Proline-Aromatic Sequences Stabilize Turns via C–H/π interactions in both *cis*-Proline and *trans*-Proline

Himal K. Ganguly, Michael B. Elbaum, and 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.

Abstract

In proteins, proline-aromatic sequences exhibit increased frequencies of *cis*-proline amide bonds, via proposed C–H/π interactions between the aromatic ring and either the proline ring or the backbone C–Hα of the residue prior to proline. These interactions would be expected to result in tryptophan, as the most electron-rich aromatic residue, exhibiting the highest frequency of *cis*-proline. However, prior results from bioinformatics studies on proteins and experiments on proline-aromatic sequences in peptides have not revealed a clear correlation between the properties of the aromatic ring and the population of *cis*-proline. An investigation of the effects of aromatic residue (aromatic ring properties) on the conformation of proline-aromatic sequences was conducted using three distinct approaches: (1) NMR spectroscopy in model peptides of the sequence Ac-TGPAr-NH_2 (Ar = encoded and unnatural aromatic amino acids); (2) bioinformatics analysis of structures in proline-aromatic sequences in the PDB; and (3) computational investigation using DFT and MP2 methods on models of proline-aromatic sequences and interactions. C–H/π and hydrophobic interactions were observed to stabilize local structures in both the *trans*-proline and *cis*-proline conformations, with both proline amide conformations exhibiting C–H/π interactions between the aromatic ring and Hα of the residue prior to proline (Hα-*trans*-Pro-aromatic and Hα-*cis*-Pro-aromatic interactions) and/or with the proline ring (*trans*-ProH-aromatic and *cis*-ProH-aromatic interactions). These C–H/π interactions were strongest with tryptophan (Trp) and weakest with cationic histidine ($HisH^+$). Aromatic interactions with histidine were modulated in strength by His ionization state. Proline-aromatic sequences were associated with specific conformational poses, including type I and type VI β-turns. C–H/π interactions at the pre-proline Hα, which were stronger than interactions at Pro, stabilize normally less favorable conformations, including the α or α_L conformations at the pre-proline residue, *cis*-proline, and/or the g^+ χ_1 rotamer or α_L conformation at the aromatic residue. These results indicate that proline-aromatic sequences, especially Pro-Trp sequences, are loci to nucleate turns, helices, loops, and other local structures in proteins. These results also suggest that mutations that introduce proline-aromatic sequences, such as the R406W mutation that is associated with protein misfolding and aggregation in the microtubule-binding protein tau, might result in substantial induced structure, particularly in intrinsically disordered regions of proteins.

Introduction

Proline-aromatic sequences, in which an aromatic residue follows proline, exhibit increased populations of *cis*-proline amide bonds compared to sequences in which a non-aromatic residue follows proline (Figure 1).[1-7] Stabilization of *cis* amide bonds in proline-aromatic sequences has been proposed to occur through C–H/π interactions between the aromatic ring and the proline ring (Hβ, Hγ, and/or Hδ) (a *cis*-proline-aromatic interaction) and/or with Hα or other hydrogens of the residue preceding the proline (an Hα-*cis*-proline-aromatic interaction) (Figure 1b).[8-14]

Proline-aromatic sequences are broadly present in proteins and within the bioactive sequence of peptide ligands of receptor proteins. For example, the μ-opioid receptor binds the opioid peptides endomorphin-1 (Y**PW**F-NH_2) and endomorphin-2 (Y**PF**F-NH_2).[15] These peptides prefer turn conformations, with different conformational preferences in an aqueous versus membrane environment.[16-19] The preference for multiple, distinct turn conformations provides these peptides with the ability to bind different targets. Endomorphin-1 is bound by the μ-opoid receptor with a *cis*-proline amide conformation (pdb 8f7r), while the bound structure of endomorphin-2 to dipeptidyl peptidase-3 (DPP3) exhibits *trans*-proline (pdb 5ehh), among the small number of structures of ligand-bound endomorphins.[15,20] These structural results recapitulate solution data on endomorphins, which show significant populations of *cis*-proline and distinct turn conformations in water versus in a membrane environment.[16-19,21] Other bioactive peptides with proline-aromatic sequences include neurotensin (with a PY sequence), gastrin (PW and PH), galanin (PH), helicokinin (PW), the vasodilation promoter bradykinin (PF), the vasoconstriction promoter angiotensin II (PF), and the antibacterial peptide indolicidin (3 PW sequences).[22,23] More generally, turn conformations are broadly implicated in ligand

binding by receptors, and thus sequences that promote turns are overrepresented in receptor ligands.[24] Beyond their direct biological importance, proline-aromatic sequences are also important in the design of peptide therapeutics, chemical probes, nanoparticles, and biomaterials.

In addition, proline-aromatic sequences are sometimes observed as disease-inducing mutations within proteins, without a clear mechanistic basis for the induction of disease phenotype. For example, in the microtubule-binding protein tau, the R406W mutation at the Ser_{404}-Pro_{405}-(Arg_{406} Trp_{406}) sequence is observed in familial FTDP-17 (frontotemporal dementia and parkinsonism linked to chromosome 17) and results in early-onset dementia associated with misfolding and aggregation of the tau protein.[25-29] The tau R406W mutation is also widely employed in mouse models of Alzheimer's disease, due to the early-onset aggregation and dementia phenotype.[30-33] The R406W mutation, which introduces a proline-aromatic (proline-tryptophan, Pro-Trp) sequence, occurs in an intrinsically disordered region of tau, and there is an absence of data supporting mechanisms by which this mutation leads to protein misfolding.

In the inverse aromatic-proline sequences (an aromatic residue *preceding* proline) (Figure 1c), the *cis*-proline conformation is stabilized by a C–H/ interaction between the aromatic ring and H and/or H of the Pro. This C–H/ interaction is tunable by the electronic properties of the aromatic ring, with more electron-rich aromatic residues significantly stabilizing the *cis* amide bond (with Trp-Pro sequences, the *cis* amide conformation is stabilized by > 1 kcal mol^{-1} compared to Ala-Pro).[11,34-36] A comparable analysis of the effect of aromatic electronic properties has not been conducted for proline-aromatic sequences (an aromatic residue *following* proline), despite the substantially greater likelihood of a *cis* amide bond in proline-aromatic sequences compared to proline-nonaromatic sequences and its potential importance in protein-protein

interactions, protein folding, and protein misfolding and disease. Herein, we examine the effects of aromatic residue identity and electronics on structure within proline-aromatic sequences, in order to better understand how proline-aromatic sequences influence protein structure, dynamics, and function.

Results

In order to further understand the roles of aromatic residues after proline in stabilizing the *cis* amide bond, we investigated these effects within the context of the model peptide Ac-TGPX-NH_2.[9] Gly was employed here at the pre-proline residue in order to focus these studies on inherent interactions of the aromatic ring with the protein backbone, and to avoid potential artefacts that could arise due to the presence of a side-chain group.

We initially examined the series of peptides Ac-TGPX-NH_2, where X = Arg, Phe, Asn, and cyclohexylalanine (Cha) (Figure 2, Figure 3, Table 1). Arg was employed as a polar post-proline residue which does not interact to stabilize *cis*-Pro, either in peptides by NMR or in proteins by bioinformatics.[1,4] Arg406 is also the native post-proline residue at the Ser404-Pro405 sequence of tau. Phe was employed as the simplest and most common aromatic amino acid, in order to examine the specific effects of aromatic residues at Pro-aromatic sequences. Cha is a hydrophobic, non-aromatic analogue of Phe. Cha allows the separation of the roles of the hydrophobic effect (which both Cha and aromatic amino acids exhibit) and C–H/π interactions, which are not possible in the aliphatic residue Cha.[11,34,35] Asn was also examined as a reference amino acid that, when following Pro, stabilizes *cis* amide bonds better than any non-aromatic and non-proline amino acid.[1,9] Asn also stabilizes *cis*-proline in TSPN sequences within the RNA

polymerase II C-terminal domain, as part of a code for phosphorylation-dependent proline *cis*-*trans* isomerization.[37]

All peptides were synthesized and subsequently analyzed by NMR spectroscopy, which allows quantification of the relative populations of species with *trans* and *cis* proline rotamers [$K_{trans/cis}$ = ([peptide with *trans*-proline]/[peptide with *cis*-proline])] as well as the relative energetics ($\Delta G_{trans/cis} = -RT \ln K_{trans/cis}$). These data indicated that Asn and Phe have specific interactions stabilizing a *cis* amide bond relative to Arg, consistent with prior results.[1,4] In addition, Cha was observed to stabilize the *cis* amide conformation relative to Arg and to Asn, though less than Phe. These data suggest that *cis*-proline-aromatic structures are stabilized *both* by the hydrophobic effect *and* by specific aromatic effects, in contrast to observations in aromatic-*cis*-proline structures, where Cha does not significantly stabilize the *cis* amide conformation.

The observation that Phe stabilizes the *cis* amide conformation relative to Asn was also observed in Ac-TAPX-NH_2 model peptides, with Ala prior to proline (Table 2 and Supporting Information). These results suggest that these observations are inherent to proline-aromatic sequences and are not unique to the Gly residue used in the Ac-TGPX-NH_2 model peptides. However, Phe and Trp had similar effects in stabilizing the *cis* amide conformation in proline-aromatic sequences. These results stand in contrast to aromatic-*cis*-Pro sequences, where the aromatic effect of Trp is far greater than that of Phe in stabilizing *cis*-Pro.[11,34,35] These results suggest a fundamental difference in how C–H/π interactions are operative in these two proline amide isomerization equilibria, *trans*-proline-aromatic/*cis*-proline-aromatic versus aromatic-*trans*-proline/aromatic-*cis*-proline interactions. Notably, in Ac-TAPW-NH_2, the Hα of Ala exhibited a significant upfield shift in both *trans*-Pro and *cis*-Pro. These results are consistent

with a ring current effect due to interaction of Ala Hα with the Trp aromatic ring in *trans*-Pro and *cis*-Pro.

In order to further understand the nature of *cis*-proline-aromatic interactions, an additional series of peptides was synthesized in the Ac-TGPX-NH_2 context, in which X = all canonical aromatic amino acids, including histidine in both physiologically relevant ionization states (Figure 2c). In addition, peptides were synthesized in which a series of X= 4-Z-phenylalanine amino acids was incorporated at the C-terminal amino acid, in order to understand the mechanisms by which aromatic residues after proline stabilize a *cis* amide bond.

All peptides were analyzed by NMR spectroscopy (Figure 4, Figure 5, Figure 6, Table 1, Table 3). The NMR data revealed that more electron-rich aromatic amino acids stabilized the *cis* amide conformation relative to Arg, Asn, Cha, 4-NO_2-Phe, 4-CF_3-Phe, or His. In particular, the peptide with cationic His ($HisH^+$) exhibited $K_{trans/cis}$ equivalent to that with Arg, indicating that the electron-poor aromatic group of cationic histidine does not stabilize the *cis* amide conformation via C–H/π interactions. However, 4-I-Phe, Phe, Tyr, and Trp exhibited similar values of $K_{trans/cis}$, despite the substantially different electronic properties of their aromatic rings. These results suggested that the roles of aromatic residues following proline might be due to the effects of the aromatic rings on multiple conformations.

The NMR spectra of the peptides in the Ac-TGPX-NH_2 series were further analyzed as a function of aromatic electronic properties. In particular, the chemical shifts (δ) of Gly Hα and Pro Hβ and Hγ were examined for evidence of interactions with the aromatic rings at these hydrogens (Figure 6, Table 3), the resonances of which would be shifted upfield due to aromatic ring current effects. Thus, stronger interactions with the aromatic ring result in larger upfield shifts at an interacting Gly or Pro resonance. In the *cis*-proline conformations, the Gly Hα

resonances exhibit dramatic upfield shifts, whose magnitude correlates with the electronic properties of the aromatic ring. For peptides with neutral histidine or 4-CF_3-Phe, relatively smaller upfield shifts (0.2 to 0.5 ppm upfield relative to the peptide with Arg) were observed. In contrast, these are nearly identical for peptides with Arg, Asn, or Cha. These upfield changes in are substantially greater in peptides with more electron-rich aromatic amino acids (Phe and Tyr, up to 1.0 ppm). In the peptide with X = Trp, one of the Gly H resonances is at 2.48 ppm, which is an upfield shift of 1.27 ppm in Gly H relative to the peptide with X= Arg.

In addition, the effect of aromatic interactions in inducing order can be observed via the dispersion in the chemical shifts of the diastereotopic Gly H. In general, greater dispersion was observed in *cis*-Pro than *trans*-Pro, with the most dispersion observed in the peptide with Trp. The extent of Gly H chemical shift dispersion correlated with the electronic properties of the aromatic ring, with more electron-rich aromatic residues exhibiting greater dispersion. Notably, cationic $HisH^+$ and neutral His were quite distinct, with larger aromatic effects on Gly H chemical shift observed for the more electron-rich aromatic ring of neutral His than the electron-poor aromatic ring of cationic $HisH^+$.

Upfield shifts were also observed in the *cis*-proline conformation at Pro H resonances, though these changes were substantially smaller. Collectively, these data indicate that, in *cis*-proline-aromatic conformations, the aromatic ring interacts strongly with the H of the residue prior to proline (an H-*cis*-proline-aromatic interaction), in a manner that correlates with aromatic electronic properties, as well as interacting less closely with the Pro H and Pro H in the *cis*-proline conformation (a *cis*-proline-aromatic interaction).

Given the strong correlation between aromatic electronic properties and the upfield δ change in Gly Hα in peptides with *cis*-proline, with the largest shifts in the most electron-rich aromatic amino acids, why is the correlation between aromatic electronics and induced *cis*-proline amide bond ($K_{trans/cis}$) weak? Stated alternatively, Trp stabilizes a *cis*-proline amide bond via a particularly favorable interaction with Hα of the residue prior to proline; why doesn't Trp even more strongly promote a *cis* amide bond compared to Phe, since the interaction with Trp is stronger?

Analysis of the equivalent chemical shifts with the *trans*-proline conformation suggested that aromatic amino acids following proline *also* interact with Hα of the residue prior to proline (Hα-*trans*-Pro-aromatic interaction*)* and with Hα of proline (*trans*-ProHα-aromatic interaction), with stronger interactions (= more upfield changes in δ) with more electron-rich aromatic residues. Chemical shift changes were also observed at Pro Hβ and Hγ, though these had trends that were less clear. Thus, aromatic amino acids following proline interact with Hα of the residue prior to proline and with the proline ring to stabilize *both* the *cis and* the *trans* conformations of proline. Moreover, both of these interactions are stronger with more electron-rich aromatic amino acids, in particular with Trp. These results indicate that Pro-Trp sequences might be particularly potent in inducing local structures, in both the *trans*-proline and *cis*-proline conformations.

Bioinformatics analysis of proline-aromatic sequences in the PDB. In order to obtain further insights into the conformational biases of proline-aromatic sequences in proteins, we conducted a bioinformatics investigation of high-resolution protein structures in the PDB with *trans*-proline-aromatic and *cis*-proline-aromatic conformations. This analysis examined the frequencies of *cis* amide bonds and the geometries of the interactions as a function of aromatic

amino acid (Trp, Tyr, Phe, His). These results were compared to Pro-Arg sequences, as a reference state that lacks local interactions stabilizing either proline amide conformation. Full details and extensive additional analysis are in the Supporting Information. The data (Table 4) indicated that proline-aromatic sequences are substantially more likely than Pro-Arg (4.8%) or generic Pro-Xaa sequences (5.3%) to adopt a *cis*-proline amide bond, consistent with prior analyses.[4,5] Notably, Pro-Tyr (9.2%) and Pro-Phe (8.1%) sequences are particularly likely adopt a *cis* amide bond.[12,13] Pro-Trp sequences exhibit an increased frequency (6.6%) for a *cis* amide bond relative to Arg or to a generic amino acid, but a reduced frequency relative to Tyr or Phe. These structural data are consistent with multiple factors in proline-aromatic sequences that can stabilize both *cis* and *trans* proline amide bonds.[8] In addition, these data from globular proteins are not necessarily representative of propensities in intrinsically disordered proteins.

