Helical Twists and β-Turns in Structures at Serine–Proline Sequences: Stabilization of *cis*-Proline and type VI β-turns via C–H/O interactions

Harrison C. Oven, Glenn P. A. Yap, 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*. Co-author contact information: HCO: overh@udel.edu; GPAY: gpyap@udel.edu.

Data availability statement: NMR spectra used for compound characterization are in the Supporting Information. Coordinates of crystal structures are deposited with the CCDC. Coordinates of results from DFT calculations are in the Supporting Information. All other data involved in this work are available from the authors upon request.
Funding: NSF (CHE-2004110). Instrumentation support was provided by NIH (GM110758, S10 OD026896A) and NSF (CHE-1229234).
Conflict of interest statement: The authors declare no conflicts of interest.
Ethical statement: All standard ethical guidelines were followed in the conduct of this work.
Running title: Twists and Turns at Serine-Proline
Keywords: beta-turn, proline *cis-trans* isomerism, intrinsically disordered proteins, protein structure, hydrogen bonds, noncovalent interactions, bioinformatics

Abstract

Structures at serine-proline sites in proteins were analyzed using a combination of peptide synthesis with structural methods and bioinformatics analysis of the PDB. Dipeptides were synthesized with the proline derivative (2*S*,4*S*)-(4-iodophenyl)hydroxyproline [hyp(4-I-Ph)]. The crystal structure of Boc-Ser-hyp(4-I-Ph)-OMe had two molecules in the unit cell. One molecule exhibited *cis*-proline and a type VIa2 β-turn (BcisD). The *cis*-proline conformation was stabilized by a C–H/O interaction between Pro C–H_α and the Ser side-chain oxygen. NMR data were consistent with stabilization of *cis*-proline by a C–H/O interaction in solution. The other crystallographically observed molecule had *trans*-Pro and both residues in the PPII conformation. Two conformations were observed in the crystal structure of Ac-Ser-hyp(4-I-Ph)-OMe, with Ser adopting PPII in one and the β conformation in the other, each with Pro in the δ conformation and *trans*-Pro. Structures at Ser-Pro sequences were further examined via bioinformatics analysis of the PDB and via DFT calculations. Ser–Pro *versus* Ala-Pro sequences were compared to identify bases for Ser stabilization of local structures. C–H/O interactions between the Ser side-chain O_γ and Pro C–H_α were observed in 45% of structures with Ser-*cis*-Pro in the PDB, with nearly all Ser-*cis*-Pro structures adopting a type VI β-turn. 53% of Ser-*trans*-Pro sequences exhibited main-chain C=O_i•••H–N_{i+3} or C=O_i•••H–N_{i+4} hydrogen bonds, with Ser as the *i* residue and Pro as the *i*+1 residue. These structures were overwhelmingly either type I β-turns or N-terminal capping motifs on α-helices or 3_{10}-helices. These results indicate that Ser-Pro sequences are particularly potent in favoring these structures. In each, Ser is in either the PPII or β conformation, with the Ser O_γ capable of engaging in a hydrogen bond with the amide N–H of the *i*+2 (type I β-turn or 3_{10}-helix; Ser χ_1 *t*) or *i*+3 (α-helix; Ser χ_1 g^+) residue. Non-proline *cis* amide bonds can also be stabilized by C–H/O interactions.

Introduction

Serine–proline (Ser–Pro, SP) sequences are pervasive in protein structures. There are over 68,000 SP sequences in the human proteome, with SP sequences present in 76% of all human proteins.[1] Proline residues are inherently sites of structural modulation in proteins, capable of inducing disorder (via disruption of hydrogen-bonded α-helical and β-sheet secondary structures) or order (via stabilization of turns, loops, or the PPII conformation).[2-4] Ser-Pro sequences are important both in globular proteins and in intrinsically disordered regions of proteins (IDPs).[5,6] Of particular note, the human RNA Polymerase II C-terminal domain contains 52 imperfect repeats of the YSPTSPS sequence, with two SP sequences in each consensus repeat.[7-9] In addition, Ser–Pro sequences represent 25% of all phosphorylation sites *in vivo*.[10] A better understanding of the structural preferences at Ser–Pro sequences would provide insights into structures in IDPs, into inherent questions in protein structure and folding, and into how protein phosphorylation at SP sites can change protein structure.[11]

