*_{C-Br} functioning as a strong electron acceptor (C–Br•••O angles 174°, 178°) and the δ^- periphery of the Br functioning as an electron donor, here for C–H/Br interactions (Figure S1).

Global analysis of crystal assembly via C–H/O interactions in hydroxyproline nitrobenzoates and iodophenylhydroxyproline crystal structures. In all three crystal structures, we observed that proline assembly involved multiple C–H/O interactions that were substantially below the 2.72 Å sum of the van der Waals radii of H and O. These results suggested that C–H/O interactions might be a general mechanism for assembly at proline residues. We therefore investigated all reported crystal structures from our group that included iodophenyl hydroxyprolines or hydroxyproline nitrobenzoates. Including the structures reported herein, a total of 26 structures was examined.

Remarkably, every single structure included at least one C–H/O interaction at proline with an H•••O distance below 2.72 Å (Figure 6, Tables 1–5). Close interactions were observed at all proline ring positions. The closest observed proline C–H/O interaction (in FAc-Hnb-OMe, at C–Hδ) had an H•••O distance of 2.27 Å, with seven other close proline C–H/O interactions in this crystal structure (2.38–2.52 Å). The largest number of interactions and the closest average interactions were observed at Hδ. While all of these molecules have an additional electron-withdrawing group at Cγ compared to proline, the ubiquity of C–H/O interactions across all of these crystal structures suggests that C–H/O interactions are an *inherent*, general basis for interactions at proline rings.

Informatics analysis of assembly via C–H/O interactions at proline in the Cambridge Structural Database. In order to examine the generality of proline assembly via C–H/O interactions, we conducted a complete search of proline residues in the Cambridge Structural Database (CSD).[26] Each carbon position on the proline ring was then examined individually. Interaction distances here were determined with normalization of bond lengths to hydrogen.[27]

The results indicated that C–H/O interactions are widely observed in crystal structures with proline and proline derivatives (Figure 7, Figure 8, Table 6).[8c, 28] H•••O distances significantly below the 2.72 Å sum of the van der Waals radii of H and O were typical, at each of C–Hα, C–Hβ, C–Hγ, and C–Hδ, with a majority of structures exhibiting at least one C–H/O interaction. The closest observed H•••O distances approached the 2.0 Å distance of typical traditional hydrogen bonds (Figure 7), although in general the distances were longer than in standard hydrogen bonds, consistent with C–H/O interactions being weaker than classical hydrogen bonds. Notably, close C–H/O interactions were more frequent at C–Hα and C–Hδ, which have the most δ^+ character that should make these positions most favorable for C–H/O

interactions. Interactions at Hα were statistically significantly overrepresented, as there is only one Hα at each proline. In contrast, at C–Hγ, the greatest frequency of interactions was around the 2.72 Å sum of the van der Waals radii of H and O, although many interactions were observed that were substantially closer. Overall, the analysis of the CSD indicated that close C–H/O interactions are widely observed at intermolecular assembly interfaces of proline residues. These results further suggest that C–H/O interactions are an inherent feature of proline recognition.

Discussion

Proline is a unique amino acid because of its cyclic structure on the backbone, which causes it to lack a traditional hydrogen-bond donor within proteins. However, due to their proximity to backbone amides, the proline C–Hα and C–Hδ bonds are substantially more polarized than simple aliphatic C–H bonds (e.g. on the side chains of Ala, Val, Leu, or Ile). Moreover, because of the ring structure, these proline C–H bonds are collectively more sterically accessible, with less entropic cost of interaction, than the polarized C–Hα of other non-Gly amino acids.[5]

These proline C–H bonds have previously been identified as potential sites for specific noncovalent interactions, such as C–H/O interactions and C–H/π interactions, in addition to their ability to interact with other nonpolar side chains via the hydrophobic effect. Herein, we examined the generality of C–H/O interactions as bases for intermolecular assembly at proline residues. We analyzed intermolecular C–H/O interactions in a series of related crystal structures from molecules with 4*R*- or 4*S*-hydroxyproline derivatives. We found that every single crystal structure had at least one intermolecular C–H/O interaction with a H•••O distance below the 2.72 Å sum of the van der Waals radii of H and O. Notably, the median interaction distances were

~2.4-2.5 Å, and the closest distances were around 2.3 Å (Table 5). A more comprehensive analysis of proline residues in the CSD indicated that these trends were general (Figure 7, Figure 8, Table 6): a C–H/O interaction at a proline C–H bond was the typical, expected observation, with close interactions commonly observed. These results suggest that C–H/O interactions are an important, inherent mode for recognition of proline.

Moreover, close C–H/O interactions were observed not only at C–Hα and C–Hδ, the most polarized C–H bonds, with the most δ^+ charge on hydrogen, but also at C–Hβ and C–Hγ. The frequency of close interactions suggested a significant role beyond simple electrostatics, or simple van der Waals interactions, as a basis for C–H/O interactions at proline. Indeed, analysis of the partial charges on proline hydrogens in Ac-Pro-NMe_2, using either the Hirshfeld or CM5 methods to assign partial charges to atoms, revealed only very small partial charges on the hydrogens (Hirshfeld δ^+ +0.03 to +0.05; CM5 δ^+ +0.09 to +0.14).[29] If the C–H/O interaction were based entirely on electrostatics, these interactions would be expected to be exceptionally weak outside of the gas phase or the most nonpolar solvents. For example, a gas-phase favorable interaction energy of 2 to 4 kcal mol^{-1}, which has been estimated[30] as a typical C–H/O interaction energy at Hα in proteins, would, if entirely electrostatic in nature, correspond to an interaction energy of only 0.4–0.8 kcal mol^{-1} in chloroform, and less than 0.05 kcal mol^{-1} in water, since electrostatics energy (E) scales with dielectric constant (ε) as $E \sim 1/\varepsilon$. Moreover, a purely electrostatic description of C–H/O interactions would not explain the H•••O distances being substantially below the van der Waals radii of the atoms. Indeed, prior computational analysis of model compounds for C–H/O interactions in implicit solvent indicated only relatively modest reductions in interaction strength in water versus vacuum or non-polar solvents.[31]

In order to examine the relative importance of electrostatics in C–H/O interactions, we examined two model structures (Figure 9). First, we conducted a complete geometry optimization on a minimal version of the dimer of **2** (Figure 3d), with truncation of the *t*-butoxycarbonyl (Boc-) to methoxycarbonyl (Moc-) and of the hydroxyproline iodophenyl ether (–O-PhI) to hydroxyproline (–OH). This minimal structure exhibited an intermolecular hydrogen bond and four intermolecular C–H/O interactions stabilizing the dimer (Figure 9a), with all distances significantly below 2.72 Å. Second, we examined the crystal structure[32] of DMF•$CHCl_3$ as a model small-molecule C–H/O interaction, with the carbonyl O of DMF as the electron donor and the C–H of $CHCl_3$ as the electron acceptor. From the crystal coordinates, which indicated a particularly close 2.0 Å H•••O distance, a restrained geometry optimization calculation was conducted to properly position the hydrogens and to minimize bond distortions inherently present in the crystal structure (Figure 9b). In addition, the individual component compounds (Moc-hyp-NHMe, DMF, $CHCl_3$) were subjected to full, unrestrained geometry optimizations. Both complexes (Moc-hyp-NHMe dimer and DMF•$CHCl_3$) were analyzed for the energies of the complex (compared to the individually optimized components) as a function of solvent.[33] Complex energies were examined in solvation conditions from vacuum ($\varepsilon = 1$) through H_2O ($\varepsilon = 78$), plus computationally generated solvents with $\varepsilon = 1000$ and $\varepsilon = 1{,}000{,}000$, which would essentially fully neutralize the effect of any purely electrostatics-based interactions. If C–H/O interactions were predominantly stabilizing due to electrostatics, then only a minimal interaction strength should be present in H_2O, and fundamentally no interaction strength should be present with $\varepsilon = 1{,}000{,}000$. In contrast to these expectations, the computational data on both complexes (Figure 9; Tables S6 and S7) indicate that the majority of the complex interaction

energy is present even in H_2O or with ε = 1,000,000, with only a modest change in interaction energy between $CHCl_3$ (ε = 4.9) or H_2O as a solvent.

These results are consistent with a significant role for stereoelectronic effects in C–H/O interactions, which has been suggested previously.[7h, 16a-e] The C–H/O interactions have a partially covalent character due to intermolecular electron delocalization between an oxygen and the proline C–H bonds. This stereoelectronic component to noncovalent interactions can be conceptualized considering the frontier molecular orbitals that would overlap to form new, more favorable molecular orbitals upon close assembly. Considering a carbonyl O interaction with a C–H bond, an electron donor on the carbonyl (any of the O p-like lone pair (O_p), the O s-like lone pair (O_s), and/or the carbonyl π orbital) exhibits orbital overlap with the C–H σ^* orbital (Figure 9b), resulting in intermolecular electron delocalization, in a manner analogous to Lewis base-Lewis acid interactions. Electron-withdrawing groups on the C–H (e.g. the backbone amide atoms bound to Cα and Cδ) further enhance the C–H/O interaction, not only via direct effects on electrostatics (more δ^+ on the associated H), but also via lowering the orbital energy of the C–H σ^*, which results in greater stabilization due to electron delocalization. Similar effects of electron-donor groups on the orbital energies of dienes, and of electron-withdrawing groups on the orbital energies of dienophiles, explain the exceptional sensitivity of the *kinetics* of Diels-Alder reactions to the electronics of substituents on either the diene or the dienophile, despite no significant change to the overall *thermodynamics* of the reaction as a result of substitution.[34]

4-Substituted prolines, including 4*R*-hydroxyproline in collagen, are particularly favorable sites for assembly via C–H/O interactions, due to the enhancement of stereoelectronic effects via interaction at the hydrogens. For example, in the crystal structure of Glyc-Hnb-OMe (Figure 4), both diastereotopic Hβ are *anti*-periplanar to an electron-withdrawing group: the *pro-*

S Hβ is *anti* to the Oγ nitrobenzoate (an interaction that stabilizes the Cγ-*exo* ring pucker of the 4*R*-nitrobenzoate via a $\sigma_{C-H} \rightarrow \sigma^*_{C-O}$ stereoelectronic effect), while the *pro-R* Hβ is anti-periplanar to the backbone amide N (stabilizing via a $\sigma_{C-H} \rightarrow \sigma^*_{C-N}$ stereoelectronic effect).[18a, 19a, 35] Notably, the latter effect is inherently present at one C–Hβ and one C–Hγ in *all* proline residues. A C–H/O interaction at these hydrogens should stabilize those stereoelectronic effects, by making the C–H bonds more electron-rich and better σ electron donors. Thus, 5 of the 7 C–H bonds of proline are primed for C–H/O interactions due to bond polarization and/or their roles in stereoelectronic effects that stabilize a proline ring pucker. In addition, in 4-substituted prolines the C–Hγ bond is significantly polarized due to the electronegative atoms on Cγ. Thus, in 4-substituted prolines, *all* C–H bonds exhibit polarization that makes them particularly prone to intermolecular assembly via C–H/O interactions.