A more detailed analysis was conducted of local proline-aromatic interactions that can stabilize both the *trans*-proline and *cis*-proline conformations (Figure 7). While Pro-His sequences also exhibited local proline-aromatic interactions with *trans*-Pro and *cis*-Pro in each interaction mode (Tables S29 and S33), Pro-His sequences were not included in the subsequent analysis due to the difficulty in assigning His ionization and tautomeric states in structures in the PDB. In conducting this analysis, we considered interactions of the aromatic ring with either the proline ring or with the Hα of the prior residue, each with either a *trans*-proline or a *cis*-proline. In bulk bioinformatics analysis, the positions of the heavy atoms were used to quantify the presence or absence of an interaction. When hydrogens are shown in structures herein, the hydrogen positions were determined using DFT calculations, with the positions of the heavy atoms fixed based on those observed crystallographically. Determination of the hydrogen

positions reveals that they often exhibit H•••C distances with aromatic ring carbons that are below the 2.90 Å sum of the van der Waals radii of H and C.

For all aromatic residues, Pro-aromatic sequences were characterized by close interactions of the aromatic ring in *both* the *trans*-proline *and* the *cis*-proline conformations (Figure 7, Figure 8). In Pro-aromatic structures with *trans*-proline, the aromatic ring exhibited close interactions with the Hα and/or Hδ of the pre-proline residue (Hα-*trans*-Pro-aromatic interaction, or more generally an interaction with the P–1 residue) and/or with one face of the proline ring, typically on Hβ, Hγ, or Hδ (*trans*-ProH-aromatic interaction, an interaction on Pro). In addition, in *trans*-proline structures, the aromatic ring was capable of significantly interacting with *both* the proline ring and with the prior residue (dual interactions, with both P–1 and Pro). 35% of *trans*-Pro-Trp sequences exhibited a *trans*-ProH-aromatic interaction and 33% exhibited an interaction with the P–1 residue. 21% of the structures exhibited close interactions of the aromatic ring both at the pre-proline residue and at proline (dual interactions). Similarly, 27% of structures with *cis*-Pro-Trp exhibited an interaction of the aromatic ring with Hα of the pre-proline residue (Hα-*cis*-Pro-aromatic interaction), with an additional 47% of structures exhibiting an interaction of Trp with the proline ring (*cis*-ProH-aromatic interaction). In *cis*-proline structures, dual interactions with both Pro and P–1 were not observed as defined geometrically herein. Overall, 43% of Pro-*trans*-Trp structures and 75% of Pro-*cis*-Trp structures exhibited at least one C–H/π interaction of the Trp aromatic ring with Pro or with the P–1 residue.

Similar analysis was conducted on structures with Tyr and Phe (Figure 8). For *trans*-Pro structures, the frequency of aromatic interactions was in order of aromatic electronics, Trp > Tyr > Phe. For *cis*-Pro structures, the frequencies of interactions were greater overall, and again had

the interaction frequency Trp > Tyr > Phe. These data are consistent with our observations by NMR spectroscopy in model peptides: in Pro-aromatic sequences, aromatic amino acids showed evidence of interaction with both the pre-proline and the proline residues, in both the *cis* and *trans* amide conformations, with interaction strength (based on changes in chemical shift compared to the peptide with Arg) related to aromatic electronic properties, Trp > Tyr ~ Phe > His > $HisH^+$. Trp exhibits the most frequent interactions with both pre-proline and proline, both in structures with *trans* and with *cis* amide bonds. The ability of Trp to stabilize structures with both *trans* and *cis* amide bonds explains why Trp promotes the *cis* amide conformation with lower frequency than Tyr or Phe in the PDB, despite the stronger aromatic interactions inherent to Trp.

The bioinformatics data also provided insights into how Pro-aromatic structures induce conformational biases in protein main-chain conformations, in both the *trans*-Pro (Figure 9) and *cis*-Pro (Figure 10) conformations.[38] In structures with *trans*-Pro amide bonds (Figure 9), C–H/π interactions were observed as apparently stabilizing in structures with turns (α-turn, β-turn, and γ-turn), helices (α- and 3_{10}-), and in less well-defined structures (coils, loops, bends, and extended conformations). C–H/π interactions were particularly associated with turn conformations, which are substantially overrepresented in structures with interactions and underrepresented in structures without interactions. More broadly, *trans*-Pro-aromatic C–H/π interactions were associated with a significant preference for Pro to be in the β conformation ((ϕ,ψ) ~ (–90°, 0°)), which is associated with type I and type II' β-turn conformations. These C–H/π interactions were less likely in PPII or extended conformations, especially for Trp (Figure 9e). In addition, the uncommon α conformation ((ϕ,ψ) ~ (–80°, +60°)) was associated with a

substantial frequency of *trans*-Pro-aromatic C–H/π interactions. For Trp, but less so for other aromatic amino acids, *trans*-Pro-aromatic C–H/π interactions were also significantly observed in the β-helix conformation. Bend structures were also predominantly associated with interactions.

The conformational preferences in structures with *cis*-Pro-aromatic C–H/π interactions were similarly analyzed (Figure 10). Here, C–H/π interactions were relatively more strongly associated with Pro adopting the PPII conformation. The site of interaction of the aromatic ring was closely associated with the main chain conformation: when the aromatic residue interacted with the P–1 residue, Pro was predominantly in the PPII region of the Ramachandran plot (Figure 11). In contrast, when the aromatic ring was interacting with the proline, Pro was almost exclusively in the α conformation. These conformations are closely associated with type VI β-turns, which have been previously observed to be stabilized by *cis*-proline-aromatic interactions.[2,8,9,39,40] When the aromatic ring interacts with proline, Pro is predominantly in a β conformation, which is present in type VIa1 and type VIa2 β-turns (PcisD and BcisD in the Dunbrack β-turn nomenclature).[41,42] In contrast, when the aromatic ring interacts with the P–1 residue, Pro is primarily in the PPII conformation as present in type VIb β-turns (BcisP), as well as in the α conformation.

Ramachandran plots of the aromatic amino acids as a function of C–H/π interaction mode (Figure 11, bottom) more broadly exhibited substantial differences from a standard Ramachandran plot. These results caused us to analyze the specific structures that were observed to be associated with C–H/π interactions, in both *trans*-Pro (Figure 12) and *cis*-Pro (Figure 13), as a function of aromatic amino acid. With *trans*-Pro, type I β-turns and γ-turns were associated with interactions of the aromatic ring with the P–1 residue, while type II β-turns were associated

with interactions with Pro. □-Turns and □-helix capping motifs were also observed with stabilizing interactions. With *cis*-Pro, type VIa1 □-turns could involve interactions at either the P–1 or Pro residue, while □-turns (type I-$_{□C}$) were associated with interactions with the proline ring (Figure 13). These interaction modes in different turns were observed for all aromatic amino acids. Most generally, a major effect of a Pro-aromatic sequence (with *trans*-Pro or *cis*-Pro) was to reduce the preferred conformational space via the interactions of the aromatic ring.

In addition, this analysis indicated that certain conformational poses were associated with different interaction modes (Figure 14, Figure 15). In particular, interactions of the aromatic ring with the P–1 residue appeared to be associated with conformations that are generally considered disfavored. With *trans*-Pro, interactions with the P–1 residue most commonly involved two unfavorable conformations: (1) the sterically disfavored[43,44] g^+ χ_1 rotamer at the aromatic amino acid; (2) the pre-proline amino acid in either the uncommon □ conformation or in the □$_L$ conformation, which is on the sterically disfavored right side of the Ramachandran plot. Although these conformations are disfavored, they are frequently observed when the aromatic ring interacts with the P–1 residue. In contrast, interaction of the aromatic ring at proline with *trans*-Pro can be accomplished with the more favorable g^- rotamer on the aromatic residue.

Similarly, with *cis*-proline, interaction on the P–1 residue was associated with having the aromatic amino acid in the higher-energy □$_L$ conformation. In contrast, interaction with Pro did not require a higher-energy conformation other than a *cis* amide bond. More broadly, interactions were infrequent (*trans*-Pro) or absent (*cis*-Pro) with the *t* rotamer on the aromatic residue.

Most generally, an interaction of the aromatic ring with the P–1 residue required *two* higher energy conformations: with *trans*-Pro, the combination of g^+ at the aromatic residue and wither the □$_L$ or □ conformation at the P–1 residue; and with *cis*-Pro, the combination of *cis*-Pro

and the α_L conformation at the aromatic residue. These observations suggest that the interaction of the aromatic ring with the P–1 residue can stabilize locally unfavorable conformations, and thus is likely a particularly favorable interaction to overcome the unfavorable conformational energy required to adopt these interaction geometries.

Computational investigation of C–H/π interactions in proline-aromatic sequences. In order to provide further insights into the effects of Pro-aromatic sequences on local protein structure, *trans*-Pro-aromatic and *cis*-Pro-aromatic interactions were also examined computationally. These minimal models were used to analyze Pro–aromatic or Hα-Pro–aromatic interactions as a function of interaction type and proline amide conformation.

We initially examined Ac-Ala-Pro-Phe-NHMe sequences, using minimal models derived from the structures identified via bioinformatics, with truncation of the P–1 side chain to Ala in order to identify general interaction modes rather than those that might primarily occur through interaction with the P–1 side chain, which were also observed in structures in the PDB. These initial models were then subjected to full geometry optimization calculations, using the M06-HF DFT functional and the Def2TZVP basis set in implicit water.[45-47] Calculations were focused on structures with local hydrogen bonds, as structures of loops or with longer-range interactions would be expected to be unstable during geometry optimization. A series of structures was obtained, with either *trans* or *cis* amide bonds, and for each with interaction either at the P–1 residue or at Pro (Figure 16). In addition, as controls, geometry optimization was performed on these peptides with all residues in a PPII conformation, and with the Pro residue in either an *exo* or *endo* conformation, with all three Phe χ_1 rotamers.

The resultant structures differ in the identities of their intramolecular hydrogen bonds and other noncovalent interactions, and the potential interactions of their main-chain atoms with

solvent molecules, precluding direct comparison of the energies of different structures. In order to obtain approximate interaction energies for each mode of interaction, geometry optimization was also conducted on a variation of each structure with rotation of the interacting Phe χ_1 rotamer to the trans (*t*) rotamer ($\chi_1 \sim 180^\circ$), in which the side chain cannot interact with either the P–1 residue or the proline ring (Figure 16a). In this case, both the interacting and the non-interacting rotamers have the same hydrogen bonds and the same approximate main-chain conformations, with the primary energy difference being the presence or absence of an interaction with the aromatic ring. While side-chain rotamers are not inherently identical in energy, typically all three rotamers are within 1.0 kcal mol^{-1} (as was also observed herein in the Ac-Ala-*trans*-Pro-Phe-NHMe structures in the PPII conformation, where the *t* rotamer was lowest in energy in the absence of any interactions [Table S73]).[43,44] Thus, the differences in rotamer energies are relatively small compared to the observed interaction energies, allowing approximation of the interaction energies.

Local energy minima with aromatic interactions were observed in both *trans*-Pro and *cis*-Pro conformations, each with an interaction either at the P–1 residue or at Pro (Figure 16b). Structures with interactions were observed with □-turn, □-turn, and □-turn geometries, including structures where the proline-aromatic interaction would stabilize 3_{10}- or □-helical conformations. These results specifically indicate that proline-aromatic sequences can nucleate local order, in both *trans*-Pro and *cis*-Pro, consistent with bioinformatics results.

The calculations also indicated that interactions of Phe with P–1 H□ were highly favorable (~ –5 kcal mol^{-1}), in both *trans*-Pro and *cis*-Pro (Figure 16b). In contrast, interactions at Pro, while favorable, were significantly weaker (–1.8 to –2.7 kcal mol^{-1}). However, the interactions at Pro are expected to come at a significantly lower entropic cost, due to fewer

required bond rotations to adopt a conformation where the Phe ring can interact with Pro. In addition, the interactions at P–1 all had a significant conformational energy cost: a combination of at least *two* disfavored conformations (α_L or β at (ϕ,ψ); g^+ at χ_1; and/or *cis*-Pro at ω) was observed in all interactions at P–1. In contrast, interactions at Pro occur with the favored g^- rotamer and with 0 or 1 disfavored conformations in the main chain. These observations are consistent with the larger inherent interaction strength at P–1 Hα, which allows these conformations to be observed (e.g. as is seen via bioinformatics) despite the unfavorable conformational energy. Indeed, these results suggest the aromatic interactions at P–1 can specifically promote the adoption of inherently disfavored conformations.

A subset of these interacting conformational poses was also examined in an Ac-Ala-Pro-Trp-NHMe context, in order to understand differences in proline-aromatic sequences with Phe versus with Trp (Figure 17). Interaction energies were again examined via comparison to the same structure with a non-interacting *t* Trp χ_1 rotamer. These structures indicated the particular favorability of dual interactions, with the aromatic ring interacting at both P–1 Hα and at Pro, for *trans*-Pro-Trp. Notably, for all interactions at the P–1 Hα, the interaction energies were significantly greater for Trp than for Phe, indicating that Pro-Trp sequences are particularly potent nuclei of local order. These results are consistent with the more electron-rich nature of the Trp ring, as well as its larger surface area for interaction. We also note that these electronic interaction energies do not include additional potential contributions due to the hydrophobic effect.

These calculations indicated that both *trans*-Pro and *cis*-Pro had the potential for strong interactions, with Hα-*trans*-Pro-Hβ-Trp and Hα-*cis*-Pro-Trp geometries exhibiting the largest complex interaction strengths. In *trans*-Pro, the Trp indole can interact *both* with Hα of the pre-

proline residue *and* with one face of the proline ring. In contrast, with *cis*-Pro, the Trp indole ring interacts *either* with Hα of the pre-proline residue *or* with one face of the proline ring. In addition, due to the size of the indole aromatic ring, additional interactions were observed with Trp in both proline amide conformations, including with P–1 Hα and with the main chain amide N–H (N–H/π interactions[48]). Notably, many of these aromatic interactions exhibited hydrogen-carbon or hydrogen-nitrogen distances that were well below the 2.90 Å and 2.75 Å sums of the van der Waals radii of H and C or of H and N, respectively. These short distances are consistent with a substantial stereoelectronic component to these C–H/π and N–H/π interactions. The closest C–H/π interactions were observed at C–Hα, which is the most polarized C–H bond on proline, with the most δ^+ charge on H and with the lowest energy σ^*_{C-H}, which would be most favorable for interaction via stereoelectronic effects due to more similar orbital energies with the (electron-rich) aromatic π molecular orbitals.