In addition to main-chain conformational preferences at the ϕ and ψ torsion angles, Pro sequences exhibit *cis-trans* isomerism about the X–Pro amide bond (ω torsion angle) (Figure 1). The identity of the residue before Pro significantly impacts the frequency of *cis*-Pro and the structures observed, in part via local interactions of the X residue with Pro or with adjacent residues.[2,12-15] We recently proposed that *cis*-Pro is stabilized in phosphoserine–proline (pSer-Pro) and Glu–Pro sequences via a C–H/O interaction between the side-chain anionic oxygen and Pro C–H_α.[16] More broadly, Pro residue C–H bonds are particularly likely to interact with oxygens via C–H/O interactions, due to the presence of multiple polarized C–H bonds on the proline ring.[17] C–H/O interactions are also observed with neutral oxygens of water molecules,

alcohols, and carbonyls.[18-24] Thus, we hypothesized that the Ser side-chain oxygen might modulate structure via C–H/O interactions with the proline ring C–H bonds.

Herein, we conduct a comprehensive analysis of protein structure at Ser–Pro sequences, including potential structural stabilization mediated by the Ser side-chain hydroxyl.[25-28] This work was performed using a combination of small-molecule synthesis and structural analysis in the solid and solution states; via bioinformatics analysis of Ser–Pro sequences in the PDB; and via DFT calculations on structures observed in Ser–Pro motifs.

Results and Discussion

Synthesis of Boc-Ser-hyp(4-I-Ph)-OMe and Ac-Ser-hyp(4-I-Ph). In order to examine Ser–Pro structures via small-molecule X-ray crystallography, we synthesized dipeptides with the unnatural proline derivative (2*S*,4*S*)-(4-iodophenyl)hydroxyproline [hyp(4-I-Ph)] (Scheme 1).[17,29,30] This amino acid promotes crystallization via the aryl iodide. The aryl iodide has also been employed for further amino acid modification via Suzuki or Sonogashira cross-coupling reactions, which can be conducted on peptides in water.[29,30] In addition, hyp(4-I-Ph) increases the likelihood of a *cis*-proline amide bond via the 4*S*-substitution of the aryloxy group, which leads to a greater preference for the proline *endo* ring pucker, the predominant ring pucker observed in *cis*-proline.[31-33]

Ser–Pro dipeptides were synthesized in solution from Boc-hyp(4-I-Ph)-OMe (**1**), whose synthesis was described previously.[30] Boc deprotection and amide coupling with Boc-Ser-OH and EDCI generated Boc-Ser-hyp(4-I-Ph)-OMe (**3**) (Scheme 1). In addition, **3** was converted into Ac-Ser-hyp(4-I-Ph) (**5**) in two steps via Boc deprotection and acetylation. In order to understand the specific effects of the Ser hydroxyl on structure, the Ala derivatives Boc-Ala-

hyp(4-I-Ph)-OMe (**6**) and Ac-Ala-hyp(4-I-Ph)-OMe (**8**) were synthesized as controls (Scheme S2).

X-ray crystallography of Ser-Pro dipeptides. Boc-Ser-hyp(4-I-Ph)-OMe crystallized from acetone and the crystal structure was determined. Interestingly, two molecules with distinct conformations were present in the unit cell, with one molecule exhibiting a *cis*-proline amide bond and the other molecule exhibiting a *trans*-proline amide bond (Figure 2). Crystal assembly was mediated by intermolecular hydrogen bonds between Ser hydroxyls; by C–H/O interactions at Pro C–H_δ and at both diastereotopic Ser C–H_β; and by a particularly close (2.94 Å) C=O•••I distance consistent with a strong halogen bond[34] between a Boc carbonyl and an aryl iodide (Figure 2a).