Indeed, via the native post-translational modification 4*R*-hydroxyproline, collagen assembly includes interstrand C–H/O interactions between a proline carbonyl and a hydroxyproline C–Hα.[23] More broadly, polyproline helix bundles and other PPII-rich tertiary structures, which fold despite lacking the extensive hydrogen bonds and/or hydrophobic core of typical globular proteins, also assemble via C–H/O interactions.[36] In addition, C–H/O interactions at Hδ stabilize the ζ conformation that is commonly observed at the residue prior to proline (pre-proline conformation) (Figure 1d).[13b, 13c] Interstrand C–H/O interactions at proline can also stabilize β-sheets.[7a, 37] C–H/O interactions may also be involved in the folding of peptoids, oligomers of N-alkyl glycines that can functionally mimic proline, either via the glycine Hα or via the hydrogens on carbons attached to nitrogen, that is, equivalent to Pro Hδ.[38]

C–H/O interactions are common in many structural elements involving proline, including conformational preferences, secondary structure stabilization, and tertiary structure. In addition,

C–H/O interactions are observed as loci for recognition at protein-protein and protein-ligand complexes. Herein, we demonstrated the generality of C–H/O interactions at proline residues in the assembly of intermolecular complexes of proline, via a combination of X-ray crystallography, computational analysis, and informatics analysis of structures in the CSD.

Methods

Small molecule synthesis. The syntheses of **1** and **3** were described previously.[18a, 18c, 19b] The details of the syntheses and characterization of compounds **2**, **4**, and **5** are described in the Supporting Information.

X-ray crystallography. Details of X-ray crystallography are described in the Supporting Information. The structures of **2**, **4**, and **5** have been deposited in the CSD as CCDC 2160105, 2160106, and 2160107.

Computational chemistry. Calculations were conducted with Gaussian 09 or Gaussian 16.[39] The lengths of C–H bonds are in general systematically biased to shorter bond lengths in small-molecule crystal structure determination, due to the lower electron density at hydrogen compared to heavy atoms.[27] For the analysis of interactions at hydrogen, the direct use of hydrogen atom position from CIF files can thus lead to misrepresentation of the distances and geometries of interactions involving these hydrogens. This systematic effect can be addressed in part via normalization of C–H bond lengths,[27] which was conducted using Conquest when applied to the informatics analysis of the CSD. However, this effect inherently ignores the context of the hydrogens and how it can subtly modulate C–H bond lengths and hydrogen positions. Therefore, for the analysis of all crystal structures generated within the group, we

determined the most likely hydrogen atom position using DFT calculations on the crystal structures, in which the positions of the heavy atoms determined crystallographically were fixed, but the positions of the hydrogens were optimized. Hydrogen atom position optimization was conducted iteratively, with final hydrogen position geometry optimization conducted using the M06-2X functional and the Def2TZVP basis set in implicit chloroform (IEFPCM solvation model).[40] For each structure, close interactions were initially identified with bond lengths to hydrogen normalized within Mercury, and the dimeric structures extracted, followed by geometry optimization of hydrogen atom positions. The coordinates of the hydrogen-positions-optimized structures of intermolecular complexes are included in the Supporting Information.

Full geometry optimization was also conducted on **2**, **4**, and **5**, in order to understand the inherent conformational preferences of these proline derivatives. For **2**, full geometry optimization was conducted using each combination of *exo* and *endo* ring pucker, *trans* or *cis* proline, and α_R/δ and PPII/β regions of the Ramachandran plot. For **4** and **5**, geometry optimization was conducted using the crystallographically determined structure as an initial geometry (PPII for **4**, α_R for **5**; *exo* ring pucker and *trans* amide bond for both), with additional analysis of the other minimum-energy conformation in Ramachandran space, all with a *trans* amide and *exo* ring pucker. These structures were then optimized after truncation of the hydroxyproline nitrobenzoate to a fluoroproline or to a proline, and completing geometry optimization of both the α_R and PPII conformations of these structures using both the M06-2X functional and MP2 method[41] with the 6-311++G(2d,2p) basis set[42] in implicit water, in order to compare the effects of the acyl capping groups on structure and n→π* interactions with capping groups that have been characterized previously using the same methods.[18c] Additional details, additional analysis, and atomic coordinates are in the Supporting Information.

In order to understand the nature of the C–H/O interactions in proline assembly, the structure of the dimer of **2** with assembly via a hydrogen bond and C–H/O interactions was subjected to full geometry optimization, in order to obtain a structure that represented optimized intermolecular interactions without the restraints inherently imposed by the crystal lattice. The final geometry optimization was conducted using the M06-2X functional and the Def2TZVP basis set in implicit chloroform. The final coordinates and additional details are in the Supporting Information.

Informatics analysis of the Cambridge Structural Database (CSD). Database searches were conducted using Conquest. Searches were conducted using a generic proline molecule as a starting point, with any substituent at the 4th position, on the proline nitrogen, and on the C-terminal carbonyl carbon. Polycyclic proline residues were excluded by specifying the N-terminal atom and the C-terminal carbonyl carbon as acyclic atoms. An initial search for proline crystal structures with these parameters resulted in 1545 unique structures with proline. The search was then modified to identify close proline C–H•••O contacts between proline hydrogens (examining each proline carbon independently) and any oxygen atom for a range from 0 Å to 2.8 Å, with the hydrogen positions normalized. This search with a defined H•••O distance of ≤ 2.8 Å yielded 834 unique structures. Contacts were identified intermolecularly, or intramolecularly across 5 or more bonds. 2,848 interactions were identified from 834 structures with this search. Each interaction model was sorted by identifier and C–H•••O distance to determine the total number of published crystal structures with H•••O distances ≤ 2.72 Å. This analysis revealed 2,238 C–H/O interactions ≤ 2.72 Å among 789 unique crystal structures containing proline.

Acknowledgements

We thank NSF (CHE-2004110 and BIO-1616490) for funding. Instrumentation support was provided by NIH (GM110758, S10-OD026896A) and NSF (CHE-1229234).

Supporting Information Available

This material is available free of charge via the journal web site.

Keywords

Hydrogen bonds, noncovalent interactions, protein structure, self-assembly, stereoelectronic effects

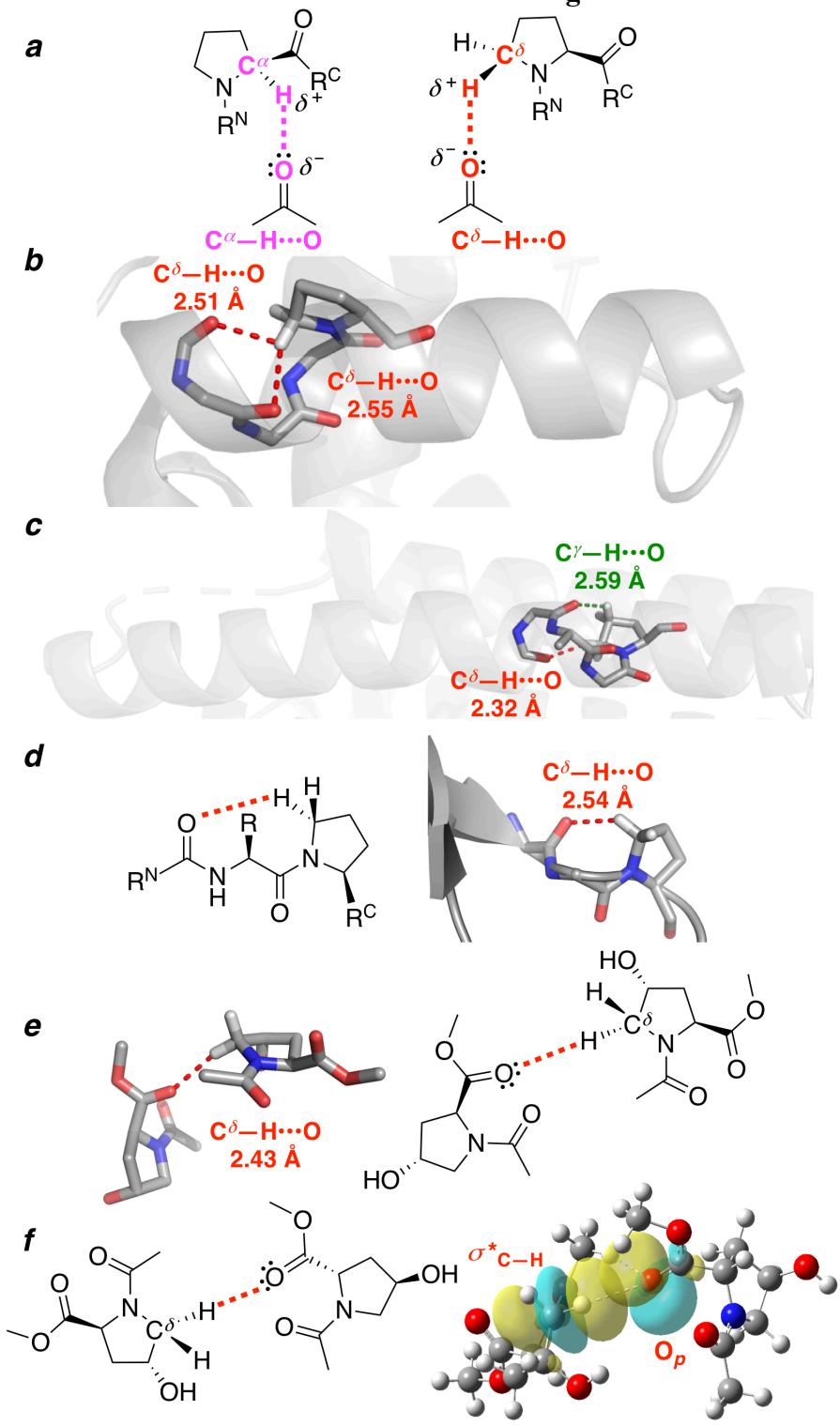


Figure 1. **C–H/O interactions defined and examples from the PDB and CSD**. (a) C–H/O interactions between a generic carbonyl and proline C–H bonds. Interactions are shown at both C^{α} and C^{δ}. (b) X-ray crystal structure of human beta-1 alcohol dehydrogenase (ADH1B) solved