Finally, in order to more directly examine the C–H/π interaction energies involved, we generated minimal structures, derived from both interaction modes in type I and type VI β-turns, that represent Hα•Phe and Pro•Phe interactions with both *trans*-Pro and *cis*-Pro (Figure 18). These truncated structures convert the intramolecular interaction into an intermolecular complex, whose complex energy can then be calculated directly. These calculations indicate that aromatic interactions at either P–1 or at Pro inherently are significantly energetically favorable. In structures derived from either *trans*-Pro or *cis*-Pro, the interactions at P–1 are stronger than those at Pro, likely due to interaction at the more polarized C–Hα. Indeed, the closest H•••$C_{aromatic}$ distances (all < 2.90 Å) are observed in P–1 interactions (Figure 18; see also Figure 17). In contrast, the contacts at Pro are somewhat more distant, with H•••C distances often > 2.90 Å. An important caveat is that these *electronic* energies do not include additional favorable

contributions due to the hydrophobic effect, nor do they include the unfavorable conformational energy required to adopt the structures, and therefore they are not free energies of interaction. However, these computational results are consistent with the experimental data, suggesting that aromatic contacts are closer or more favorable when at the P–1 residue compared to at Pro, with *either* proline amide conformation.

Discussion

Dyson, Wright, Creighton, Basu, we, and others have all identified that an aromatic residue after proline promotes a *cis*-proline amide bond.[1,3,4,7,9,10,12,13,36,39,49-52] However, a systematic analysis of the impact of the post-proline aromatic residue identity on structure which includes non-canonical amino acids has not yet been conducted. We therefore used the Ac-TGPX-NH_2 model peptide context to examine how the identity of the residue after proline can promote or stabilize *cis* amide bonds. We examined both native aromatic amino acids (Trp, Tyr, Phe, neutral His, and cationic His [$HisH^+$]) and a series of unnatural 4-substituted aromatic amino acids, which allow systematic variation in the electronic properties of the aromatic ring. These peptides were compared to peptides with C-terminal Arg (a residue that does not engage in interactions with the backbone and that is natively present post-proline in *tau*) or Asn residues. Asn promotes *cis* amide bonds and turn conformations, and is present in our type VI β-turn Ac-TXPN-NH_2 model peptides.[1,9,35,37,39]

When we examined the effect of the post-proline aromatic residue on the proline *cis*-*trans* isomerization equilibrium in Ac-TGPX-NH_2 (X = aromatic amino acid) peptides, we observed complicated effects that were not readily explained by a single mechanism. The replacement of Arg with Phe resulted in a significant increase in the population of *cis* amide

bond. However, in contrast to prior data in aromatic-proline sequences,[11,34,35] the hydrophobic non-aromatic amino acid cyclohexylalanine (Cha) also significantly increased the population of *cis* amide bond, suggesting a partial role for the hydrophobic effect in stabilizing *cis*-proline-aromatic interactions. Moreover, there was only a limited correlation between the aromatic amino acid electronic properties and the *cis* amide bond population. For example, at the post-proline position, 4-NO_2-Phe and 4-I-Phe exhibited an increased population of *cis* amide bond compared to Cha, and neutral His exhibited a higher population of *cis* amide bond than cationic His ($HisH^+$). Notably, these data represent another example where histidine protonation state can impact local structure via the effects on proline amide conformation.[53-55]

However, peptides with 4-I-Phe, Phe, Tyr, or Trp as the post-proline residue exhibited very similar populations of *cis*-Pro amide bond, despite the very substantial differences in the electronic properties of their aromatic rings. Pro-aromatic sequences stabilize the *cis* amide conformation via interactions of the aromatic rings with Hα of the residue before proline (pre-proline [P–1] residue) and/or with the face of the proline ring. Further examination of the NMR spectra revealed that in the *cis* conformation, the Gly Hα resonances were shifted upfield with an aromatic residue in the post-proline position, and that the most electron-rich aromatic residues resulted in more substantial upfield shifts (the most upfield Gly Hα of 2.48 ppm was observed in the peptide with Trp) (Figure 5, Table 3). The chemical shifts of the Pro Hα also exhibited upfield shifts consistent with their interaction with an aromatic ring, although the effects on Pro Hα were substantially less than those on Gly Hα. The conformational restriction due to aromatic interactions was also observed via increases in dispersion of the diastereotopic Gly Hα, which strongly suggests main-chain ordering at Gly. In addition to aromatic effects on the *cis* conformation, we also observed aromatic effects on the chemical shifts of both Gly Hα and Pro

Hα in the *trans* conformation. The magnitudes of the chemical shift changes observed for a single amino acid were less in *trans*-Pro than those in *cis*-Pro. However, in both cases, the effects were dependent on aromatic ring electronic properties, with the largest effects observed in the peptide with Trp.

The NMR data thus indicated that a post-proline aromatic residue interacts to stabilize structures in both the *trans* and *cis* amide conformations, via interactions of the aromatic ring with the pre-proline Hα and/or with the face of the proline ring (Figure 15). Analysis of the PDB revealed that these trends in model peptides are broadly observed in globular proteins, with Pro-Phe and Pro-Tyr sequences the most likely to adopt a *cis* amide bond (Table 4).[4,52] Moreover, Pro-aromatic sequences exhibited Hα-Pro-aromatic and/or Pro-aromatic interactions in both the *trans*-Pro and *cis*-Pro conformations, with Trp exhibiting the highest frequencies of these local interactions (Figure 9, Figure 10). In addition, these Pro-aromatic or Hα-Pro-aromatic interactions were frequently observed in turn conformations, in structures with both *trans*-Pro and *cis*-Pro, as has previously been described for aromatic-*cis*-Pro structures.

Based on the totality of the analysis, we conclude that Pro-aromatic sequences promote local ordered structures, including turns and elements that can promote secondary structure (Figure 12, Figure 13). These effects are inherent to these sequences, and result in stabilized structures and increased order in both the *trans*-Pro and *cis*-Pro conformations. The aromatic ring can locally interact either with the proline residue or with the residue prior to proline. In *trans*-Pro, the aromatic ring also is capable of dual interactions at both the pre-proline residue and at proline. These results also imply proline-aromatic sequences as important sites for nucleation of local structure in intrinsically disordered proteins (IDPs).

Interactions at the proline ring are inherently more entropically favorable, requiring fewer numbers of restrictions in bond rotation, including interaction with the aromatic residue in a g^- conformation.[3] In contrast, interactions at the pre-proline residue Hα appear to be significantly stronger (when considering only the aromatic interaction), typically involving closer interactions between the aromatic ring and the hydrogens at P–1. These closer interactions specifically stabilize conformations that are typically higher energy, including conformations on the right side of the Ramachandran plot, the β conformation, the less favorable g^+ χ_1 rotamer, and/or a proline *cis* amide bond. Indeed, in structures with an aromatic ring interaction at the P–1 residue, *two* disfavored conformations are required. Thus, the interactions at P–1 must be strong enough to overcome the unfavorable energy of those conformations. Stated alternatively, the P–1•aromatic interaction is strong enough to promote disfavored regions of conformational space.

Why is the interaction of aromatic residues at the pre-proline residue so energetically favorable? Analysis of the PDB and computational investigations indicate that the C–H/π interactions at the Hα on the P–1 residue are particularly favorable, with distances between the hydrogen and the closest carbons of the aromatic ring commonly below the 2.90 Å sum of the van der Waals radii of H and C. Distances this close are consistent with a considerable molecular orbital (stereoelectronic) component to the interaction, with the close distances associated with through-space electron delocalization. Conceptually, the electron delocalization can be considered as occurring via alignment of the σ^* antibonding orbital of the polarized C–H bond with the π molecular orbitals of the aromatic ring, resulting in interresidue electron delocalization via orbital mixing. This mode of stabilization is analogous to Lewis acid-Lewis base interactions, which involve orbital overlap of a filled molecular orbital with an empty molecular orbital to generate a more stable complex. A similar basis for stabilization of S–H/π

interactions was recently described.[56] In both C–H/π interactions and S–H/π interactions, due to the low electronegativity of the atom bonded to H, the partial positive charge (δ^+) on the H is small.

Indeed, NBO analysis of minimal models (Figure 18b) is consistent with a stereoelectronic component to these C–H/π interactions. Moreover, calculations of interaction energies of these complexes indicate only a modest reduction in interaction strength in H_2O versus in vacuum. Although C–H/π interactions are typically described in electrostatics terms (here, interaction of the δ^- π face of the aromatic ring with the δ^+ of the hydrogen), the small charges involved mean that these interactions should be exceptionally weak in water and other polar solvents if electrostatics were the primary energetic component. The common observation of these interactions in water (*vide infra*) indicates that electrostatics cannot be the primary driving force of these interactions.[57] The hydrophobic effect and dispersion certainly contribute to interaction strength in water (see, for example, the ability of the non-aromatic residue Cha to promote *cis*-proline in Ac-TGPCha-NH_2 model system peptides). However, the particularly close nature of these interactions, and their significant sensitivity to aromatic electronic effects, suggests that these explanations are insufficient to fully describe the nature of these interactions.

Interactions of aromatic residues in proline-aromatic sequences with the P–1 residue and/or the proline residue were demonstrated to be broadly observed across all canonical aromatic amino acids. The weakest interactions were observed with cationic $HisH^+$, with significantly stronger interactions present with neutral His. The strongest interactions were present with Trp. These data indicate that Pro-Trp sequences should be particularly potent sites to nucleate order in peptides and proteins.

Experimental

Peptide synthesis and characterization. Peptides were synthesized by solid-phase peptide synthesis by standard methods using Rink amide resin. All peptides were acetylated on the N-terminus and contained C-terminal amides. Additional details of peptide synthesis and characterization are in the Supporting Information.

^{1}H NMR spectroscopy. Peptides were characterized using 1-D and TOCSY ^{1}H NMR spectroscopy. NMR spectra were recorded on a 600 MHz spectrometer in solutions of 90% H_2O/10% D_2O containing 5 mM phosphate buffer (with pH 4.0, 6.5, 7.2, 8.0, or 8.5, as indicated), 25 mM NaCl, and 0.1 mM TSP. All spectra were referenced to TSP. Additional experimental details and key spectral data, including key resonance assignments for both *trans*-proline and *cis*-proline rotamers, are in the Supporting Information.

PDB analysis. The PISCES server was used to generate a data set of proteins from a subset of the PDB for analysis.[58] Parameters for selection included proteins solved by X-ray crystallography with resolution ≤ 2.0 Å, with an R-factor ≤ 0.25, and with sequence identity ≤ 25%. In-house scripts were written for analysis, with visualization by Pymol. For distinguishing between *trans*-proline and *cis*-proline amide bonds, the ω torsion angle was examined, with *trans*-proline defined as $|\omega| > 150°$ and *cis*-proline defined as $-90° < \omega < +90°$. Secondary structures were determined using PROMOTIF.[38] Additional details and analysis are in the Supporting Information.

Computational analysis. Protein structures from the PDB were truncated to minimal *trans*-Pro-aromatic, and *cis*-Pro-aromatic structures via the following general approach. Hydrogens were added to proteins using Pymol, if necessary. The resultant all-atom protein structures were reduced to tripeptide or tetrapeptide fragments within Pymol. These peptide

fragments were further reduced in size to tripeptides within GaussView 5.[59] For more accurate determination of hydrogen atom positions, these minimal structures were then subjected to iterative restrained geometry optimization calculations, in which the crystallographically determined heavy atom positions were fixed, with geometry optimization of the positions of the hydrogen atoms only, with final geometry optimization conducted using the M06-HF DFT functional and the Def2TZVP basis set in implicit water, using Gaussian09.[45,46,59] The IEFPCM continuum polarization solvation model was used for all calculations herein.[47] These models with hydrogen positions optimized were used in Figures 7, 9, 10, 12, 13, and 15.

In order to more rigorously analyze these interactions, full, unrestrained geometry optimization calculations were conducted on a subset of these identified structures. The PDB-derived structures were used as initial models, the side chain of the pre-proline residue was truncated to Ala, and iterative geometry optimization was conducted, allowing all atom positions to move during the optimizations. Final geometry optimization was conducted with the M06-HF method and the Def2TZVP basis set in implicit water.

For each of these geometry-optimized structures with an aromatic interaction (Figure 16), the structure was additionally modified by rotating the χ_1 from a position where it interacts with the P–1 position or with the proline ring (all in either the g^- or g^+ rotamers), to the *t* rotamer (χ_1 = 180˚), where the aromatic ring is non-interacting, and the resultant models subjected to full geometry optimization using the same methods. The energies of the resultant structures (interacting and non-interacting forms of each conformational pair) were determined using the M06-HF DFT functional and the Def2TZVP basis set in implicit water. The interaction energy for each conformational pose was then determined via a comparison of the electronic energy of

the structure with an interaction versus the energy of the structure with the non-interacting *t* rotamer:

$E_{int} = E_{interacting\ rotamer} - E_{non\text{-}interacting\ rotamer}$. The structures of the interacting rotamers and the E_{int} for each structural interaction mode are indicated in Figures 16 and 17. Additional details are in the Supporting Information.

Finally, as an alternative approach to determine approximate electronic interaction energies, a subset of fully geometry-optimized structures with H-aromatic or Pro-aromatic interactions, with each mode in either a type I -turn (*trans*-Pro) or type VIa1 -turn (*cis*-Pro), was further truncated. Here, atoms were removed from each structure within GaussView, with hydrogens added to open valences, until a small representative bimolecular complex was obtained that had minimal steric clashes between the component parts at the atoms not directly involved in the interaction with the aromatic ring. These structures were then subjected to a restrained geometry optimization that only allowed the positions of newly added hydrogen atoms or the modified carbonyls to change. These resultant complexes were then analyzed via energy, counterpoise, and natural bond orbital (NBO) methods. Energy and counterpoise calculations were conducted with the MP2 method and the Def2QZVP basis set.[60] Counterpoise calculations in vacuum were used to quantify gas-phase interaction energies and basis-set superposition energies (BSSE).[61,62] All other calculations were conducted in implicit water (IEFPCM continuum polarization model). NBO analysis was conducted using NBO6 with the M06-HF DFT functional and the Def2QZVP basis set.[63,64] Additional details are in the Supporting Information.

Acknowledgements

We thank the DOD CDMRP PRARP program (AZ140115), NSF (CHE-1412978 and CHE-2004110), and NIH (GM093225) for funding. Instrumentation support was provided by NIH (GM110758) and NSF (CHE-1229234).

Supporting Information Available

Synthesis and characterization data of all peptides, additional 1-D and 2-D NMR spectra, tabulation of NMR data, additional bioinformatics data, additional details on calculations, and coordinates of all structures resulting from full geometry optimization calculations. This material is available free of charge via the Internet at the journal web site.