In the molecule with *cis*-Pro (ω = –12°), Ser was in the β conformation and Pro was in the δ conformation (Figure 2c). The overall structure was a type VIa2 β-turn (BcisD in the Dunbrack β-turn nomenclature[35]).[25,26,36-38] Type VIa2 β-turns are not mediated by hydrogen bonds between the *i* and *i*+3 residues of the turn, but are defined by the torsion angles and proximity of these termini. In this structure, the *cis*-Pro conformation appears to be stabilized by a C–H/O interaction between the Ser O_γ lone pair(s) and the Pro C–H_α, mediated via the Ser *t* χ_1 rotamer. C–H/O interactions are observed in diverse contexts in protein structures, including in recognition of α-helical GXXXG motifs via interaction of a main-chain carbonyl with Gly C–H_α.[23,39-43] C–H/O interactions are also observed at proline kinks in α-helices, where interaction with Pro C–H_α replaces a typical hydrogen bonding interaction with an amide hydrogen.[22,44] C–

H/O interactions also stabilize the ξ (pre-proline) conformation [(ϕ,ψ) ~ (–130°, +60°)], where the carbonyl oxygen of the residue i–2 to proline interacts with the Pro C–H_δ.[35,45,46]

The molecule of Boc-Ser-hyp(4-I-Ph)-OMe with *trans*-Pro exhibited both Ser and Pro in the PPII conformation (Figure 2d). The $n\rightarrow\pi^*$ interaction across Ser was relatively weak,[47,48] with a 3.14 Å O_{Ac}•••C=O_{Ser} distance, consistent with the low PPII propensity of Ser.[49-51] The Ser side chain was in the g^- χ_1 conformation. Pro exhibited an *exo* ring pucker, consistent with the relatively modest *endo*/*exo* ring pucker preference of aryloxy groups.[17,30]

Ac-Ser-hyp(4-I-Ph)-OMe crystallized with two different conformations for the molecule in the unit cell (Figure 3). Crystal packing (Figure 3a) was again mediated by hydrogen bonds, C–H/O interactions, and halogen bonds. Both conformations had a *trans*-proline amide bond with Pro in the δ conformation. In one conformation (Figure 3b), Ser was in the PPII conformation, with an intraresidue side chain-amide hydrogen bond similar to (but significantly longer than) that observed for phosphothreonine and phosphoserine.[11,52,53] As had been observed for Boc-Ser-hyp(4-I-Ph)-OMe, the $n\rightarrow\pi^*$ interaction across Ser was relatively weak. This conformation also exhibited Ser in the g^+ χ_1 conformation. Thus, across all structures, all three Ser χ_1 rotamers were represented, similar to what is observed in the PDB, where a relatively even distribution of Ser χ_1 rotamers is present.[54,55]

In the other conformation in the unit cell (Figure 3c), Ser was in the β conformation. The β structure was stabilized by an intraresidue C5 hydrogen bond.[56] A relatively long (weak) $n\rightarrow\pi^*$ interaction was observed across Pro. This $n\rightarrow\pi^*$ interaction is potentially weakened by the C5 hydrogen bond reducing electron density at the Ser carbonyl, making it a weaker electron donor for $n\rightarrow\pi^*$ interactions.[57,58]

In order to further understand the structures observed crystallographically, all structures were subjected to full geometry optimization via DFT computational methods (Figure S1). These calculations can identify whether the structures observed represent local energy minima, or alternatively whether interactions identified might be artefacts resulting from crystal packing. All four geometry-optimized structures exhibited conformations very similar to those observed crystallographically. In particular, the C–H/O interaction in Ser-*cis*-Pro was confirmed to be present in the local energy minimum identified via these calculations. In all cases, the computationally optimized and X-ray structures were very similar in all torsion angles and in their intramolecular noncovalent interactions. Geometry optimization did result in closer $n \rightarrow \pi^*$ interactions across Ser than were observed in the solid state. The calculations, by allowing full optimization of the position of the Ser hydroxyl hydrogen, also identified a potential O–H•••N interaction with the Pro amide nitrogen (analogous to N–H/N interactions[59,60]) that could help stabilize the δ structure at Pro in Ac-Ser-hyp(4-I-Ph)-OMe.