at 1.60 Å resolution (pdb 1u3u).[8b] Residues 322-336 are shown as a grey α-helix. Pro328 has interactions between its H^{δ} and the carbonyls of Glu326 and Lys327 (C^{δ}–H•••O distances 2.51 Å and 2.55 Å respectively, red). (c) X-ray crystal structure of bacteriorhodopsin solved at 1.43 Å resolution (pdb 1m0k).[8b] Residues 176-190 are shown as a grey α-helix. Pro186 has a C–H/O interaction between H^{δ} and the carbonyl of Trp182 (C^{δ}–H•••O distance 2.32 Å, red) and an additional interaction between H^{γ} and the carbonyl of Ser183 (C^{γ}–H•••O distance 2.59 Å, green). (d) X-ray crystal structure of *Rhizomucor miehei* triacylglyceride lipase solved at 1.90 Å resolution (pdb 3tgl).[8c] Residues 98-100 are shown. Pro100 exhibits a C–H/O interaction between H^{δ} and the carbonyl of Ser98 (C^{δ}–H•••O distance 2.54 Å, red). The pre-proline residue (Tyr99) adopts the ξ conformation (ϕ,ψ = –91°, +119°). (e) Crystallographic data from Ac-(2*S*,4*R*)-4-hydroxyproline methyl ester (Ac-Hyp-OMe) showing an intermolecular C–H/O interaction. The crystal structure was used as an initial model and the hydrogen positions were optimized using the M06-2X DFT functional with the Def2TZVP basis set, while the positions of the heavy atoms were fixed to those observed crystallographically. All geometry optimization calculations were conducted with implicit $CHCl_3$ solvation. This structure exhibits a C^{δ}–H/O interaction between the *pro-R* H^{δ} and the Pro carbonyl of the adjacent molecule (C^{δ}–H•••O distance 2.43 Å, red). The crystal structure identifier is RISDAY.[8d] (f) Natural bond orbital (NBO) analysis of Ac-(2*S*,4*R*)-4-hydroxyproline methyl ester dimer showing orbital overlap between the p-like lone pair of the proline carbonyl and the C–H σ^* antibonding orbital. Orbitals are shown with an isovalue of 0.02.

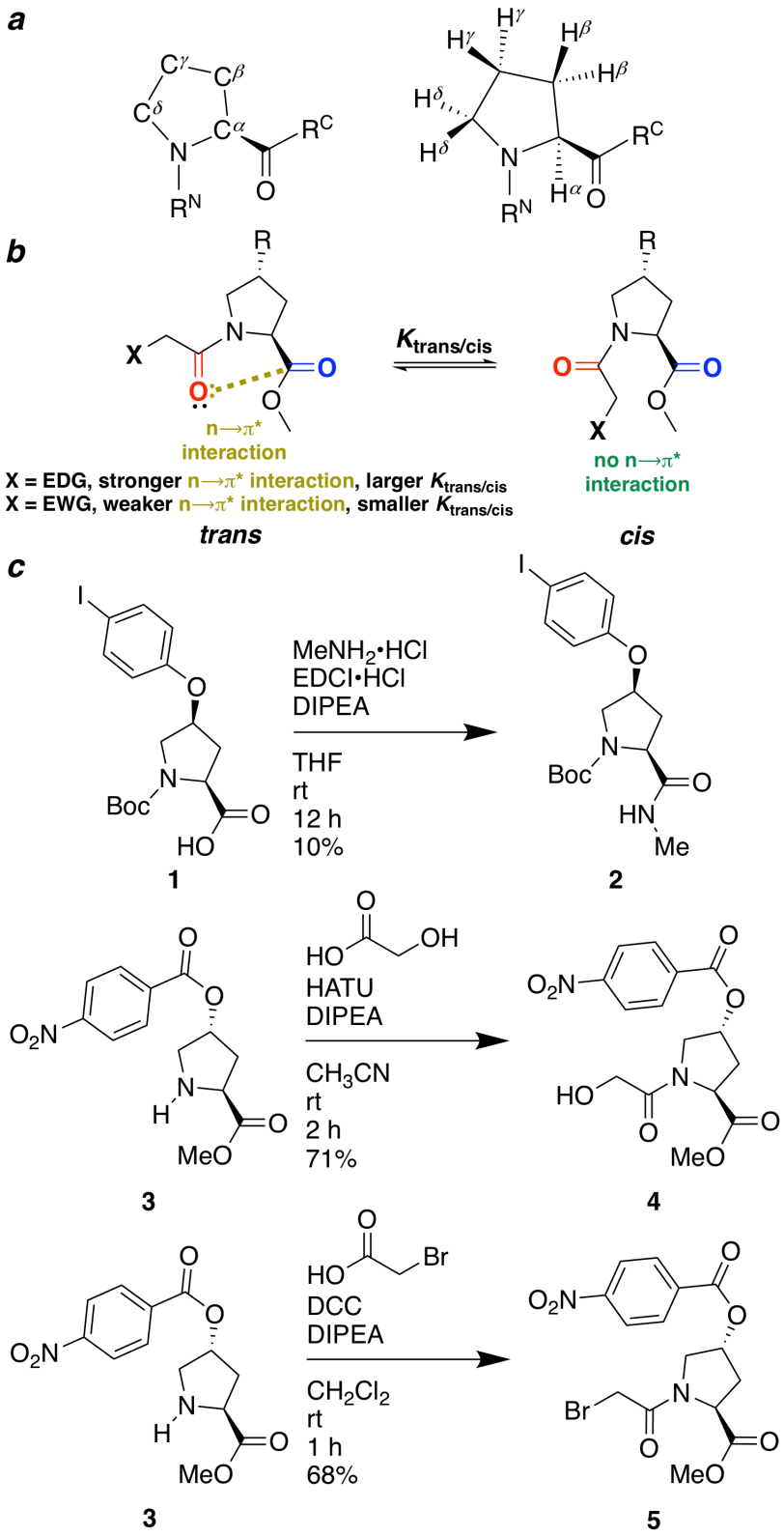


Figure 2. **Synthesis of 4-substituted proline derivatives**. (a) Proline ring positions at the carbons (left) and hydrogens (right). (b) *trans* and *cis* amide rotamers of proline. The n⟶π*

interaction (gold) stabilizes the *trans* conformer, but is not present in the *cis* conformer. When X = an electron-donating group (EDG), the n⟶π* interaction is stronger and the *trans* conformation is promoted (larger $K_{trans/cis}$, closer C•••O distance), while when X = an electron-withdrawing group (EWG), the n⟶π* interaction is weaker, resulting in a longer C•••O distance and a smaller $K_{trans/cis}$. Electron-donor carbonyl, red; electron-acceptor carbonyl, blue. (c) Synthesis of **2**, **4**, and **5** from compounds **1** and **3**, whose syntheses were described previously.[18a, 19b]

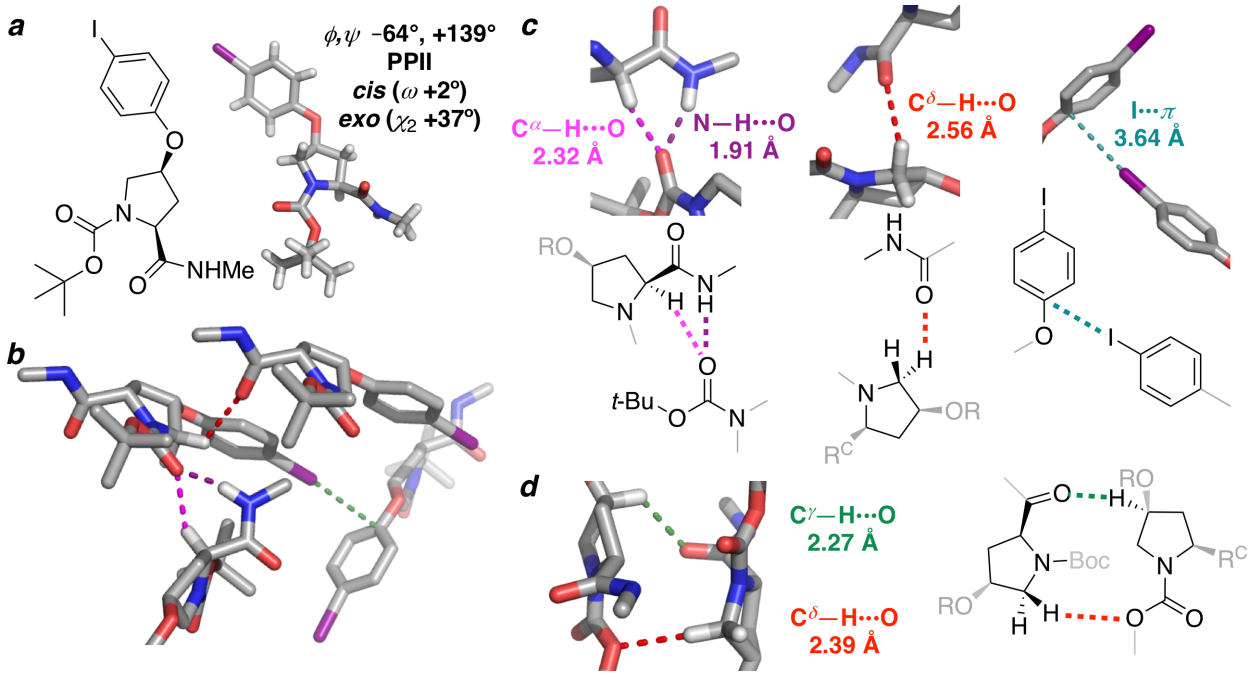


Figure 3. **Conformational analysis and intermolecular interactions within the crystal structure of 2**. (a) Crystal structure of Boc-(2*S*,4*S*)-(4-iodophenyl)-hydroxyproline methyl amide (Boc-hyp(4-I-Ph)-NHMe) (**2**). Diffractable crystals were obtained by slow evaporation at room temperature from a solution of **2** in hexanes. (b) Crystal packing, with significant intermolecular interactions highlighted. (c) These structures exhibit three distinct intermolecular interactions that are below the sum of the van der Waals radii for these atoms (sum of vdw radii of H and O = 2.72 Å).[43] A hydrogen bond is present between the amide NH and the Boc carbonyl of an adjacent molecule (N–H•••O distance 1.91 Å, purple). H^{α} participates in a C–H/O interaction with the Boc carbonyl (C^{α}–H•••O distance 2.32 Å, blue). H^{δ} interacts with the oxygen of the Pro carbonyl of the adjacent molecule (C^{δ}–H•••O distance 2.56 Å, red). An I/π interaction was also observed within the crystal structure (I•••$C_{aromatic}$ distance 3.64 Å, teal). Heavy atom positions were fixed and the positions of the hydrogens were optimized using the M06-2X DFT method with the Def2TZVP basis set and implicit $CHCl_3$ solvation. (d) Results of full geometry optimization of the dimeric structure that was observed crystallographically. This dimer was subjected to full geometry optimization with the M06-2X DFT method with the Def2TZVP basis set in implicit chloroform. Two close C–H/O interactions were observed in the fully optimized dimer. H^{γ} interacts with the Pro carbonyl (C^{γ}–H•••O distance 2.27 Å, green), and H^{δ} interacts with the oxygen of the carbamate (C^{δ}–H•••O distance 2.39 Å, red). An intermolecular hydrogen bond was also present within this optimized model, between the methyl amide hydrogen and the Boc carbonyl of the adjacent molecule (N–H•••O distance 2.03 Å).