Figures

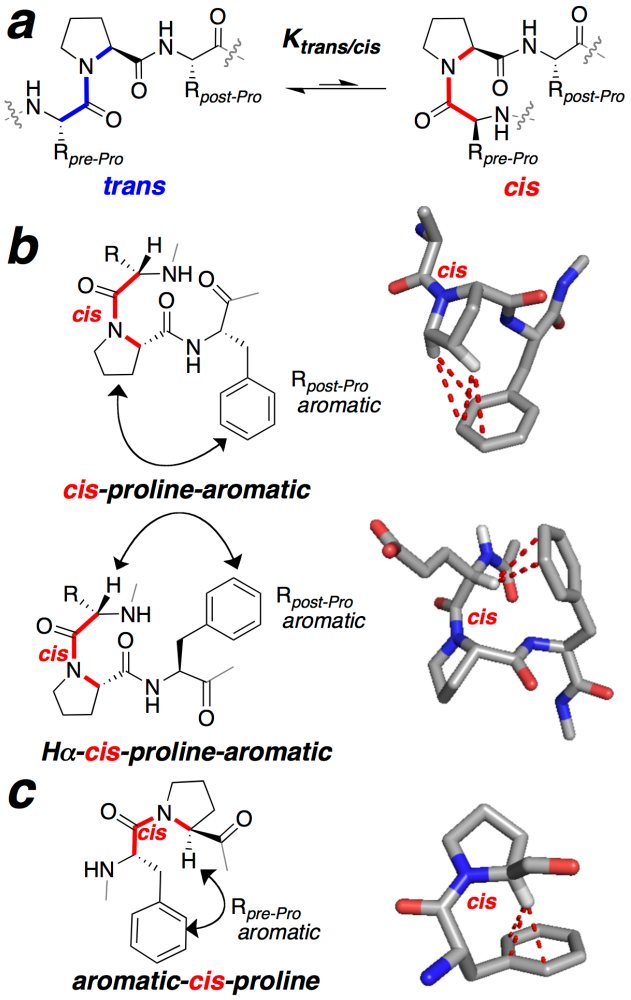

Figure 1. *cis-trans* isomerization of proline. (a) The nature of the amino acids preceding and following the proline residue influences the equilibrium constant, $K_{trans/cis}$. (b) Proline-aromatic sequences stabilize the *cis*-Pro conformer via (top) *cis*-proline-aromatic interactions (aromatic ring interacts with the proline ring) (pdb 2noo) or via (bottom) Hα-*cis*-proline-aromatic C–H/π interactions (aromatic ring interacts with the residue prior to proline) (pdb 3or1). (c) Aromatic-*cis*-proline C–H/π interactions stabilize the *cis*-proline conformer (*cis*-Pro) (pdb 3sxx).

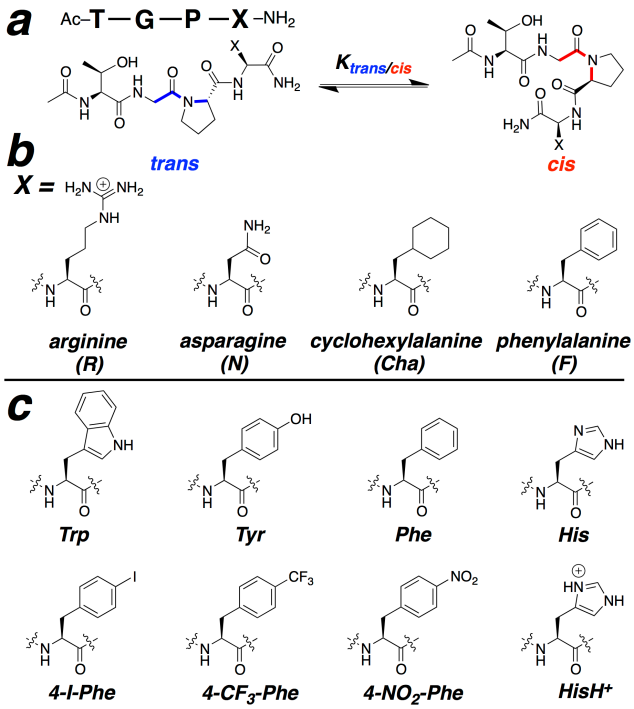

Figure 2. Peptides used in this study. (a) To investigate the roles of hydrophobicity and aromaticity of the residue following a proline residue on proline *cis-trans* isomerization, a series of tetrapeptides was synthesized with the general sequence Ac-Thr-Gly-Pro-Xaa-NH_2 (Ac-TGPX-NH_2, TGPX), where Xaa or X indicates the amino acid at the modification site. (b) The post-proline residue was systematically examined with a charged polar residue (Arg), a polar neutral residue (Asn), a non-aromatic hydrophobic residue (Cha) that mimics Phe, and an aromatic residue (Phe). (c) The effects of electron-rich and electron-deficient aromatic rings were investigated via using the canonical amino acids Trp, Tyr, Phe, His (neutral), and $HisH^+$ (cationic His), or the non-native amino acids 4-iodophenylalanine, 4-trifluoromethylphenylalanine, and 4-nitrophenylalanine.

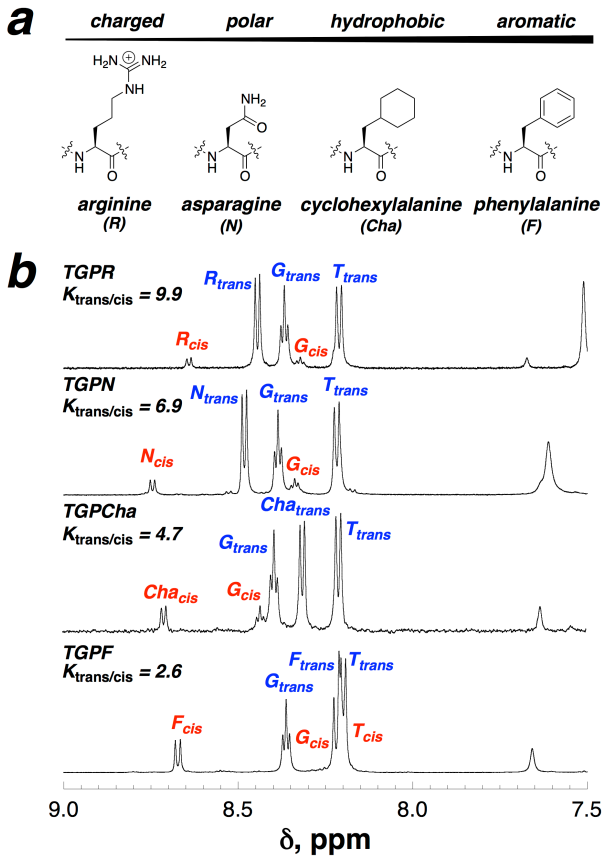

Figure 3. Effect of the residue C-terminal to proline on *cis-trans* isomerization in the model peptide sequence Ac-TGPX-NH_2 (TGPX). (a) Structures of residues X within the model peptide sequences Ac-TGPX-NH_2. The C-terminal residue was examined as the charged residue Arg (R) and the neutral, hydrophobic, aromatic residue Phe (F). Asparagine (N) is favored in type VI β-turns as the residue after proline, and increases the *cis*-Pro population in the *i*+3 position of β-turns. $K_{trans/cis}$ = the ratio of the population with a *trans*-Pro amide to the population with a *cis*-Pro. Thus, a lower $K_{trans/cis}$ value indicates a higher population of *cis*-Pro. (b) Amide region of the ^{1}H NMR spectra of the tetrapetides TGPR, TGPN, TGPCha, and TGPF. Resolved amide resonances for the residues with the assigned conformer, *trans* (blue) or *cis* (red), are annotated with their one-letter codes, except for cyclohexylalanine (3-letter code: Cha). The spectra were recorded at 298 K with peptide in 90% H_2O/10% D_2O with 5 mM sodium phosphate pH 4.0 and 25 mM NaCl.

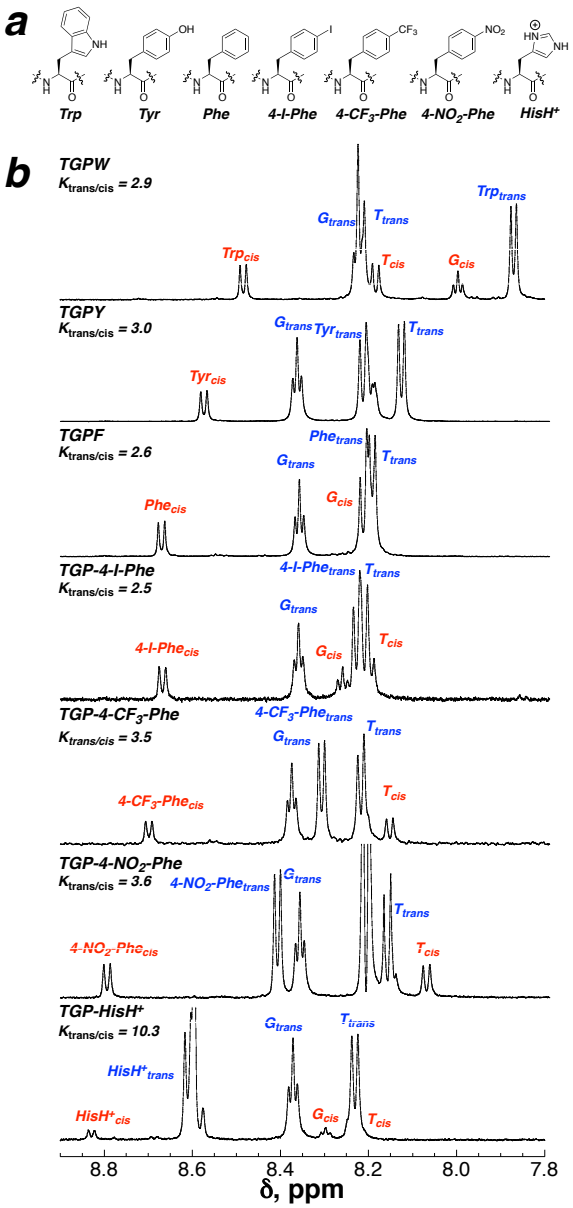

Figure 4. Effect of the identity of the aromatic residue C-terminal to proline on *cis-trans* isomerization. (a) Aromatic residues Trp (W), Tyr (Y), Phe (F), 4-iodophenylalanine (4-I-Phe), 4-trifluromethylphenylalanine (4-CF_3-Phe), 4-nitrophenylalanine (4-NO_2-Phe), and cationic histidine ($HisH^+$), arranged by electron richness of the face, were examined as residue X in model peptides with the sequence Ac-TGPX-NH_2 (TGPX). (b) Amide region of the ^1H NMR spectra of the peptides TGPX. Resolved amide resonances for the residues, with the assigned *trans* (blue) and *cis* (red) conformers indicated, are annotated with either their one-letter code or with their three-letter code (for X in TGPX) for the relevant substitution. The spectra were recorded at 298 K with peptide in 90% H_2O/10% D_2O with 5 mM sodium phosphate pH 4.0 and 25 mM NaCl. TGPHis data (neutral His) are not included due to rapid exchange of the amide hydrogens at pH 8.5.

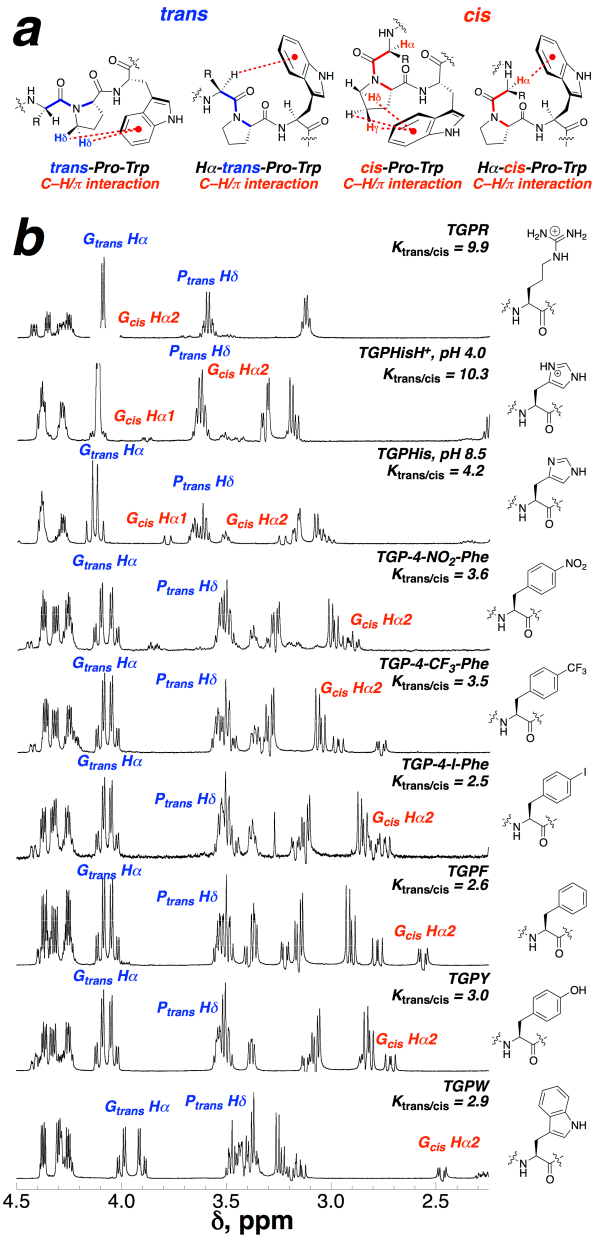

Figure 5. Competing C–H/□ interactions in the *trans*-Pro and *cis*-Pro conformations in the Ac-TGPX-NH_2 peptide context modulate the relative populations of the amide rotameric states in the *cis-trans* isomerization equilibrium. (a) In *cis*-Pro and *trans*-Pro conformations, two general types of C–H/□ interactions are possible, in which either or both (1) the side-chain hydrogen atoms of proline ring and/or (2) H□ of the residue preceding Pro interact with the aromatic ring and act as the □ electron acceptor. Hydrogens interacting with the □ electrons are shifted upfield due to ring-current effects. (b) Aliphatic region of the ^{1}H NMR spectra of the peptides Ac-TGPX-NH_2. The Gly H□1 and H□2 (red) resonances are shifted upfield in the *cis*-Pro conformation due to the ring-current effect from the aromatic residue X, consistent with an H□-*cis*-Pro-aromatic C–H/□ interaction. Two main types of C–H/□ interactions may be observed here in the *trans*-Pro conformation: both (1) a *trans*-Pro-H□-aromatic interaction (upfield shift, resonances annotated in blue) and/or (2) an H□-*trans*-Pro-aromatic interaction (significant in TGPW, resonances annotated in blue). Notably, the Gly H□1 and Gly H□2 resonances in the *cis*-Pro conformation are more upfield when the histidine side chain is neutral, consistent with a stronger C–H/□ interaction with neutral His than with cationic $HisH^+$. The spectra were obtained in 90% H_2O/10% D_2O with 5 mM sodium phosphate pH 8.5 (neutral His) or 4.0 (all other peptides) and 25 mM NaCl at 298 K.