Analysis of dipeptides by NMR spectroscopy in solution. We sought to identify whether the C–H/O interaction observed at *cis*-proline in the solid state was present in solution. Therefore, we examined the NMR spectra of Ac-Ser-hyp(4-I-Ph)-OMe and Boc-Ser-hyp(4-I-Ph)-OMe in methanol, comparing the NMR spectra to those of the dipeptides with Ala, which lacks the Ser hydroxyl (Table 1).[11,61] Similar results were obtained either with an acetyl (Ac) or Boc N-terminal acyl group. Peptides with Ser had a significantly higher population of *cis*-proline (lower $K_{trans/cis}$) than those with Ala. *cis*-Pro was relatively more stable than *trans*-Pro in the Ser peptides compared to the Ala peptides by $\Delta\Delta G = 0.4–0.5$ kcal mol^{-1}. These results indicate either that the Ser hydroxyl stabilizes *cis*-proline, or that the Ser hydroxyl destabilizes *trans*-proline. Based on the crystallographic and computational data, we propose that Ser-*cis*-Pro is stabilized

in part by a C–H/O interaction between the Ser oxygen and the Pro C–H_α. Consistent with a C–H/O interaction,[11,16] the chemical shift of Pro H_α is downfield by 0.26–0.29 ppm in Ser-*cis*-Pro compared to Ser-*trans*-Pro. In contrast, for Ala–Pro peptides, the chemical shifts of Pro H_α are similar for Ala-*cis*-Pro and Ala-*trans*-Pro, which were also similar to those in Ser-*trans*-Pro. Collectively, these results suggest that the C–H/O interaction observed crystallographically and computationally contributes to stabilizing the Ser-*cis*-Pro conformation, even in solution with competitive hydrogen-bonding groups.

Bioinformatics analysis of Ser-Pro structures in the PDB. In order to more broadly understand the structures present at Ser-Pro sequences, we conducted a bioinformatics analysis of high-resolution protein structures in the PDB. The bioinformatics analysis included quantifying distances between relevant heavy atoms that would be consistent with noncovalent interactions, as well as determining key torsion angles associated with different conformations.

In order to identify the presence and quantify the frequency of C–H/O interactions between the Ser side-chain oxygen and proline C–H bonds, the distances between Ser O_γ and Pro C_α were measured as a function of proline amide conformation (Figure 4). Potential C–H/O interactions were defined as C•••O distances < 3.8 Å (~ 1.09 Å C–H bond length + 1.20 Å van der Waals radius of H + 1.52 Å van der Waals radius of O). 45% of Ser-*cis*-Pro structures exhibited a Pro/Ser C_α–H/O_γ interaction using this definition of C–H/O interactions. These results, in combination with other data above, strongly suggest that C–H/O interactions directly stabilize the *cis*-proline conformation at Ser–Pro sequences in proteins in water. In contrast, for Ser-*trans*-Pro sequences, essentially no Pro/Ser C_α–H/O_γ interactions were observed, although

28% of all sequences exhibited the possibility of a C–H/O interaction between Ser O_γ and Pro C–H_δ (Figure S13).

Analysis of the Ser χ_1 conformations (Table 2, Figure S3) revealed that the *t* rotamer is substantially more populated in Ser-*cis*-Pro structures (60%). The g^+ χ_1 conformation was relatively underpopulated for Ser-*cis*-Pro. In contrast, in Ser-*trans*-Pro structures, the three χ_1 rotamers are relatively equally populated, as has been observed previously for Ser.[54,55] Notably, in structures with C–H/O interactions at Ser-*cis*-Pro, the *t* rotamer was observed exclusively (Figure S4), as had also been seen by small-molecule X-ray crystallography (Figure 2c).