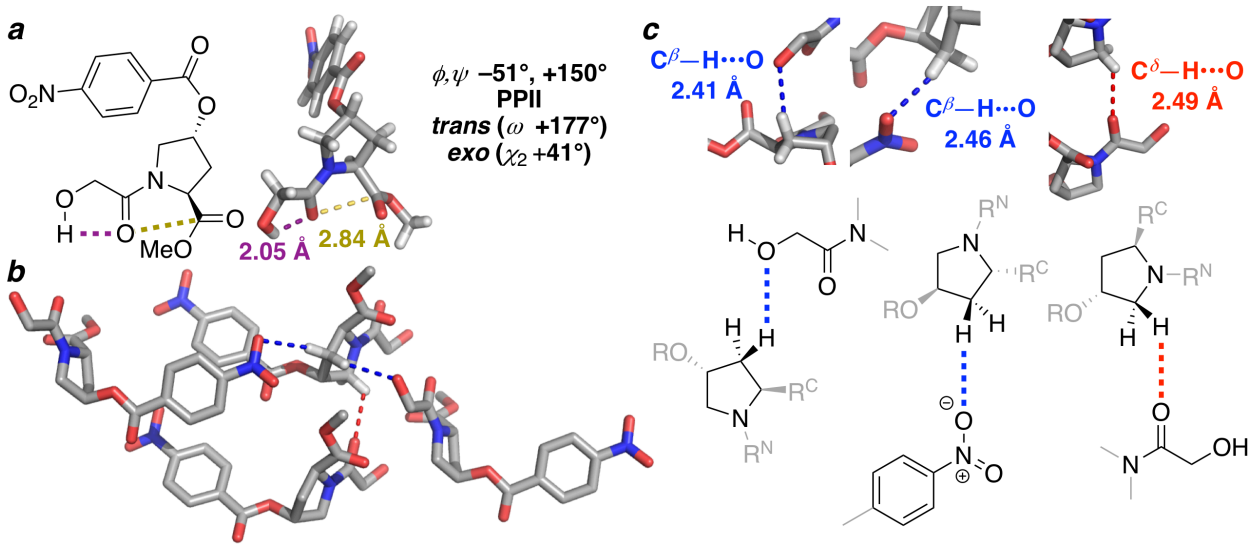


Figure 4. **Conformational analysis and intermolecular interactions within the crystal structure of 4**. (a) Crystal structure of glycolyl-(2*S*,4*R*)-(4-nitrobenzoyl)-4-hydroxyproline methyl ester (Glyc-Hyp(4-NO_2-Bz)-OMe) (**4**). Diffractable crystals were obtained by slow evaporation at room temperature from a solution of **4** in hexanes. This structure exhibits an intramolecular hydrogen bond between the glycolyl alcohol H and the glycolyl carbonyl O (purple) and an intercarbonyl n→π^* interaction ($C{=}O_{Glyc}$•••$C{=}O_{Hyp}$ distance 2.84 Å, gold). (b) Crystal packing, with significant intermolecular interactions highlighted. (c) This structure exhibits three distinct intermolecular C–H/O interactions that are below the sum of the van der Waals radii for these atoms. Two C^{β}–H/O interactions are present between $H^{\beta}$and oxygens of adjacent molecules. The *pro-S* H^{β} interacts with the oxygen of the glycolyl alcohol and the *pro-R* H^{β} interacts with the oxygen of the nitrobenzoate group (C^{β}–H•••O distances 2.41 and 2.46 Å respectively, blue). H^{δ} participates in a C^{δ}–H/O interaction with the glycolyl carbonyl of an adjacent molecule (C^{δ}–H•••O distance 2.49 Å, red). Heavy atom positions were fixed and the positions of the hydrogens were optimized using the M06-2X DFT method with the Def2TZVP basis set and implicit $CHCl_3$ solvation.

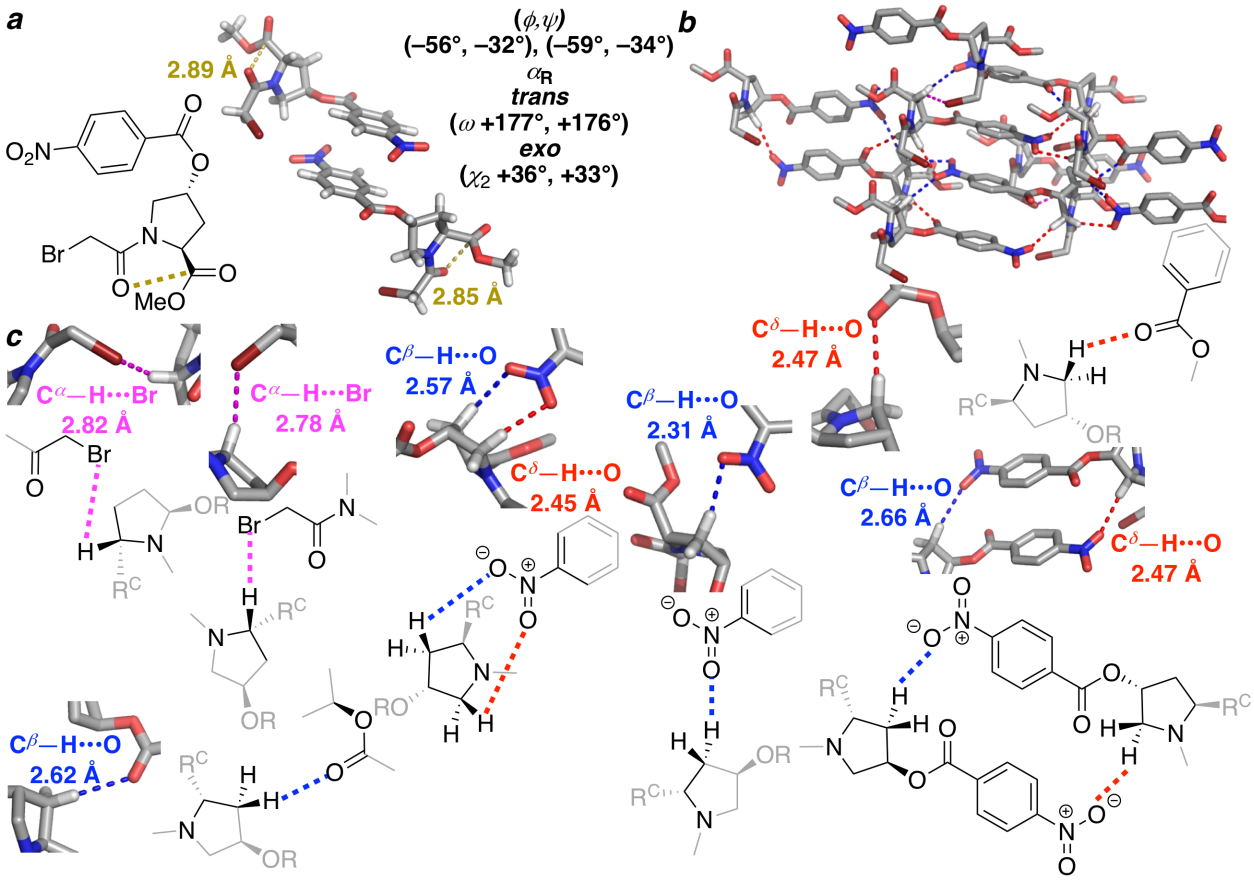


Figure 5. **Conformational analysis and intermolecular interactions within the crystal structure of 5**. (a) Crystal structure of bromoacetyl-(2*S*,4*R*)-(4-nitrobenzoyl)-hydroxyproline methyl ester (BrAc-Hyp(4-NO_2-Bz)-OMe) (**5**). Diffractable crystals were obtained by slow evaporation at room temperature from a solution of **5** in hexanes. Two independent molecules of **5** were observed within the unit cell, each in an α-helical conformation with an *exo* proline ring pucker. These molecules exhibit intercarbonyl n→π^* interactions ($C{=}O_{BrAc}$•••$C{=}O_{Hyp}$ distances 2.85 Å and 2.89 Å, gold). (b) Crystal packing, with significant intermolecular interactions highlighted. (c) These structures exhibit eight distinct intermolecular C–H/O or C–H/O-like interactions that are below the sum of the van der Waals radii for these atoms. Two C–H/O-like interactions occur between H^{α} and the bromine of an adjacent molecule (C^{α}–H•••Br distances 2.82 Å and 2.78 Å, magenta). Four C–H/O interactions occur between H^{β} and oxygens of adjacent molecules (C^{β}–H•••O distances 2.31 Å, 2.57 Å, 2.62 Å, and 2.66 Å, blue). Three interactions occur between H^{δ} and oxygens of adjacent molecules (C^{δ}–H•••O distances 2.31 Å, 2.45 Å, and 2.47 Å, red). Heavy atom positions were fixed and positions of the hydrogens were optimized using the M06-2X DFT method with the Def2TZVP basis set and implicit $CHCl_3$ solvation.