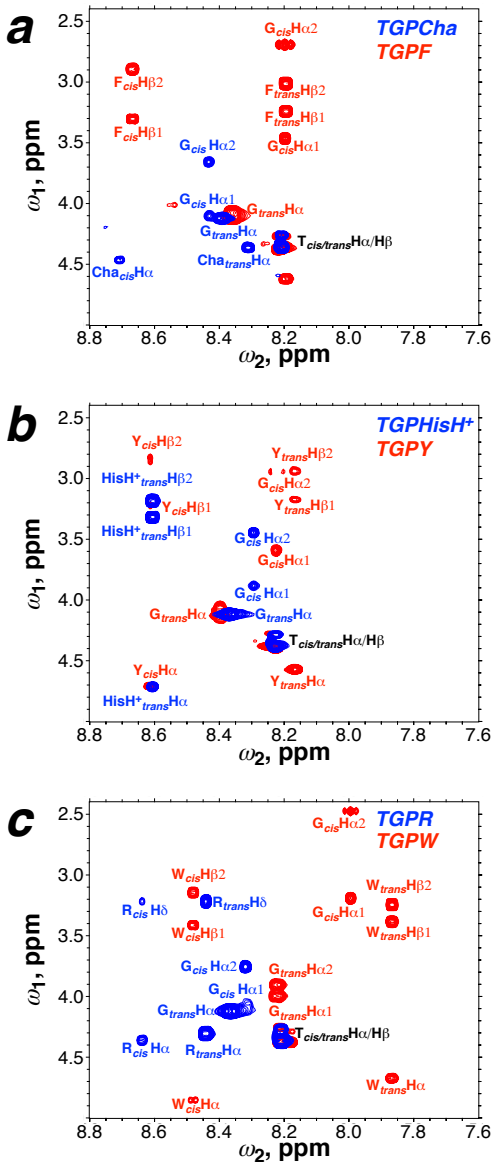

Figure 6. Superposition of the TOCSY spectra (fingerprint region) of pairs of comparator peptides (TGPX). (a) TGPCha (blue) and TGPF (red), X = hydrophobic non-aromatic and aromatic, respectively. (b) TGPHisH$^+$ (blue) and TGPY (red), X = electron-poor and electron-rich aromatic groups, respectively. (c) TGPR (blue) and TGPW (red), X = positively charged polar non-aromatic and the strongest electron-donating natural amino acid, respectively. The spectra were recorded in 90% H_2O/10% D_2O with 5 mM sodium phosphate pH 4.0 and 25 mM NaCl at 298 K.

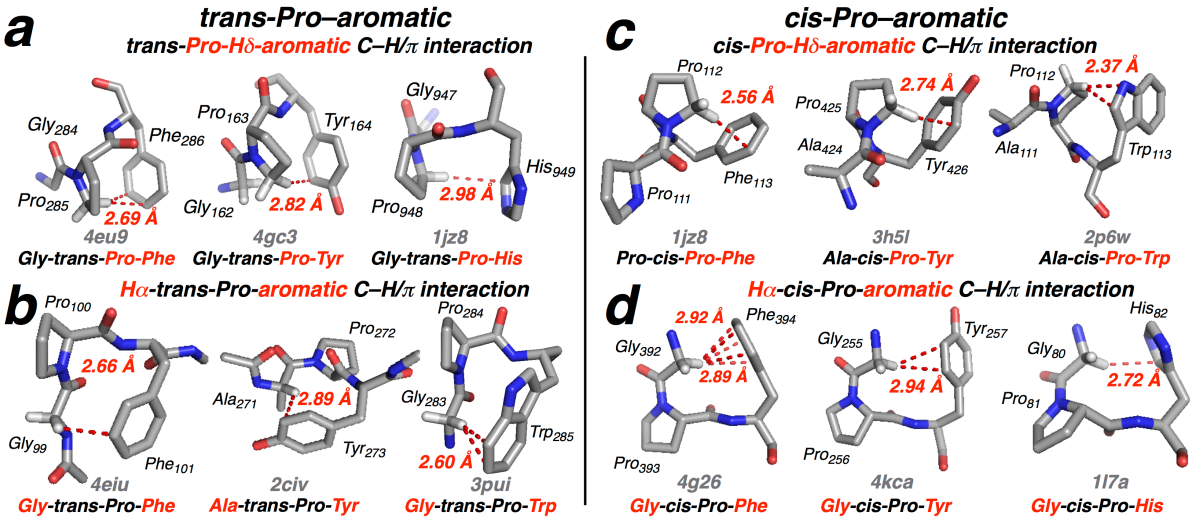

Figure 7. Local interactions observed in the (a) *trans*-Pro and (b) *cis*-Pro conformations in proline-aromatic sequences within the PDB. Distances (Å) are the distance between the indicated hydrogen atom and the nearest carbon atom(s) of the interacting aromatic ring. The hydrogen atoms were added in Pymol, and the positions of the hydrogens were subsequently optimized, while keeping the positions of the heavy atoms fixed, by DFT calculations using the M06-HF functional and the Def2TZVP basis set in implicit water. Interactions at Pro Hβ and Hγ are also observed in some structures; see later figures and the Supporting Information for details.

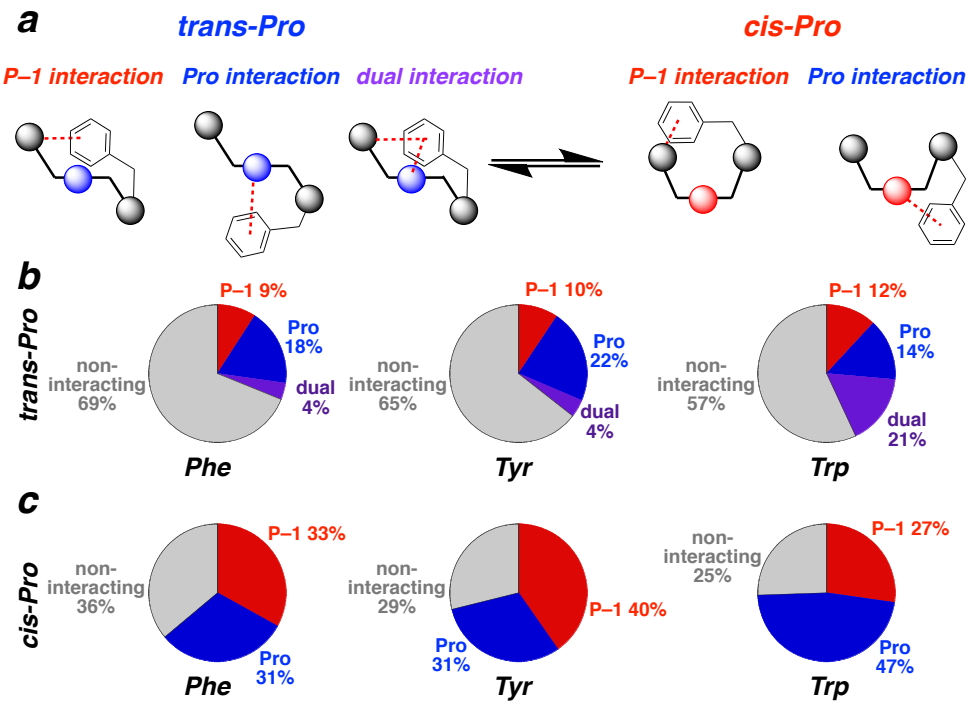

Figure 8. Modes of aromatic interactions in proline-aromatic sequences and their frequencies of occurrence in the PDB. (a) Schematic representation of interactions of the aromatic ring with the proline ring (blue [*trans*] or red [*cis*]) and/or with the residue before proline (pre-proline, P–1 residue, grey) for *trans*-Pro and *cis*-Pro conformations. (b, c) Normalized frequencies (in %) of (b) *trans*-Pro and (c) *cis*-Pro conformations involved in C–H/π interactions with hydrogen atoms exclusively from pre-proline (P–1, red) or from proline (Pro, blue), or with interactions with both pre-proline and proline (dual interactions, purple). Aromatic rings that do not exhibit a close interaction with either pre-proline or proline are indicated in grey.

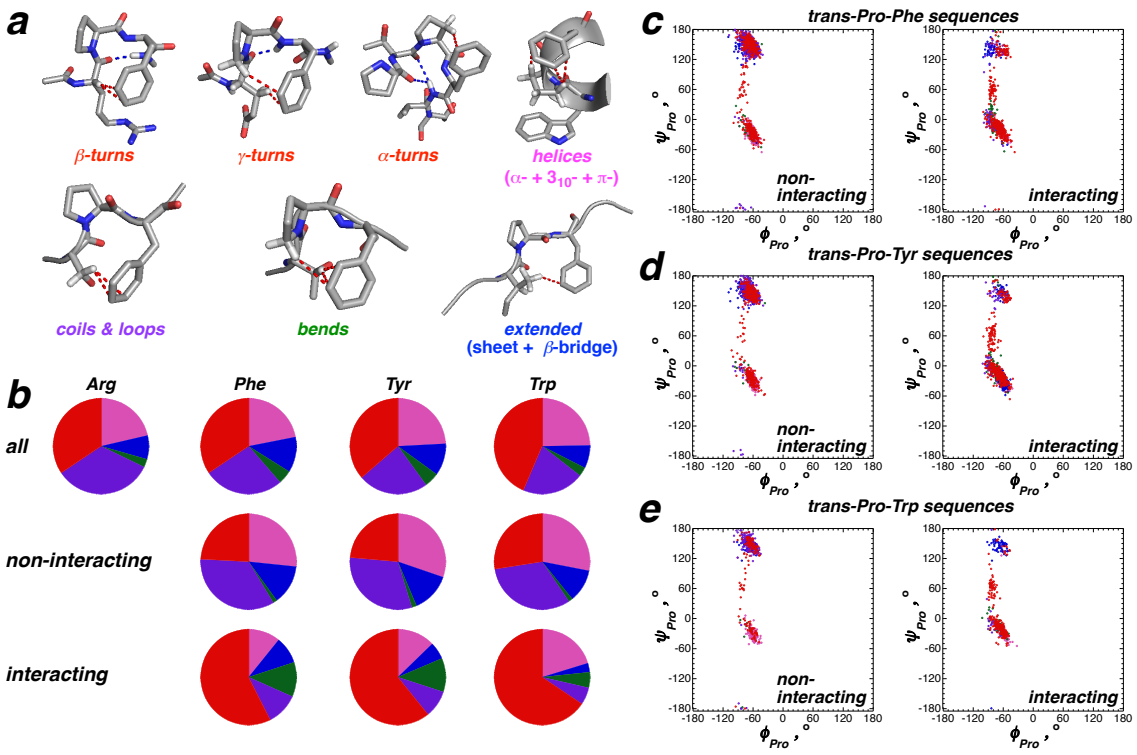

Figure 9. Secondary structures and the statistical distribution of the interaction modes within the secondary structures for *trans*-proline-aromatic sequences within the PDB, compared to the non-interacting residue Arg (*trans*-proline-arginine sequences) as a non-interacting reference state. (a) Classification of protein secondary structures was conducted and their nomenclature was assigned via PROMOTIF.[38] (b) Distribution of the assigned secondary structures in the PDB where P–1, Pro, and/or dual interactions are encountered are indicated: turns (□-, □-, and □+), red; helices (□- and 3_{10}-), magenta; bends, green; extended (□-sheets and bridges), blue; and coils and loops, purple. (c, d, e) Ramachandran plots of the proline residues in (left) non-interacting and (right) interacting structures in the PDB for (c) *trans*-Pro-Phe, (d) *trans*-Pro-Tyr, and (e) *trans*-Pro-Trp structures.

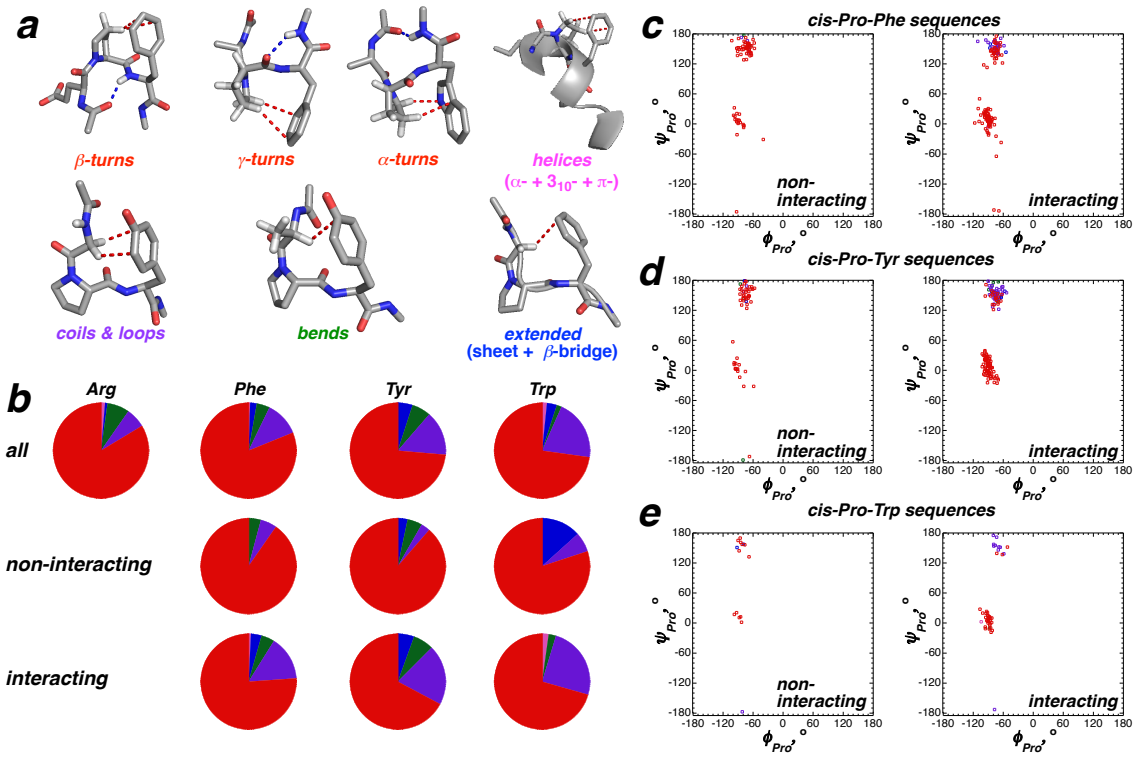

Figure 10. Secondary structures and the statistical distribution of the interaction modes within the secondary structures for *cis*-proline-aromatic sequences and *cis*-proline-arginine sequences (non-interacting reference state) within the PDB. (a) Classification of protein secondary structures was conducted and their nomenclature was assigned via PROMOTIF. (b) Distribution of the assigned secondary structures in the PDB where P–1, Pro, and/or dual interactions are encountered are indicated: turns (□-, □-, and □+), red; helices (□-, and 3_{10}-), magenta; bends, green; extended (□-sheets and bridges), blue; and coils and loops, purple. (c, d, e) Ramachandran plots of the proline residues in (left) non-interacting and (right) interacting structures in the PDB for (c) *cis*-Pro-Phe, (d) *cis*-Pro-Tyr, and (e) *cis*-Pro-Trp structures.

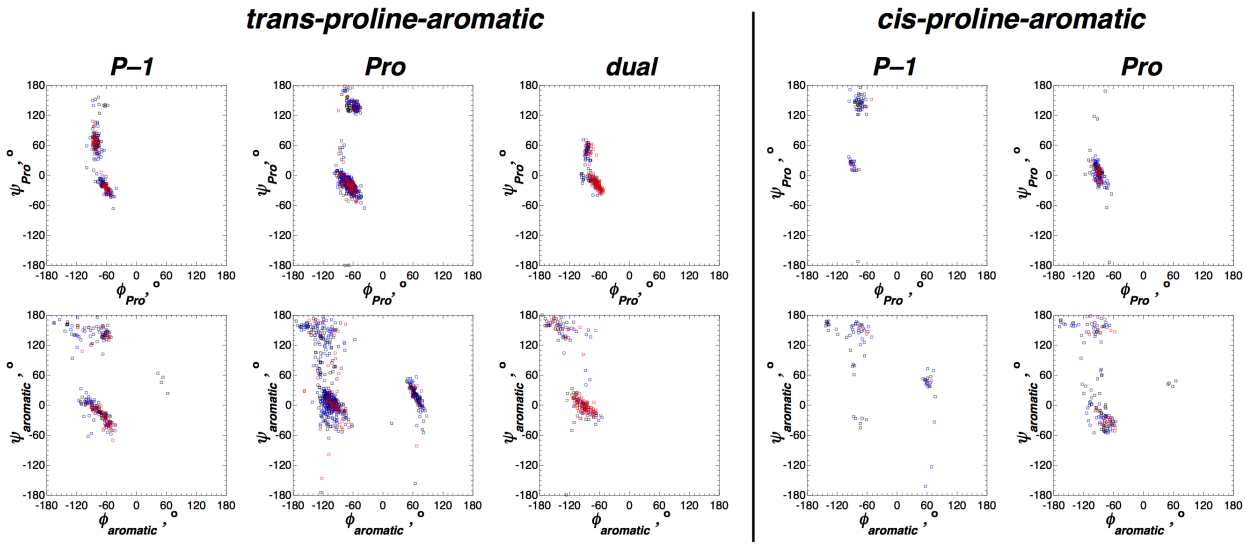

Figure 11. Composite Ramachandran plots for (top) proline and (bottom) aromatic residues (Phe: black, Tyr: blue, Trp: red) of (left) *trans*-proline-aromatic and (right) *cis*-proline aromatic sequences in turns, as a function of interaction of the aromatic ring with the P–1 residue, with Pro, or with both the P–1 and Pro residues (dual). Ramachandran plots for the individual amino acids are in the Supporting Information.

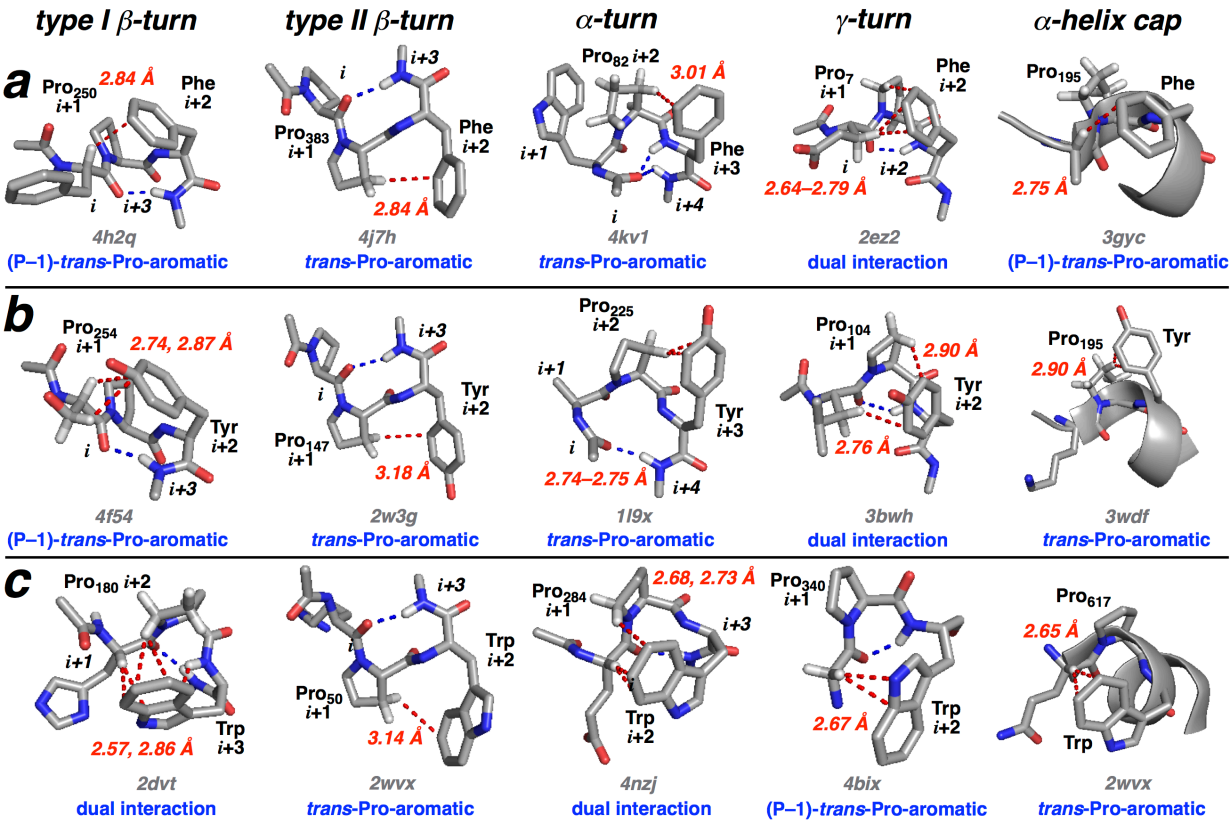

Figure 12. Suite of secondary structures stabilized by (P–1)-*trans*-Pro-aromatic [H□-*trans*-Pro-aromatic], *trans*-Pro-aromatic, and dual P–1 and Pro C–H/□ interactions with (a) Phe, (b) Tyr, and (c) Trp as the □-electron donor. C–H/□ interactions are indicated with red dashed lines, with the shortest H•••C distances indicated. For the turn structures, canonical *i* to *i*+3 hydrogen bonds are indicated by dashed blue lines. The residue numbers corresponding to the positions of Pro in the proteins are indicated. Hydrogen atoms were added in Pymol, and the positions of the hydrogens were subsequently optimized, while keeping the positions of the heavy atoms fixed, by DFT calculations using the M06-HF functional and the Def2TZVP basis set in implicit water. See Figures S30, S32, S35-S42, and S46-S51 for additional examples of modes of interactions and the correlation of interaction mode with secondary structure in *trans*-Pro.

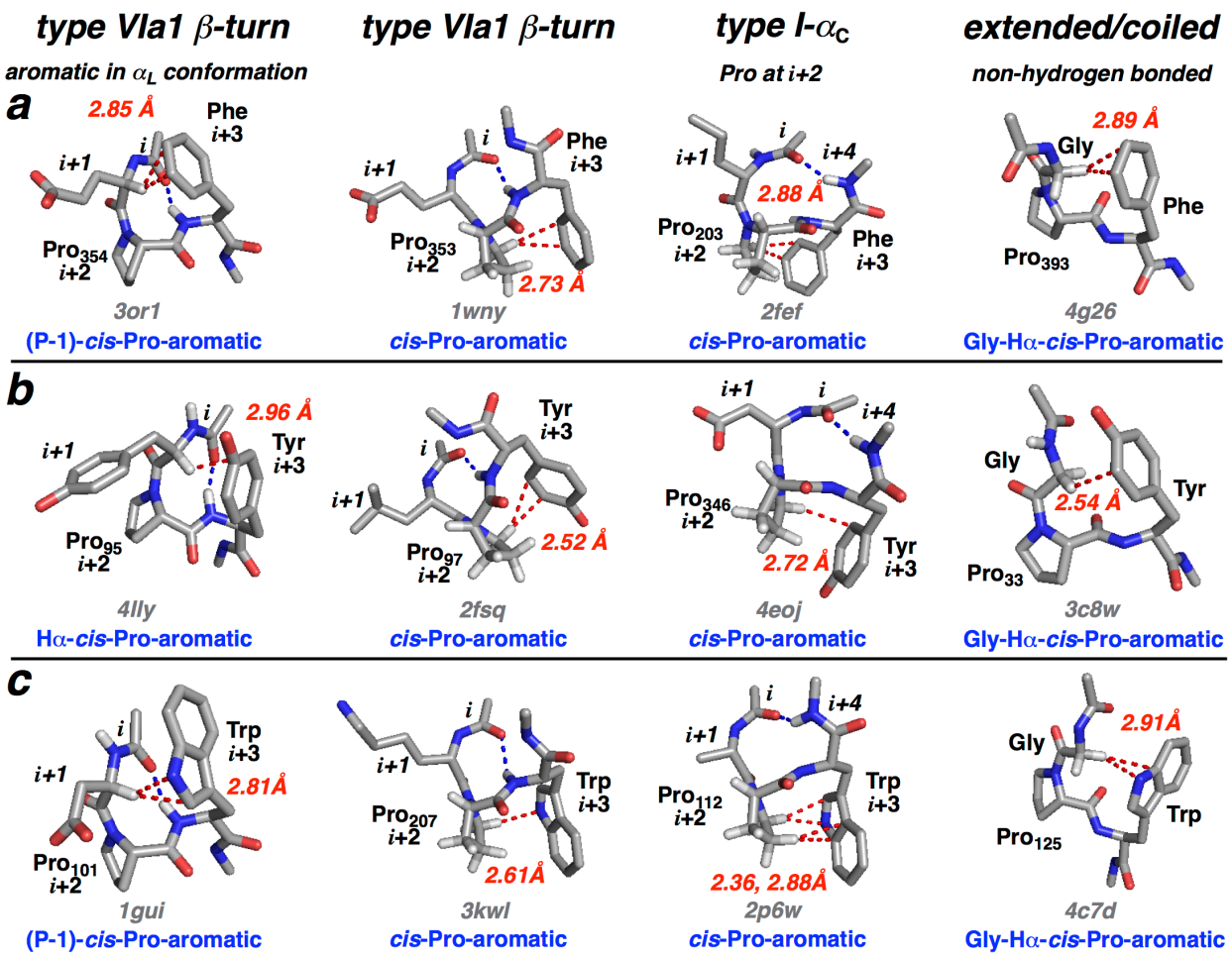

Figure 13. Suite of secondary structures stabilized by (P–1)-c*is*-Pro-aromatic [H-*cis*-Pro-aromatic] and *cis*-Pro-aromatic C–H/ interactions with (a) Phe, (b) Tyr, and (c) Trp as the -electron donor. C–H/ interactions are indicated by dashed red lines with the shortest distances between the H and C atoms indicated. For turn structures, canonical *i* to *i*+3 hydrogen bonds are indicated by dashed blue lines. The residue numbers correspond to the position of Pro residues in the proteins. The hydrogen atoms were added in Pymol, and the positions of the hydrogens were subsequently optimized, while keeping the positions of the heavy atoms fixed, by DFT calculations using the M06-HF functional and the Def2TZVP basis set in implicit water. See Figures S31, S32, S55, S60, S62, and S64 for additional examples of modes of interactions and the correlation of interaction mode with secondary structure in *cis*-Pro.

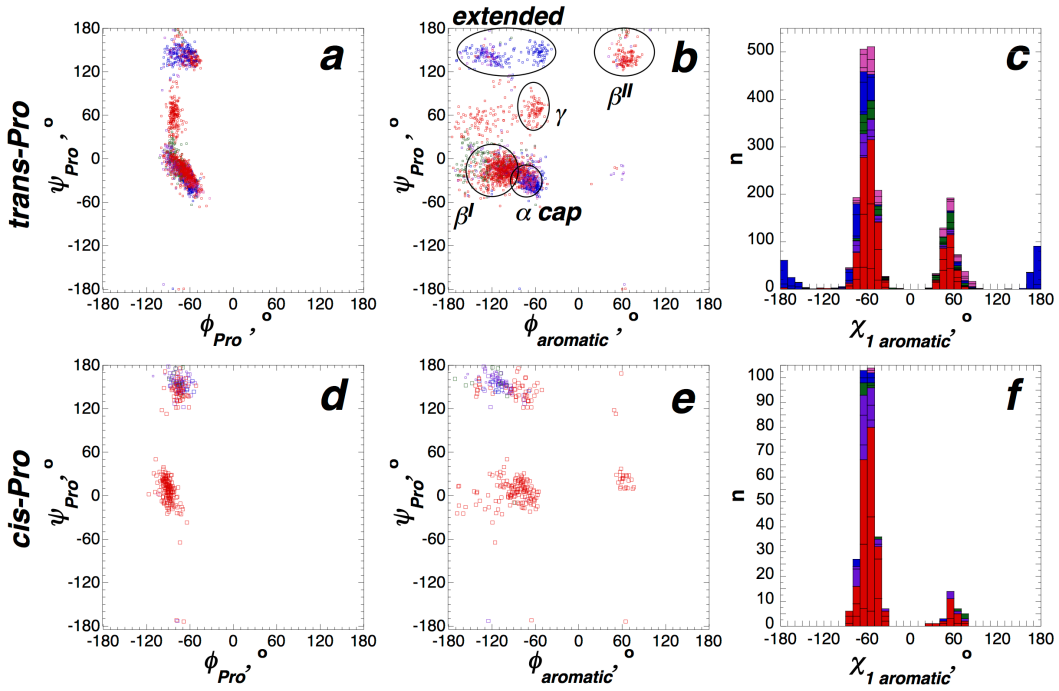

Figure 14. Conformational preferences of proline and aromatic residues in (a-c) *trans*-Pro-aromatic and (d-f) *cis*-Pro-aromatic sequences in the PDB with C–H/π interactions. (a) Ramachandran plot for *trans*-Pro. (b) Cross-correlation plot of the dihedral angle ψ from *trans*-proline and the dihedral angle ϕ from the aromatic residue in *trans*-Pro-aromatic sequences. (c) Distribution of χ_1 for the aromatic residue in *trans*-Pro-aromatic sequences. (d) Ramachandran plot for *cis*-Pro. (e) Cross-correlation plot of the dihedral angle ψ from *cis*-proline and the dihedral angle ϕ from the aromatic residue in *cis*-Pro-aromatic sequences. (f) Distribution of χ_1 for the aromatic residue in *cis*-Pro-aromatic sequences. The secondary structures corresponding to proline-aromatic sequences were classified as helix (magenta), extended (blue), bends (green), coils and loops (purple), and turns (red).

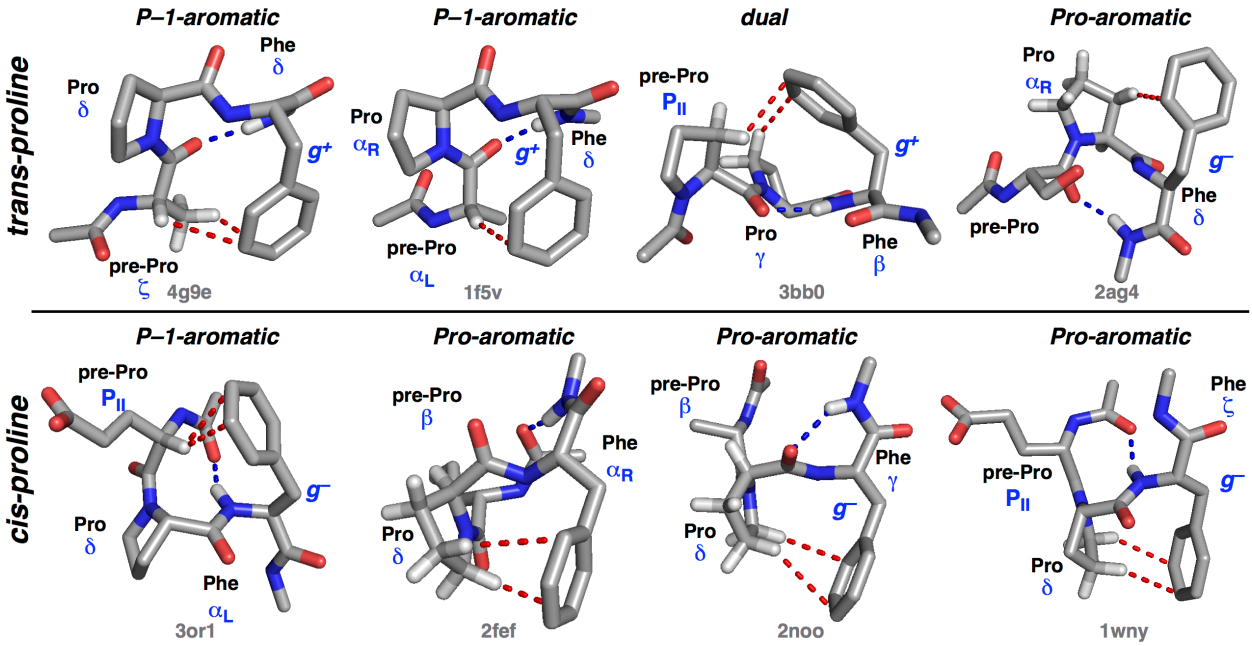

Figure 15. Conformational poses of Pro-aromatic sequences that adopt turns which are stabilized via C–H/□ interactions, for (top) *trans*-proline and (bottom) *cis*-proline. C–H/□ interactions are indicated by red lines, while hydrogen bonds of turns are indicated by blue lines.