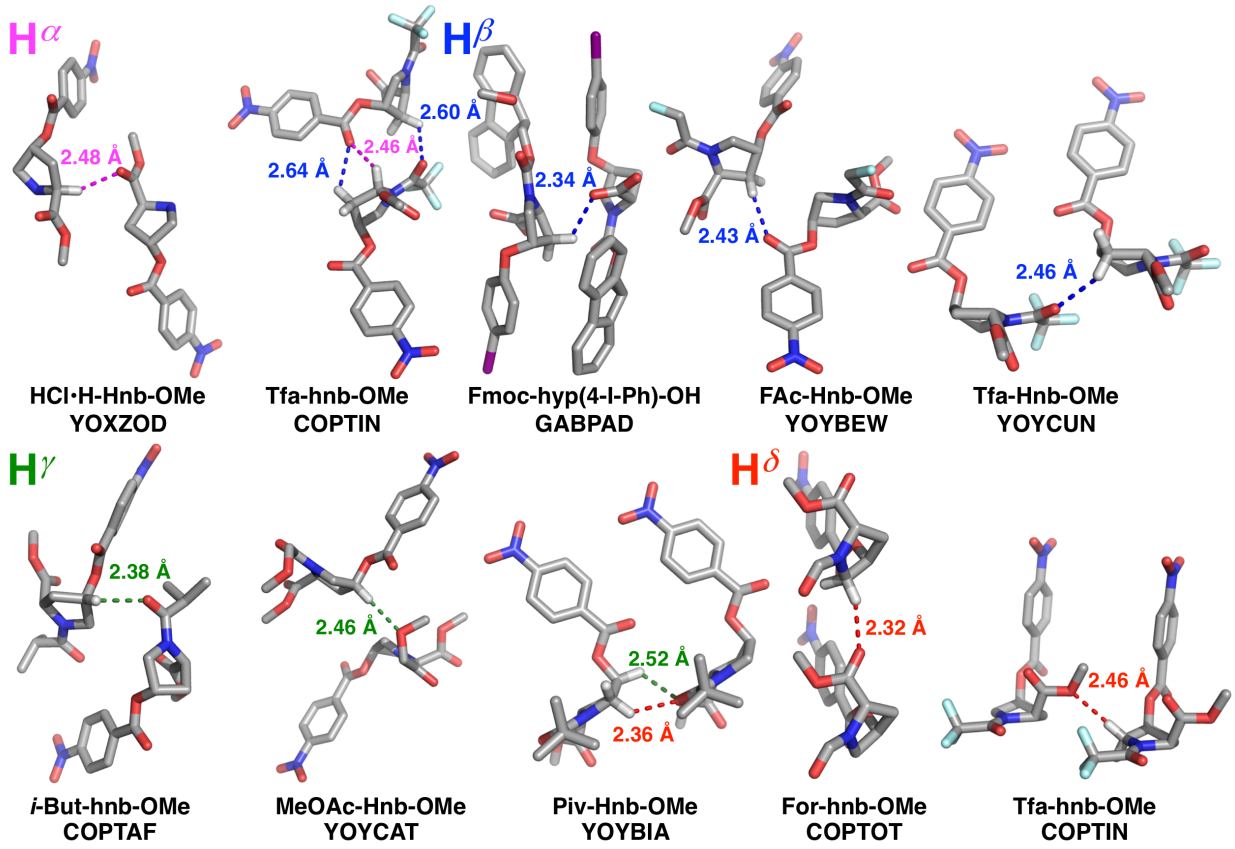


Figure 6. **Intermolecular C–H/O interactions at proline from previously reported crystal structures from the group**. Selected structures are shown with CSD identifiers. Heavy atom positions were fixed and the positions of the hydrogens were optimized using the M06-2X DFT functional with the Def2TZVP basis set and implicit $CHCl_3$ solvation. H•••O interaction distances are colored by proline ring position (H^{α}, magenta; H^{β}, blue; H^{γ}, green; H^{δ}, red). Hnb (upper case) indicates 4*R*-hydroxyproline nitrobenzoate ester, while hnb (lower case) indicates 4*S*-hydroxyproline nitrobenzoate ester. hyp(4-I-Ph) indicates the 4-iodophenyl ether of 4*S*-hydroxyproline (hyp).

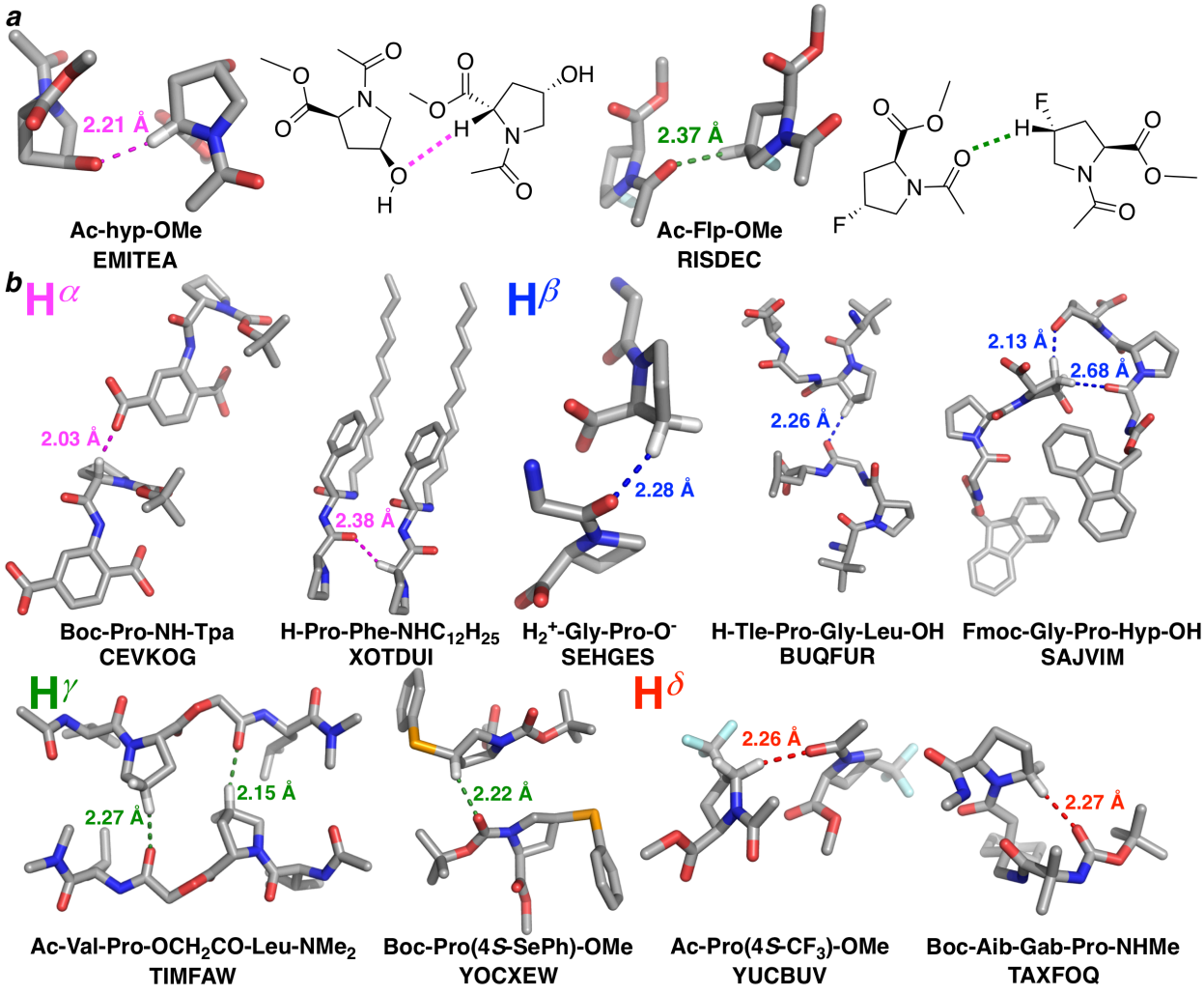


Figure 7. **Intermolecular C–H/O interactions at proline in published crystal structures**. (a) Selected structures from the CSD, with H•••O interaction distances indicated. Compounds shown are acetyl-(2*S*,4*S*)-4-hydroxyproline methyl ester (Ac-hyp-OMe) and acetyl-(2*S*,4*R*)-4-fluoroproline methyl ester (Ac-Flp-OMe) with C–H/O interactions at C^{α} and C^{γ}, respectively.[8c, 28a] (b) C–H/O interactions in the CSD. Selected structures are shown with the CSD code indicated.[28b-j] Hydrogen positions were normalized in Mercury. C–H/O interaction distances are colored by proline ring position (H^{α}, magenta; H^{β}, blue; H^{γ}, green; H^{δ}, red). Hyp (upper case) indicates 4*R*-hydroxyproline, while hyp (lower case) indicates 4*S*-hydroxyproline. Tpa = terephthalic acid, Gab = gabapentin.