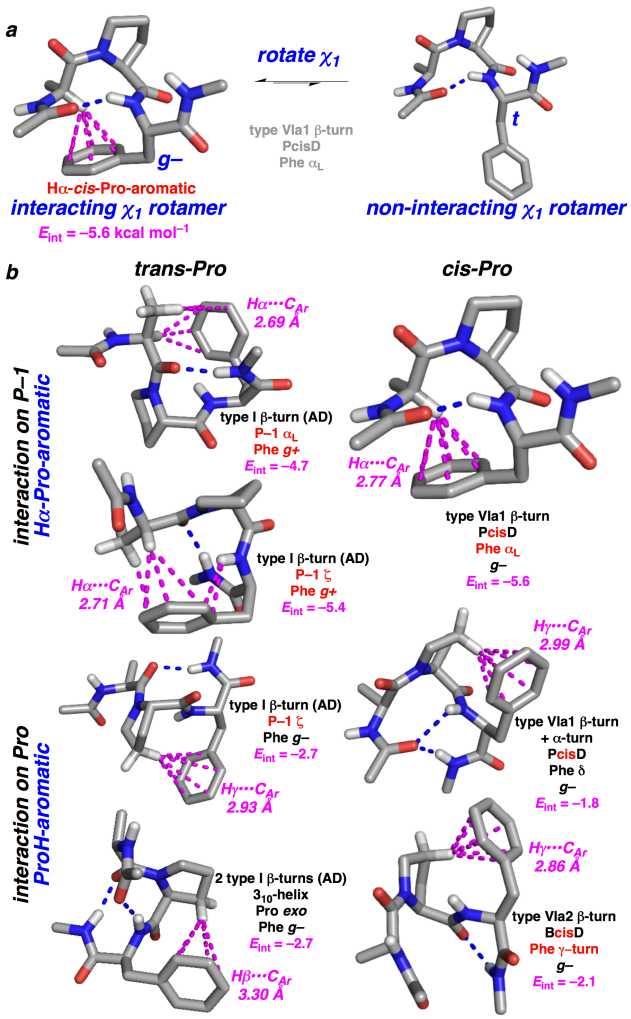

Figure 16. Computational investigation of preferred geometries of aromatic interactions in Ac-Ala-Pro-Phe-NHMe, as a function of proline amide rotamer, interaction at Pro versus at P–1, and turn geometry. C–H/π interactions are indicated in magenta, with the closest H•••$C_{aromatic}$ distance indicated. Hydrogen bonds of β-turns are indicated in blue (a) Geometry optimization calculations were conducted on truncated forms of turn structures identified in bioinformatics investigations. For each interaction pose, an alternative structure was generated in which Phe χ_1 was changed to the non-interacting *t* rotamer. After full geometry optimization on both structures (M06-HF/Def2TZVP/H_2O), the interaction geometries were analyzed. To determine the approximate interaction electronic energy for each pose, the energies of the conformations with the interacting and non-interacting rotamers of the same overall conformation were compared, $E_{int} = E_{\text{interacting rotamer}} - E_{\text{non-interacting rotamer}}$. (b) Structures of all interacting poses analyzed, as a function of site of interaction, proline amide conformation, and overall observed conformation. Interaction energies (E_{int}) (kcal mol^{-1}) are shown for each. Conformations that represent inherently higher-energy conformations are indicated in red. β-Turn types are shown using classical β-turn nomenclature and Dunbrack's β-turn nomenclature.[41,42] The geometry-optimized structures in Figure 16b were developed from original models from structures in (clockwise from upper left) pdb 1f5v, 4je5, 1wny, 2noo, 4iiy, 4g9e with χ_1 rotation to g^-, and 4g9e), with truncation of the pre-proline residue to Ala if necessary.

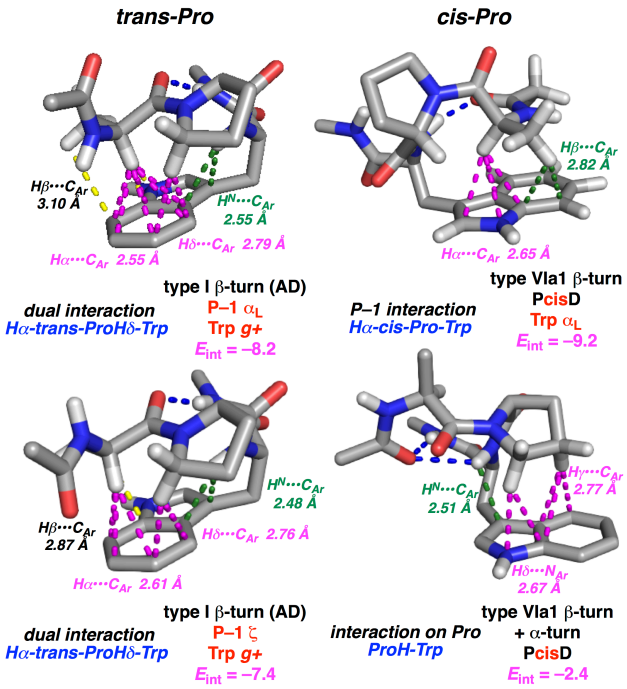

Figure 17. Computational determination of key Pro-Trp structures with C–H/π interactions, as a function of interaction site and proline amide conformation. Closest H•••C distances are indicated at each hydrogen atom type. Hydrogen bonds of β-turns are indicated in blue. Disfavored conformations are indicated in red. Additional aromatic interactions with H•••C distances below 2.90 Å are indicated in green. Interaction electronic energies (E_{int}, kcal mol^{-1}) were determined as is described in Figure 16, comparing the relative energies of the indicated structure with that of the analogous structure with the aromatic residue with the non-interacting *t* χ_1 rotamer. These geometry-optimized structures were developed from original models from structures in (clockwise from upper left) 4nzj, 1gui, 2p6w, and 2dvt, with truncation of the pre-proline residue to Ala if necessary.

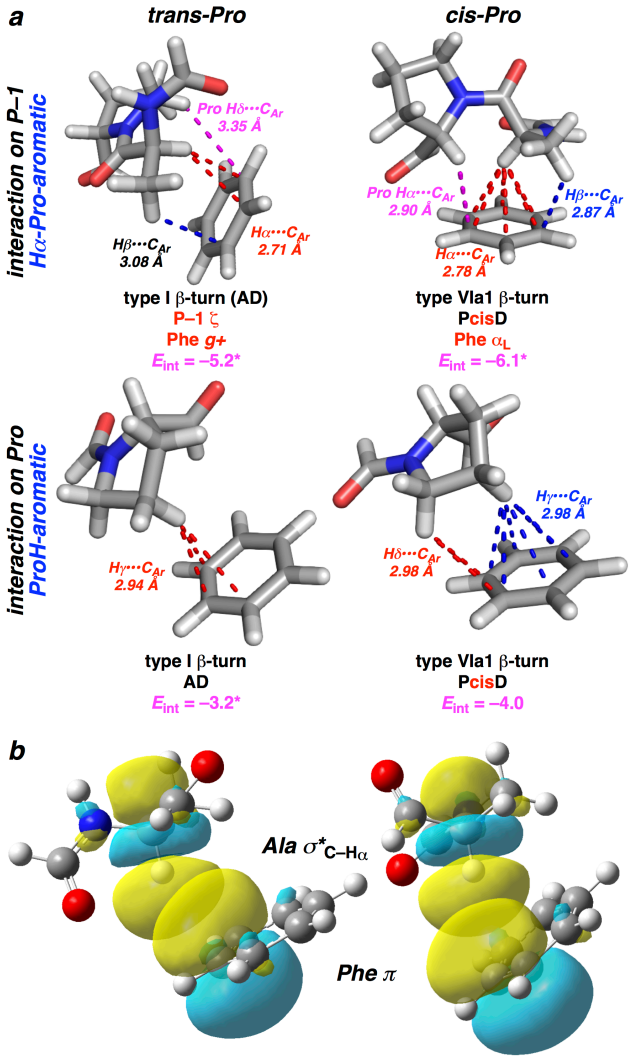

Figure 18. Minimalist bimolecular models of interaction modes in proline-aromatic sequences. (a) Interaction electronic energies (E_{int}, kcal mol^{-1}) and closest H•••C distances in each structure are indicated. E_{int} were determined using the MP2 method and the Def2QZVP basis set in implicit water, with the E_{int} corrected using the gas-phase basis-set superposition error (BSSE). These structures are derived from the equivalent structures in Figure 16 (original models from [clockwise from upper left] pdb 4g9e, 4je5, 1wny, and 4g9e with χ_1 rotation to g^-), with the conformationally unfavorable elements in the parent protein structures indicated in red text. All structures were truncated to formyl N-termini and aldehyde C-termini to simplify the calculations and to minimize steric clashes that result from the generation of a bimolecular structure. Interaction energies indicated with * have an additional interaction with the N-terminal formyl hydrogen that would not be present with a protein. In addition, these electronic energies do not include contributions from the hydrophobic effect or from unfavorable conformational energy. (b) Natural bond orbital (NBO) analysis (M06-HF/Def2QZVP/H_2O) of versions of the structures in (a) with P–1 interactions, truncated to For-Ala-H•benzene for clarity. Orbital overlap between the localized molecular orbitals shown ($\pi_{Aromatic}$ $\sigma^*_{C–H}$) results in through-space electron delocalization that stabilizes the C–H/π interaction.

Tables

Table 1. Effects of the residue C-terminal to proline on the *cis-trans* isomerization equilibrium in the peptide context Ac-TGPX-NH_2.[a]

X	$K_{trans/cis}$	ΔG[b]	$\Delta\Delta G$[c]
Arg	9.9	–1.36	0.00
Asn	6.9	–1.14	–0.22
Cha	4.7	–0.92	–0.44
4-NO_2-Phe	3.6	–0.76	–0.60
4-CF_3-Phe	3.5	–0.74	–0.62
4-I-Phe	2.5	–0.54	–0.82
Phe	2.6	–0.57	–0.79
Tyr	3.0	–0.65	–0.71
Trp	2.9	–0.63	–0.73
$HisH^+$	10.3	–1.38	+0.02
His[d]	4.2	–0.85	–0.51

[a] X = residue C-terminal to the proline. ^{1}H NMR spectra were acquired using solutions with 5 mM phosphate buffer pH 4.0 and 25 mM NaCl in 90%H_2O/10%D_2O at 298 K, unless indicated otherwise.

[b] $\Delta G = -RT \ln K_{trans/cis}$, in kcal mol^{-1}.

[c] $\Delta\Delta G = \Delta G_{TGPR} - \Delta G_{TGPX}$, the difference in the free energies of the *cis-trans* isomerization equilibria due to the substitution of arginine with the indicate amino acid at residue X in the peptides Ac-TGPX-NH_2, in kcal mol^{-1}.

[d] pH 8.0, 298 K.

Table 2. Effects of the residue C-terminal to proline on the *cis-trans* isomerization equilibrium and on the Ala Hα chemical shift in the peptide context Ac-TAPX-NH_2.[a]

peptide	$K_{trans/cis}$	ΔG[b]	$\Delta\Delta G$[c]	*trans*-Pro	*cis*-Pro
				AlaHα	AlaHα
TAPN	10.3	–1.38	0.00	4.63	4.38
TAPF	5.3	–0.99	–0.39	4.60	3.85
TAPW	5.5	–1.01	–0.37	4.47	3.89

[a] X = the residue C-terminal to proline.
[b] ΔG = –RT ln $K_{trans/cis}$, at 298 K, in kcal mol^{-1}.
[c] $\Delta\Delta G = \Delta G_{TAPX} - \Delta G_{TAPN}$, in kcal mol^{-1}, the difference in free energy for the *cis-trans* isomerization equilibrium for the indicated residue compared to asparagine when C-terminal to proline in Ac-TAPX-NH_2.

Table 3. Summary of Gly Hα and Pro Hα chemical shifts[a] as a function of proline rotamer in Ac-TGPX-NH_2 peptides.[b]

X =	$\Delta\Delta G$[c]	*trans*-Pro				*cis*-Pro			
		δ_{Gly} Hα1,	Hα2	δ_{Pro} Hα1,	Hα2	δ_{Gly} Hα1,	Hα2	δ_{Pro} Hα1,	Hα2
Arg	0.00	4.12	4.12	3.65	3.65	4.07	3.75	3.58	3.58
Asn	–0.22	4.13	4.13	3.70	3.63	4.10	3.72	3.56	3.56
Cha	–0.44	4.12	4.12	3.66	3.66	4.10	3.66	3.58	3.58
4-NO_2-Phe	–0.60	4.09	4.09	3.60	3.57	3.54	2.87	3.45	3.45
4-CF_3-Phe	–0.62	4.13	4.08	3.58	3.58	3.58	3.00	3.46	3.46
4-I-Phe	–0.82	4.09	4.09	3.58	3.58	3.53	2.92	3.66	3.66
Phe	–0.79	4.08	4.08	3.58	3.58	3.46	2.70	3.46	3.46
Tyr	–0.71	4.01	4.01	3.58	3.58	3.59	2.94	3.44	3.44
Trp	–0.73	4.00	3.90	3.45	3.45	3.19	2.48	3.36	3.36
$HisH^+$	+0.02	4.12	4.12	3.63	3.63	3.88	3.45	3.52	3.52
His[d]	–0.51	4.16	4.10	3.65	3.61	3.79	3.24	3.51	3.51

[a] Chemical shifts are indicated in ppm.
[b] Hα1 = downfield Gly Hα, Hα2 = upfield Gly Hα; Hα1 = downfield Pro Hα, Hα2 = upfield Pro Hα.
[c] $\Delta\Delta G = \Delta G_{TGPX} - \Delta G_{TGPR}$, in kcal mol^{-1}, the relative difference in the free energy of the *cis-trans* isomerization equilibrium for the substitution of arginine residue with the indicated amino acid at residue X in the peptide context Ac-TGPX-NH_2. See Table 1 for $K_{trans/cis}$ and ΔG for each peptide proline *cis-trans* isomerization equilibrium.
[d] pH 8.0, 298 K.