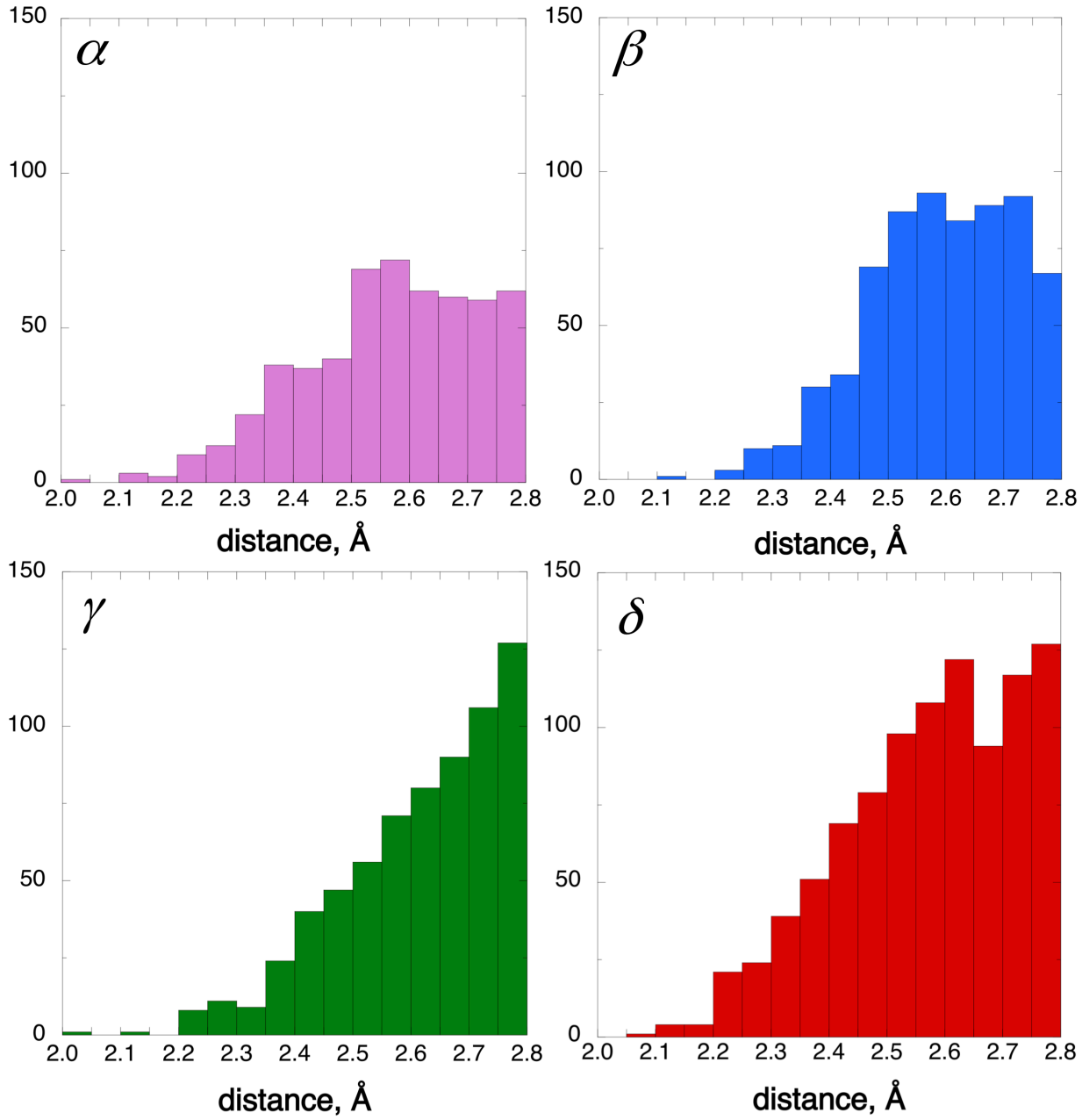


Figure 8. C–H/O interaction distances at proline within crystal structures in the CSD. Distances indicated are those between the proline hydrogens and interacting oxygen atoms. In the Cambridge Structural Database (CSD) were found the structures of 834 proline residues with hydrogens in a defined 3D contact with oxygen below the distance of 2.80 Å. Of these structures, 789 structures show an H•••O interaction distance of ≤ 2.72 Å, with many structures demonstrating multiple C–H/O interactions. 2,238 total C–H/O interactions with H•••O distances ≤ 2.72 Å were found. Search parameters excluded polycyclic proline residues. Distances were determined with the hydrogen positions normalized. The C–H/O interaction distances reported are derived from intermolecular interactions or intramolecular interactions across 5 bonds or more.

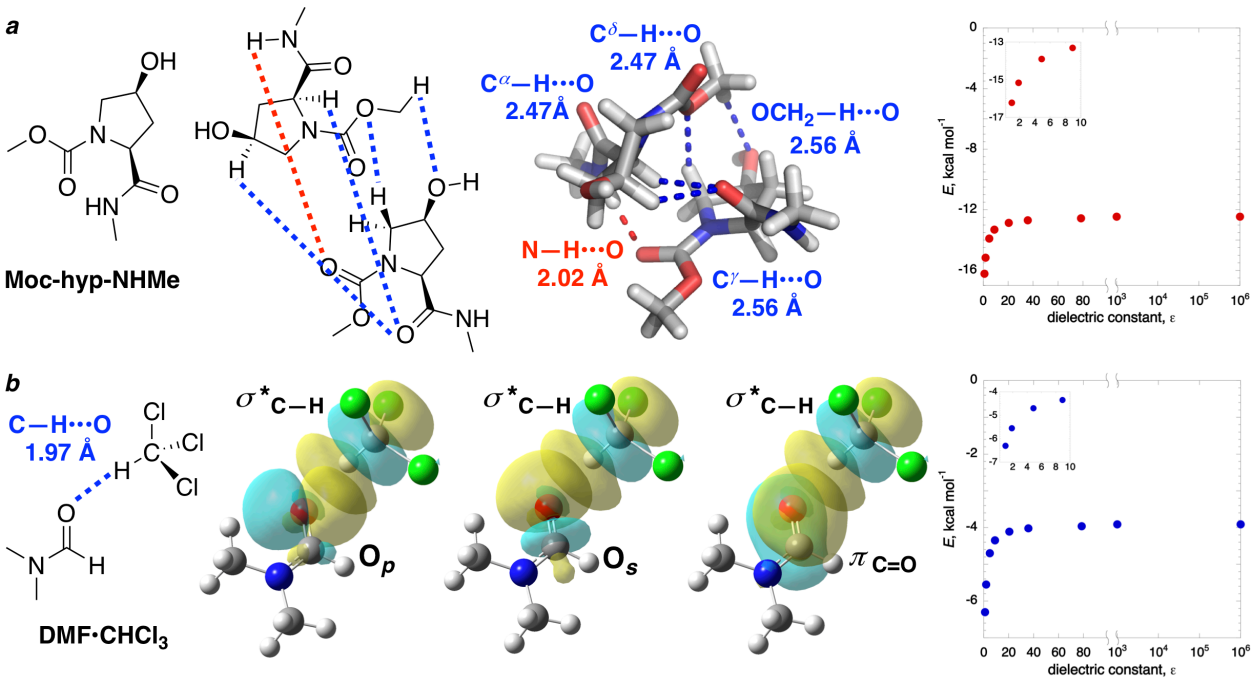


Figure 9. **Structures and energetics of model intermolecular C–H/O interactions**. (a) The geometry-optimized dimer of Boc-(2*S*,4*S*)-(4-iodophenyl)-4-hydroxyproline methyl amide [Boc-hyp(4-I-Ph)-NHMe] (**2**) was truncated to the methoxycarbonyl-(2*S*,4*S*)-hydroxyproline methyl amide [Moc-hyp(4-I-Ph)-NHMe] and the structure subjected to full geometry optimization using the M06-2X DFT functional and the Def2TZVP basis set in implicit $CHCl_3$. This structure exhibits an intermolecular hydrogen bond between the methyl amide H and the Moc carbonyl (N–H•••O distance 2.02 Å, red). Four distinct intermolecular C–H/O interactions are present within this dimer: H^{α} with the O of the Pro carbonyl, H^{γ} with the O of the Pro carbonyl, H^{δ} with an O of the methoxycarbonyl, and a H of the methoxy group with the hydroxyl oxygen of the adjacent molecule (C–H•••O distances 2.47 Å, 2.56 Å, 2.47 Å, and 2.56 Å respectively, blue). Energy calculations were performed on the geometry-optimized dimer of Moc-hyp-NHMe with solvation of dielectric constant varying from vacuum to an arbitrary solvent with ε = 1,000,000 using the MP2 method with the Def2TZVP basis set. Identical calculations were performed on the geometry-optimized monomer. The monomer energies were subtracted from the dimer energy to determine the intermolecular interaction stabilization energies for the Moc-hyp-NHMe dimer for each solvent [$E_{interaction} = E_{dimer} - 2(E_{monomer})$]. (b) The initial structure of the DMF•$CHCl_3$ complex was generated from the crystal structure with the CSD identifier VUPRII.[32] The (Cl_3)C•••O distance and the O-H-C(Cl_3) and C(O)-O-C(Cl_3) bond angles, and the (Cl_3)C-O-(O)C-H torsion angles were fixed to ensure the C–H/O interaction geometry present in the crystal structure was retained, while allowing optimization of bond lengths and angles to obtain a structure with minimal distortions. The model was subjected to this restrained geometry optimization using the MP2 method and the aug-cc-pVTZ basis set with implicit $CHCl_3$ (IEFPCM). The structures of $CHCl_3$ and DMF were independently subjected to full, unrestrained

geometry optimization using the same approach. The restrained-optimized complex was then subjected to NBO analysis, showing electron delocalization via orbital overlap between 4 occupied orbitals of the carbonyl [O_p, O_s, π, and σ (not shown)] and the $\sigma^*_{C–H}$ antibonding orbital. Energy calculations were conducted to determine the complex interaction energy of the restrained-geometry-optimized DMF•$CHCl_3$ complex as a function of solvent, with energies of the DMF•$CHCl_3$ complex and of the individual molecules of DMF and $CHCl_3$ determined using the MP2 method and the jul-cc-pV5Z[33] basis set and the indicated solvent. Using this combination of method and large basis set, the gas-phase basis set superposition error (BSSE) for DMF•$CHCl_3$ is 0.20 kcal mol^{-1}, indicating only a minor contribution of BSSE to these calculated electronic interaction energies. The dielectric constant of solvent was varied between vacuum ($\varepsilon = 1$) and an arbitrary solvent with $\varepsilon = 1{,}000{,}000$. Calculated energies for DMF and $CHCl_3$ were subtracted from the total energy of the DMF•$CHCl_3$ complex in each solvent condition to determine the intermolecular electronic interaction energies as a function of solvent.

Table 1. **C^{α}–H/O interactions from compounds with solved crystal structures from the group**.

molecule	CSD Identifier	C^{α}–H•••O distance, Å
Boc-hyp(4-I-Ph)-NHMe	[a]	2.32
Tfa-hyp(4-NO_2-Bz)-OMe	COPTIN	2.46
H_2^+-Hyp(4-NO_2-Bz)-O^-	DIZXOB	2.47, 2.57
HCl•H-Hyp(4-NO_2-Bz)-OMe	YOXZOD	2.48
F_3CCOOH•H-hyp(4-NO_2-Bz)-OMe	COPSOS	2.51
Boc-hyp(4-I-Ph)-OH	GABNUV	2.52
For-hyp(4-NO_2-Bz)-OMe	COPTOT	2.60
Boc-hyp(4-NO_2-Bz)-OMe	COPSEI	2.66
Ac-hyp(4-NO_2-Bz)-OMe	COPTEJ	2.68
Fmoc-hyp(4-I-Ph)-OH	GABPAD	2.69

[a] reported herein.

Table 2. **C^{β}–H/O interactions from compounds with solved crystal structures from the group**.

molecule	CSD Identifier	C^{β}–H•••O distance, Å
BrAc-Hyp(4-NO_2-Bz)-OMe	[a]	2.31, 2.57, 2.62, 2.66
Fmoc-hyp(4-I-Ph)-OH	GABPAD	2.34
ClAc-Hyp(4-NO_2-Bz)-OMe	YOYCIB	2.35, 2.66, 2.68, 2.69
Prp-Hyp(4-NO_2-Bz)-OMe	YOYBOG	2.37
F_2Ac-Hyp(4-NO_2-Bz)-OMe	YOYCEX	2.38, 2.55
For-Hyp(4-NO_2-Bz)-OMe	YOYBAS	2.40, 2.69
Ac-Hyp(4-NO_2-Bz)-OMe	YOXZUJ	2.40, 2.57, 2.68
Glyc-Hyp(4-NO_2-Bz)-OMe	[a]	2.41, 2.46
FAc-Hyp(4-NO_2-Bz)-OMe	YOYBEW	2.43, 2.46, 2.47, 2.52
H_2^+-Hyp(4-NO_2-Bz)-O^-	DIZXOB	2.44, 2.46
Tfa-Hyp(4-NO_2-Bz)-OMe	YOXCUN	2.46
F_3CCOOH•H-hyp(4-NO_2-Bz)-OMe	COPSOS	2.46
Boc-hyp(4-NO_2-Bz)-OMe	COPSEI	2.48
Piv-hyp(4-NO_2-Bz)-OMe	COPSUY	2.50
For-hyp(4-NO_2-Bz)-OMe	COPTOT	2.52
Boc-hyp(4-I-Ph)-NHMe	[a]	2.56
MeOAc-Hyp(4-NO_2-Bz)-OMe	YOYCAT	2.56, 2.67
Tfa-hyp(4-NO_2-Bz)-OMe	COPTIN	2.61

[a] reported herein.

Table 3. **C^{γ}–H/O interactions from compounds with solved crystal structures from the group**.

molecule	CSD Identifier	C^{γ}–H•••O distance, Å
i-But-hyp(4-NO_2-Bz)-OMe	COPTAF	2.38, 2.52
MeOAc-Hyp(4-NO_2-Bz)-OMe	YOYCAT	2.46, 2.51
H_2^+-Hyp(4-NO_2-Bz)-O^-	DIZXOB	2.47
For-Hyp(4-NO_2-Bz)-OMe	YOYBAS	2.50
Boc-Hyp(4-I-Ph)-OMe	GABPEH	2.50
H-hyp(4-NO_2-Bz)-OMe	COPSIM	2.50
Piv-Hyp(4-NO_2-Bz)-OMe	YOYBIA	2.52
Piv-hyp(4-NO_2-Bz)-OMe	COPSUY	2.54, 2.58
Glyc-Hyp(4-NO_2-Bz)-OMe	[a]	2.67

[a] reported herein.

Table 4. C^{δ}–H/O interactions from compounds with solved crystal structures from the group.

molecule	CSD Identifier	C^{δ}–H•••O distance, Å
FAc-Hyp(4-NO_2-Bz)-OMe	YOYBEW	2.27, 2.38, 2.39, 2.43
MeOAc-Hyp(4-NO_2-Bz)-OMe	YOYCAT	2.29, 2.34
H_2^+-Hyp(4-NO_2-Bz)-O^-	DIZXOB	2.30
Tfa-hyp(4-NO_2-Bz)-OMe	COPTIN	2.31, 2.46
BrAc-Hyp(4-NO_2-Bz)-OMe	[a]	2.31, 2.45, 2.47
For-hyp(4-NO_2-Bz)-OMe	COPTOT	2.32
For-Hyp(4-NO_2-Bz)-OMe	YOYBAS	2.33, 2.52
Ac-Hyp(4-NO_2-Bz)-OMe	YOXZUJ	2.33, 2.36, 2.40, 2.49
Piv-Hyp(4-NO_2-Bz)-OMe	YOYBIA	2.36
F_2Ac-Hyp(4-NO_2-Bz)-OMe	YOYCEX	2.36, 2.37, 2.39
ClAc-Hyp(4-NO_2-Bz)-OMe	YOYCIB	2.39, 2.43, 2.48, 2.53
Prp-Hyp(4-NO_2-Bz)-OMe	YOYBOG	2.40, 2.42, 2.45, 2.47
i-But-Hyp(4-NO_2-Bz)-OMe	YOYCOH	2.40, 2.42, 2.50, 2.71
Boc-hyp(4-I-Ph)-OH	GABNUV	2.45
Glyc-Hyp(4-NO_2-Bz)-OMe	[a]	2.49
Boc-Hyp(4-I-Ph)-OMe	GABPEH	2.50
F_3CCOOH•H-hyp(4-NO_2-Bz)-OMe	COPSOS	2.53
Boc-hyp(4-NO_2-Bz)-OMe	COPSEI	2.55
Ac-hyp(4-NO_2-Bz)-OMe	COPTEJ	2.55

[a] reported herein.

Table 5. **Summary data of C–H/O interactions within solved crystal structures with proline from the group**.

Group	number of structures	number of C–H/O interactions (≤ 2.72 Å)	minimum interaction distance	average C–H/O interaction length (≤ 2.72 Å)
All structures	26	102	2.27 Å	2.47 ± 0.11 Å
C^{α}–H/O	10	11	2.32 Å	2.54 ± 0.11 Å
C^{β}–H/O	18	34	2.31 Å	2.51 ± 0.11 Å
C^{γ}–H/O	9	12	2.38 Å	2.51 ± 0.07 Å
C^{δ}–H/O	19	41	2.27 Å	2.42 ± 0.09 Å

Table 6. **Summary data of C–H/O interactions within solved crystal structures of proline from the CSD**.

Position on Pro ring	number of structures with C–H/O interactions (≤ 2.80 Å)	number of structures with C–H/O interactions (≤ 2.72 Å)	number of C–H/O interactions with H•••O distances ≤ 2.80 Å	number of C–H/O interactions with H•••O distances ≤ 2.72 Å	average C–H/O interaction length (≤ 2.72 Å)
All structures	834	789	2849	2238	2.54 ± 0.13 Å
C^{α}–H/O	423	369	548	451	2.52 ± 0.13 Å
C^{β}–H/O	446	396	670	548	2.55 ± 0.11 Å
C^{γ}–H/O	459	350	672	472	2.55 ± 0.13 Å
C^{δ}–H/O	556	482	959	767	2.52 ± 0.14 Å

References

[1] a K. T. O'Neil, W. F. Degrado, *Science* **1990**, *250*, 646-651; b S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, R. L. Baldwin, *Nature* **1990**, *344*, 268-270; c C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **1994**, *33*, 5510-5517; d D. L. Minor, P. S. Kim, *Nature* **1994**, *367*, 660-663.

[2] a M. W. MacArthur, J. M. Thornton, *J. Mol. Biol.* **1991**, *218*, 397-412; b H. Reiersen, A. R. Rees, *Trends Biochem. Sci.* **2001**, *26*, 679-684; c R. Bhattacharyya, P. Chakrabarti, *J. Mol. Biol.* **2003**, *331*, 925-940.

[3] a A. A. Adzhubei, M. J. E. Sternberg, *J. Mol. Biol.* **1993**, *229*, 472-493; b A. L. Rucker, C. T. Pager, M. N. Campbell, J. E. Qualls, T. P. Creamer, *Proteins* **2003**, *53*, 68-75; c Z. Shi, K. Chen, Z. Liu, A. Ng, W. C. Bracken, N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17964-17968; d A. M. Brown, N. J. Zondlo, *Biochemistry* **2012**, *51*, 5041-5051.

[4] a D. S. Kemp, T. P. Curran, J. G. Boyd, T. J. Allen, *J. Org. Chem.* **1991**, *56*, 6683-6697; b D. A. E. Cochran, S. Penel, A. J. Doig, *Protein Sci.* **2001**, *10*, 463-470.

[5] P. Chakrabarti, S. Chakrabarti, *J. Mol. Biol.* **1998**, *284*, 867-873.

[6] a A. Bundi, K. Wüthrich, *Biopolymers* **1979**, *18*, 285-297; b D. S. Wishart, B. D. Sykes, F. M. Richards, *J. Mol. Biol.* **1991**, *222*, 311-333; c G. Merutka, H. J. Dyson, P. E. Wright, *J. Biomol. NMR* **1995**, *5*, 14-24.

[7] a Z. S. Derewenda, L. Lee, U. Derewenda, *J. Mol. Biol.* **1995**, *252*, 248-262; b M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel, R. Hilgenfeld, *J. Mol. Biol.* **2001**, *307*, 357-377; c N. J. Zondlo, *Acc. Chem. Res.* **2013**, *46*, 1039-1049; d S. Krimm, *Science* **1967**, *158*, 530-531; e G. R. Desiraju, *Acc. Chem. Res.* **1996**, *29*, 441-449; f R. K. Castellano, *Curr. Org. Chem.* **2004**, *8*, 845-865; g C. R. Jones, P. K. Baruah, A. L. Thompson, S. Scheiner, M. D. Smith, *J. Am. Chem. Soc.* **2012**, *134*, 12064-12071; h R. W. Driver, T. D. W. Claridge, S. Scheiner, M. D. Smith, *Chem. Eur. J.* **2016**, *22*, 16513-16521; iL. Jiang, L. H. Lai, *J. Biol. Chem.* **2002**, *277*, 37732-37740.

[8] a B. Schobert, J. Cupp-Vickery, V. Hornak, S. Smith, J. Lanyi, *J. Mol. Biol.* **2002**, *321*, 721-726; b B. J. Gibbons, T. D. Hurley, *Biochemistry* **2004**, *43*, 12555-12562; c A. M. Brzozowski, Z. S. Derenwenda, E. J. Dodson, G. G. Dodson, J. P. Turkenberg, *Acta Cryst.* **1992**, *B48*, 307-319; d N. Panasik, E. S. Eberhardt, A. S. Edison, D. R. Powell, R. T. Raines, *Int. J. Peptide Protein Res.* **1994**, *44*, 262-269.

[9] a K. Manikandan, S. Ramakumar, *Proteins* **2004**, *56*, 768-781; b S. Horowitz, R. C. Trievel, *J. Biol. Chem.* **2012**, *287*, 41576-41582; c A. Senes, I. Ubarretxena-Belandia, D. M. Engelman, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9056-9061; d B. K. Mueller, S. Subramaniam, A. Senes, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E888-E895.