Table 4. Frequencies of *trans*-proline and *cis*-proline conformations in the PDB, with aromatic residues (Phe, Tyr, Trp, or His) or arginine C-terminal to proline.

sequence	$N_{trans\text{-}Pro}$[a]	$N_{cis\text{-}Pro}$[a]	total[b]	% *cis*[c]
Pro-Phe	2328	204	2532	8.1
Pro-Tyr	2014	205	2219	9.2
Pro-Trp	861	61	922	6.6
Pro-His	1242	79	1321	6.0
Pro-Arg	2537	127	2664	4.8

[a] $N_{trans\text{-}Pro}$ = number of *trans*-proline residues. $N_{cis\text{-}Pro}$ = number of *cis*-proline residues.
[b] total = $N_{trans\text{-}Pro} + N_{cis\text{-}Pro}$, the total number of proline residues in the individual dataset.
[c] % *cis* = $N_{cis\text{-}Pro}/(N_{trans\text{-}Pro} + N_{cis\text{-}Pro}) \times 100$; frequency of occurrence of a *cis*-Pro residue with a Pro-Arg or with Pro-aromatic sequences in protein structures.

References

(1) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. Folding of Immunogenic Peptide-Fragments of Proteins in Water Solution .1. Sequence Requirements for the Formation of a Reverse Turn. *J. Mol. Biol.* **1988**, *201*, 161-200.

(2) Yao, J.; Dyson, H. J.; Wright, P. E. 3-Dimensional Structure of a Type-VI turn in a Linear Peptide in Water Solution - Evidence for Stacking of Aromatic Rings as a Major Stabilizing Factor. *J. Mol. Biol.* **1994**, *243*, 754-766.

(3) Nardi, F.; Kemmink, J.; Sattler, M.; Wade, R. C. The cisproline(i-1)-aromatic(i) interaction: Folding of the Ala-cisPro-Tyr peptide characterized by NMR and theoretical approaches. *J. Biomol. NMR* **2000**, *17*, 63-77.

(4) Pal, D.; Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations. *J. Mol. Biol.* **1999**, *294*, 271-288.

(5) MacArthur, M. W.; Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **1991**, *218*, 397-412.

(6) Kemmink, J.; Creighton, T. E. The Physical Properties of Local Interactions of Tyrosine Residues in Peptides and Unfolded Proteins. *J. Mol. Biol.* **1995**, *245*, 251-260.

(7) Houry, W. A.; Scheraga, H. A. Nature of the unfolded state of ribonuclease A: Effect of cis-trans X-pro peptide bond isomerization. *Biochemistry* **1996**, *35*, 11719-11733.

(8) Bhattacharyya, R.; Chakrabarti, P. Stereospecific Interactions of Proline Residues in Protein Structures and Complexes. *J. Mol. Biol.* **2003**, *331*, 925-940.

(9) Meng, H. Y.; Thomas, K. M.; Lee, A. E.; Zondlo, N. J. Effects of *i* and *i*+3 Residue Identity on Cis-Trans Isomerism of the Aromatic$_{i+1}$-Prolyl$_{i+2}$ Amide Bond: Implications for Type VI b-turn Formation. *Biopolymers (Peptide Sci.)* **2006**, *84*, 192-204.

(10) Brown, A. M.; Zondlo, N. J. A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline-Aromatic Interactions. *Biochemistry* **2012**, *51*, 5041-5051.

(11) Zondlo, N. J. Aromatic-Proline Interactions: Electronically Tunable CH/pi Interactions. *Acc. Chem. Res.* **2013**, *46*, 1039-1049.

(12) Ganguly, H. K.; Majumder, B.; Chattopadhyay, S.; Chakrabarti, P.; Basu, G. Direct Evidence for CH...pi Interaction Mediated Stabilization of Pro-cisPro Bond in Peptides with Pro-Pro-Aromatic Motifs. *J. Am. Chem. Soc.* **2012**, *134*, 4661-4669.

(13) Ganguly, H. K.; Kaur, H.; Basu, G. Local Control of cis-Peptidyl−Prolyl Bonds Mediated by CH···π Interactions: The Xaa-Pro-Tyr Motif. *Biochemistry* **2013**, *52*, 6348-6357.

(14) Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P. CH-pi hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12648-12683.

(15) Wang, Y.; Zhuang, Y.; DiBerto, J. F.; Zhou, X. E.; Schmitz, G. P.; Yuan, Q.; Jain, M. K.; Liu, W.; Melcher, K.; Jiang, Y.; Roth, B. L.; Xu, H. E. Structures of the entire human opioid receptor family. *Cell* **2023**, *186*, 413-427.

(16) Podlogar, B. L.; Paterlini, M. G.; Ferguson, D. M.; Leo, G. C.; Demeter, D. A.; Brown, F. K.; Reitz, A. B. Conformational analysis of the endogenous mu-opioid agonist endomorphin-1 using NMR spectroscopy and molecular modeling. *FEBS Lett.* **1998**, *439*, 13-20.

(17) Fiori, S.; Renner, C.; Cramer, J.; Pegoraro, S.; Moroder, L. Preferred conformation of endomorphin-1 in aqueous and membrane-mimetic environments. *J. Mol. Biol.* **1999**, *291*, 163-175.

(18) Keller, M.; Boissard, C.; Patiny, L.; Chung, N. N.; Lemieux, C.; Mutter, M.; Schiller, P. W. Pseudoproline-containing analogues of morphiceptin and endomorphin-2: Evidence for a cis Tyr-Pro amide bond in the bioactive conformation. *J. Med. Chem.* **2001**, *44*, 3896-3903.

(19) Schwartz, A. C.; Jay, D. W.; Parnham, S.; Giuliano, M. W. Sequential and Environmental Dependence of Conformation in a Small Opioid Peptide. *J. Org. Chem.* **2019**, *84*, 13299-13312.

(20) Kumar, P.; Reithofer, V.; Reisinger, M.; Wallner, S.; Pavkov-Keller, T.; Macheroux, P.; Gruber, K. Substrate complexes of human dipeptidyl peptidase III reveal the mechanism of inhibition. *Sci. Rep.* **2016**, *6*, 23787.

(21) Leitgeb, B.; Toth, G. Aromatic-aromatic and proline-aromatic interactions in endomorphin-1 and endomorphin-2. *Eur. J. Med. Chem.* **2005**, *40*, 674-686.

(22) Zdobinsky, T.; Scherkenbeck, J.; Zerbe, O.; Antonicek, H.; Chen, H. Structures of Micelle-Bound Selected Insect Neuropeptides and Analogues: Implications for Receptor Selection. *ChemBioChem* **2009**, *10*, 2644-2653.

(23) Leitgeb, B. Characteristic Structural Features of Indolicidin: Effects of the cis-trans Isomerism on its Conformation. *Chem. Biol. Drug Des.* **2014**, *83*, 132-140.

(24) Tyndall, J. D. A.; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2005**, *105*, 793-826.

(25) Hong, M.; Zhukareva, V.; Vogelsberg-Ragaglia, V.; Wszolek, Z.; Reed, L.; Miller, B. I.; Geschwind, D. H.; Bird, T. D.; McKeel, D.; Goate, A.; Morris, J. C.; Wilhelmsen, K. C.; Schellenberg, G. D.; Trojanowski, J. Q.; Lee, V. M. Y. Mutation-specific functional impairments in distinct Tau isoforms of hereditary FTDP-17. *Science* **1998**, *282*, 1914-1917.

(26) Strang, K. H.; Golde, T. E.; Giasson, B. I. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. *Lab. Investig.* **2019**, *99*, 912-928.

(27) de Calignon, A.; Polydoro, M.; Suarez-Calvet, M.; William, C.; Adamowicz, D. H.; Kopeikina, K. J.; Pitstick, R.; Sahara, N.; Ashe, K. H.; Carlson, G. A.; Spires-Jones, T. L.; Hyman, B. T. Propagation of Tau Pathology in a Model of Early Alzheimer's Disease. *Neuron* **2012**, *73*, 685-697.

(28) von Bergen, M.; Barghorn, S.; Li, L.; Marx, A.; Biernat, J.; Mandelkow, E. M.; Mandelkow, E. Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local beta-structure. *J. Biol. Chem.* **2001**, *276*, 48165-48174.

(29) Miyasaka, T.; Morishima-Kawashima, M.; Ravid, R.; Heutink, P.; van Swieten, J. C.; Nagashima, K.; Ihara, Y. Molecular analysis of mutant and wild-type tau deposited in the brain affected by the FTDP-17 R406W mutation. *Am. J. Pathol.* **2001**, *158*, 373-379.

(30) Tatebayashi, Y.; Miyasaka, T.; Chui, D. H.; Akagi, T.; Mishima, K.; Iwasaki, K.; Fujiwara, M.; Tanemura, K.; Murayama, M.; Ishiguro, K.; Planel, E.; Sato, S.; Hashikawa, T.; Takashima, A. Tau filament formation and associative memory deficit in aged mice expressing mutant (R406W) human tau. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13896-13901.

(31) Zhang, B.; Higuchi, M.; Yoshiyama, Y.; Ishihara, T.; Forman, M. S.; Martinez, D.; Joyce, S.; Trojanowski, J. Q.; Lee, V. M. Y. Retarded axonal transport of R406W mutant tau in transgenic mice with a neurodegenerative tauopathy. *J. Neurosci.* **2004**, *24*, 4657-4667.

(32) Perez, M.; Lim, F.; Arrasate, M.; Avila, J. The FTDP-17-linked mutation R406W abolishes the interaction of phosphorylated tau with microtubules. *J. Neurochem.* **2000**, *74*, 2583-2589.

(33) Ikeda, M.; Shoji, M.; Kawarai, T.; Kawarabayashi, T.; Matsubara, E.; Murakami, T.; Sasaki, A.; Tomidokoro, Y.; Ikarashi, Y.; Kuribara, H.; Ishiguro, K.; Hasegawa, M.; Yen, S. H.; Chishti, M. A.; Harigaya, Y.; Abe, K.; Okamoto, K.; George-Hyslop, P. S.; Westaway, D. Accumulation of filamentous tau in the cerebral cortex of human tau R406W transgenic mice. *Am. J. Pathol.* **2005**, *166*, 521-531.

(34) Wu, W.-J.; Raleigh, D. P. Local Control of Peptide Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions. *Biopolymers* **1998**, *45*, 381-394.

(35) Thomas, K. M.; Naduthambi, D.; Zondlo, N. J. Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* **2006**, *128*, 2216-2217.

(36) Pandey, A. K.; Thomas, K. M.; Forbes, C. R.; Zondlo, N. J. Tunable Control of Polyproline Helix (PPII) Structure via Aromatic Electronic Effects: An Electronic Switch of Polyproline Helix. *Biochemistry* **2014**, *53*, 5307-5314.

(37) Gibbs, E. B.; Lu, F. Y.; Portz, B.; Fisher, M. J.; Medellin, B. P.; Laremore, T. N.; Zhang, Y. J.; Gilmour, D. S.; Showalter, S. A. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nature Commun.* **2017**, *8*.

(38) Hutchinson, E. G.; Thornton, J. M. PROMOTIF - a program to identify and analyze sctructural motifs in proteins. *Protein Sci.* **1996**, *5*, 212-220.

(39) Yao, J.; Feher, V. A.; Espejo, B. F.; Reymond, M. T.; Wright, P. E.; Dyson, H. J. Stabilization of a Type-VI Turn in a Family of Linear Peptides in Water Solution. *J. Mol. Biol.* **1994**, *243*, 736-753.

(40) Dasgupta, R.; Ganguly, H. K.; Modugula, E. K.; Basu, G. Type VIa beta-turn-fused helix N-termini: A novel helix N-cap motif containing cis proline. *Peptide Sci.* **2017**, *108*, e22919.

(41) Wilmot, C. M.; Thornton, J. M. Analysis and Prediction of the Different Types of Beta-Turn in Proteins. *J. Mol. Biol.* **1988**, *203*, 221-232.

(42) Shapovalov, M.; Vucetic, S.; Dunbrack, R. L., Jr. A new clustering and nomenclature for beta turns derived from high-resolution protein structures. *PLoS Comput. Biol.* **2019**, *15*, e1006844.

(43) Dunbrack, R. L., Jr.; Karplus, M. Backbone-dependent Rotamer Library for Proteins: Application to Side-chain prediction. *J. Mol. Biol.* **1993**, *230*, 543-574.

(44) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40*, 389-408.

(45) Zhao, Y.; Truhlar, D. G. Density functional for spectroscopy: No long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states. *J. Phys. Chem. A* **2006**, *110*, 13126-13130.

(46) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations: The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.

(47) Tomasi, J.; Mennucci, B.; Cances, E. The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211-226.

(48) Biswal, H. S.; Wategaonkar, S. Sulfur, Not Too Far Behind O, N, and C: SH center dot center dot center dot pi Hydrogen Bond. *J. Phys. Chem. A* **2009**, *113*, 12774-12782.

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

(50) Schutkowski, M.; Bernhardt, A.; Zhou, X. Z.; Shen, M. H.; Reimer, U.; Rahfeld, J. U.; Lu, K. P.; Fischer, G. Role of phosphorylation in determining the backbone dynamics of the serine/threonine-proline motif and Pin1 substrate recognition. *Biochemistry* **1998**, *37*, 5566-5575.

(51) Montelione, G. T.; Arnold, E.; Meinwald, Y. C.; Stimson, E. R.; Denton, J. B.; Huang, S. G.; Clardy, J.; Scheraga, H. A. Chain-Folding Initiation Structures in Ribonuclease-a - Conformational-Analysis of Trans-Ac-Asn-Pro-Tyr-NHMe and Trans-Ac-Tyr-Pro-Asn-NHMe in Water and in the Solid-State. *J. Am. Chem. Soc.* **1984**, *106*, 7946-7958.

(52) Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and role of cis peptide-bonds in protein structures. *J. Mol. Biol.* **1990**, *214*, 253-260.

(53) Torbeev, V. Y.; Hilvert, D. Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate beta 2-microglobulin amyloid assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 20051-20056.

(54) Williamson, D. M.; Elferich, J.; Shinde, U. Mechanism of Fine-tuning pH Sensors in Proprotein Convertases IDENTIFICATION OF A pH-SENSING HISTIDINE PAIR IN THE PROPEPTIDE OF PROPROTEIN CONVERTASE 1/3. *J. Biol. Chem.* **2015**, *290*, 23214-23225.

(55) Gupta, S. K.; Banerjee, S.; Prabhakaran, E. N. Understanding the anomaly of cis-trans isomerism in Pro-His sequences. *Bioorg. Med. Chem. Lett.* **2022**, *76*, 128985.

(56) Forbes, C. R.; Sinha, S. K.; Ganguly, H. K.; Bai, S.; Yap, G. P. A.; Patel, S.; Zondlo, N. J. Insights into Thiol-Aromatic Interactions: A Stereoelectronic Basis for S–H/π Interactions. *J. Am. Chem. Soc.* **2017**, *139*, 1842-1855.

(57) Daniecki, N. J.; Bhatt, M. R.; Yap, G. P. A.; Zondlo, N. J. Proline C-H Bonds as Loci for Proline Assembly via C-H/O Interactions. *ChemBioChem* **2022**, *23*, e202200409.

(58) Wang, G.; Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589-1591.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.: Gaussian 09, Revision D.01. Gaussian, Inc.: Wallingford, CT, 2013.

(60) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. Direct MP2 gradient method. *Chem. Phys. Lett.* **1990**, *166*, 275-280.

(61) Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553-566.

(62) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J. Chem. Phys.* **1996**, *105*, 11024-11031.

(63) Glendening, C. R.; Landis, C. R.; Weinhold, F. Natural bond orbital methods. *WIREs Comput. Mol. Sci.* **2012**, *2*, 1-42.

(64) Glendening, E. D.; Landis, C. R.; Weinhold, F. NBO 6.0: Natural bond orbital analysis program. *J. Computational Chem.* **2013**, *34*, 1429-1437.

Graphical Table of Contents