[10] S. Scheiner, *J. Phys. Chem. B* **2005**, *109*, 16132-16141.

[11] a D. J. Barlow, J. M. Thornton, *J. Mol. Biol.* **1988**, *201*, 601-619; b J. Dhar, P. Chakrabarti, *Proteins* **2022**, *90*, 1159-1169.

[12] a G. Kleiger, R. Grothe, P. Mallick, D. Eisenberg, *Biochemistry* **2002**, *41*, 5990-5997; b A. Senes, D. E. Engel, W. F. DeGrado, *Curr. Opin. Struct. Biol.* **2004**, *14*, 465-479; c M. G. Teese, D. Langosch, *Biochemistry* **2015**, *54*, 5125-5135.

[13] a P. A. Karplus, *Protein Sci.* **1996**, *5*, 1406-1420; b S. A. Hollingsworth, P. A. Karplus, *Biomol. Concepts* **2010**, *1*, 271-283; c B. K. Ho, R. Brasseur, *BMC Struct. Biol.* **2005**, *5*.

[14] a D. J. Sutor, *Nature* **1962**, *195*, 68-69; b D. J. Sutor, *J. Chem. Soc.* **1963**, 1105-1110.

[15] a R. Taylor, O. Kennard, *J. Am. Chem. Soc.* **1982**, *104*, 5063-5070; b S. Sarkhel, G. R. Desiraju, *Proteins* **2004**, *54*, 247-259; c A. Ghosh, M. Bansal, *J. Mol. Biol.* **1999**, *294*, 1149-1158; d B. P. Klaholz, D. Moras, *Structure* **2002**, *10*, 1197-1204; e A. C. Pierce, K. L. Sandretto, G. W. Bemis, *Proteins* **2002**, *49*, 567-576; f D. B. McConnell, *J. Med. Chem.* **2021**, *64*, 16319-16327.

[16] a S. Scheiner, *Phys. Chem. Chem. Phys.* **2011**, *13*, 13860-13872; b C. R. Jones, P. K. Baruah, A. L. Thompson, S. Scheiner, M. D. Smith, *J. Am. Chem. Soc.* **2012**, *134*, 12064-12071; c O. O. Brovarets, Y. P. Yurenko, D. M. Hovorun, *J. Biomol. Struct. Dyn.* **2014**, *32*, 993-1022; d S. Scheiner, *J. Phys. Chem. A* **2015**, *119*, 9189-9199; e A. M. Vibhute, U. D. Priyakumar, A. Ravi, K. M. Sureshan, *Chem. Commun.* **2018**, *54*, 4629-4632; f T. Steiner, G. R. Desiraju, *Chem. Commun.* **1998**, 891-892.

[17] a C. R. Glendening, C. R. Landis, F. Weinhold, *WIREs Comput. Mol. Sci.* **2012**, *2*, 1-42; b E. D. Glendening, C. R. Landis, F. Weinhold, *J. Computational Chem.* **2013**, *34*, 1429-1437.

[18] a A. K. Pandey, G. P. A. Yap, N. J. Zondlo, *J. Org. Chem.* **2014**, *79*, 4174-4179; b N. V. Costantini, H. K. Ganguly, M. I. Martin, N. A. Wenzell, G. P. A. Yap, N. J. Zondlo, *Chem. Eur. J.* **2019**, *25*, 11356-11364; c N. A. Wenzell, H. K. Ganguly, A. K. Pandey, M. R. Bhatt, G. P. A. Yap, N. J. Zondlo, *ChemBioChem* **2019**, *20*, 963-967.

[19] a A. K. Pandey, D. Naduthambi, K. M. Thomas, N. J. Zondlo, *J. Am. Chem. Soc.* **2013**, *135*, 4333-4363; b C. R. Forbes, A. K. Pandey, H. K. Ganguly, G. P. A. Yap, N. J. Zondlo, *Org. Biomol. Chem.* **2016**, *14*, 2327-2346.

[20] a C. A. Hunter, J. K. M. Sanders, *J. Am. Chem. Soc.* **1990**, *112*, 5525-5534; b R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, F. M. Boeckler, *J. Med. Chem.* **2013**, *56*, 1363-1388.

[21] R. W. Newberry, R. T. Raines, *Acc. Chem. Res.* **2017**, *50*, 1838-1846.

[22] a E. J. Milner-White, L. H. Bell, P. H. Maccallum, *J. Mol. Biol.* **1992**, *228*, 725-734; b Y. K. Kang, H. Y. Choi, *Biophys. Chem.* **2004**, *111*, 135-142.

[23] J. Bella, H. M. Berman, *J. Mol. Biol.* **1996**, *264*, 734-742.

[24] N. J. Zondlo, *Phys. Chem. Chem. Phys.* **2022**, *24*, 13571-13586.

[25] G. J. Bartlett, A. Choudhary, R. T. Raines, D. N. Woolfson, *Nat. Chem. Biol.* **2010**, *6*, 615-620.

[26] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Cryst.* **2016**, *B72*, 171-179.

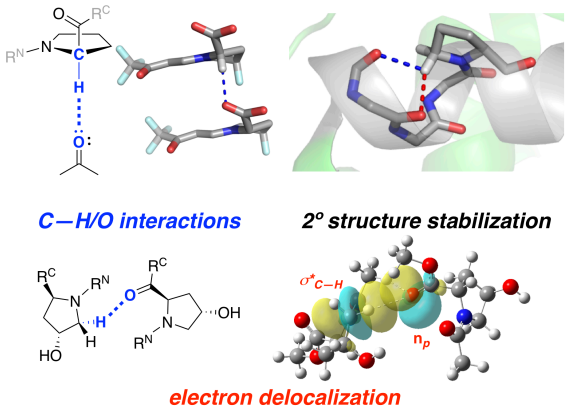
[27] G. Gilli, P. Gilli, *The Nature of the Hydrogen Bond: Outline of a Comprehensive Hydrogen Bond Theory*, Oxford University Press, New York, **2009**.

[28] a M. D. Shoulders, F. W. Kotch, A. Choudhary, I. A. Guzei, R. T. Raines, *J. Am. Chem. Soc.* **2010**, *132*, 10857-10865; b T. S. Haque, J. C. Little, S. H. Gellman, *J. Am. Chem. Soc.* **1996**, *118*, 6975-6985; c M. Marsch, C. Priem, A. Geyer, *CSD Communications* **2019**; d V. Kubyshkin, S. Afonin, S. Kara, N. Budisa, P. K. Mykhailiuk, A. S. Ulrich, *Org. Biomol. Chem.* **2015**, *13*, 3171-3181; e S. A. Moggach, S. Parsons, L. Sawyer, *Acta Crystallogr. E: Crystallogr. Commun.* **2006**, *62*, O1046-O1048; f N. A. Wani, V. K. Gupta, U. P. Singh, S. Aravinda, R. Rai, *Chemistry Select* **2016**, *1*, 1674-1677; g V. V. Veselovsky, A. V. Lozanova, V. I. Isaeva, A. A. Lobova, A. N. Fitch, V. V. Chernyshev, *Acta Crystallogr. C Struct. Chem.* **2018**, *74*, 248-+; h M. A. Dolan, P. N. Basa, O. Zozulia, Z. Lengyel, R. Lebl, E. M. Kohn, S. Bhattacharya, I. V. Korendovych, *ACS Nano* **2019**, *13*, 9292-9297; i Z. H. Du, B. X. Tao, M. Yuan, W. J. Qin, Y. L. Xu, P.

Wang, C. S. Da, *Org. Lett.* **2020**, *22*, 4444-4450; j M. Ghosh, S. Bera, S. Schiffmann, L. J. W. Shimon, L. Adler-Abramovich, *ACS Nano* **2020**, *14*, 9990-10000.
[29] a F. L. Hirshfeld, *Theor. Chem. Acc.* **1977**, *44*, 129-138; b A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 527-541.
[30] a R. Vargas, J. Garza, D. A. Dixon, B. P. Hay, *J. Am. Chem. Soc.* **2000**, *122*, 4750-4755; b S. Scheiner, T. Kar, Y. L. Gu, *J. Biol. Chem.* **2001**, *276*, 9832-9837; c H. Park, J. Yoon, C. Seok, *J. Phys. Chem. B* **2008**, *112*, 1041-1048.
[31] S. Scheiner, T. Kar, *J. Phys. Chem. B* **2005**, *109*, 3681-3689.
[32] D. S. Yufit, J. A. K. Howard, *CrysEngComm* **2010**, *12*, 737-741.
[33] E. Papajak, J. J. Zheng, X. F. Xu, H. R. Leverentz, D. G. Truhlar, *J. Chem. Theory Comput.* **2011**, *7*, 3027-3034.
[34] J. Sauer, R. Sustmann, *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 779-807.
[35] M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines, J. L. Markley, *J. Am. Chem. Soc.* **2002**, *124*, 2497-2505.
[36] a M. A. Trevino, D. Pantoja-Uceda, M. Menendez, M. V. Gomez, M. Mompean, D. V. Laurents, *J. Am. Chem. Soc.* **2018**, *140*, 16988-17000; b Z. P. Gates, M. C. Baxa, W. Yu, J. A. Riback, H. Li, B. Roux, S. B. H. Kent, T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 2241-2246.
[37] H. Yin, J. S. Slusky, B. W. Berger, R. S. Walters, G. Vilaire, R. I. Litvinov, J. D. Lear, G. A. Caputo, J. S. Bennett, W. F. DeGrado, *Science* **2007**, *315*, 1817-1822.
[38] A. W. Wijaya, A. I. Nguyen, L. T. Roe, G. L. Butterfoss, R. K. Spencer, N. K. Li, R. N. Zuckermann, *J. Am. Chem. Soc.* **2019**, *141*, 19436-19447.
[39] a M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford, CT, **2013**; b M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian, Inc., Wallingford CT, **2019**.

[40] a Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215-241; b F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305; c J. Tomasi, B. Mennucci, E. Cances, *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211-226.
[41] M. J. Frisch, M. Head-Gordon, J. A. Pople, *Chem. Phys. Lett.* **1990**, *166*, 275-280.
[42] K. Raghavachari, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650-654.
[43] A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441-451.

Graphical Table of Contents



C–H bonds of proline are particularly favorable sites for C–H/O interactions, including intermolecular assembly and intramolecular structure stabilization. These C–H/O interactions, which occur at all proline ring positions, exhibit short H•••O distances, and are significantly stabilized by stereoelectronic effects, with only a minor electrostatic contribution in solution.