The Ramachandran plots for Ser-Pro and Ala-Pro sequences were analyzed as a function of proline amide conformation (Figure S11, Table 3). The data on structures with *trans*-proline indicated that, overall, both Ser and Ala populated the PPII (53% *versus* 50%), ζ (8% versus 10%), and α_L (1.1% *versus* 1.4%) conformations with similar frequencies. In contrast, Ala was more likely to adopt the α_R (14% *versus* 5%) conformation, while Ser was more likely in the β conformation (32% *versus* 23%). At proline, SP sequences were far more likely than AP sequences to have Pro in the α_R or δ conformation (65% for SP *versus* 37% for AP). In contrast, AP sequences were much more likely to have Pro in the PPII conformation (60% for AP *versus* 34% for SP).[62-64]

Analysis of the combinations of conformations of Ser and Pro at SP sites in proteins, compared to those in AP sequences, revealed substantial differences in structures of the dipeptides (Table 4). In structures with *trans*-Pro, AP sequences were significantly more likely to have both residues in the same secondary structure (both α-helix or both PPII). Overall, 47% of AP sequences had the same secondary structure at both residues, while only 20% of SP

sequences did. These results are consistent with the high propensities of Ala for both α-helix and PPII.[49-51,65,66] The observation that 80% of SP sequences had different conformations at Ser and Pro is consistent with SP sequences being particularly prominent in turns, loops, bends, and disordered structures.

In structures with *cis*-Pro (Table 3), the β conformation was substantially more likely for Ser than Ala (66% *versus* 33%), while Ala was far more likely to adopt PPII (60% *versus* 30%). The geometric constraints of *cis*-proline inherently cause the vast majority of structures with *cis*-proline to be β-turns, with Cα•••Cα *i*/*i*+3 distances ≤ 7.0 Å (92% of Ser-*cis*-Pro, 78% of Ala-*cis*-Pro).[15,35] Type VI β-turns include both backbone-hydrogen-bonded (type VIa1, PcisD) and non-backbone-hydrogen-bonded types (others) (Figure S5). SP sequences appear particularly likely to adopt type VIb (BcisP) β-turns (55% of SP *versus* 28% of AP).

We examined the role of C–H/O interactions as a function of type VI β-turn subtype in Ser-*cis*-Pro structures (Table 5, Figure S5). The majority of structures in type VIa1 (PcisD), VIa2 (BcisD), and PcisP (one type of VIb) β-turns exhibited C–H/O interactions crystallographically between the Ser O_γ and Pro C–H_α. C–H/O interactions were mediated exclusively via the *t* χ_1 rotamer (Figure S4a), but could be observed in different secondary structures of both Ser and Pro (Table 5). In contrast, BcisP type VIb β-turns, while the most commonly observed, exhibited the lowest frequency of C–H/O interactions (21% of structures).

The presence of hydrogen-bonded β-turns was examined for *trans*-Pro and *cis*-Pro structures, via quantifying the O_i•••H–N_{i+3} distance, with distances ≤ 2.7 Å indicating possible hydrogen-bonded β-turns (Table 6, Figure S14, Figure S16). This hydrogen bond pattern is also present in 3_{10}-helices and at the N-terminus of α-helices. Potential β-turns were examined with

two different registers: with Ser or Ala as the *i* residue of the β-turn (Pro–X central residues of the β-turn; Ser_i/Ala_i–Pro-X–X_{i+3} register), or with Ser or Ala as the *i*+1 residue of the β-turn (Ser-Pro or Ala-Pro central residues of the β-turn; X_i–Ser_{i+1}/Ala_{i+1}–Pro–X_{i+3} register). In registers with Ser–Pro or Ala–Pro as the central residues, 14% of *trans*-Pro structures with Ala as the *i*+1 residue were hydrogen-bonded β-turns or related hydrogen-bonded structures, while only 5% of structures were with Ser. For *cis*-Pro structures with this register, 24% and 37% of structures were hydrogen-bonded for Ser and Ala, respectively.

In contrast, in the register with Pro–X as the central residues, 51% of structures with Ser as the *i* residue had *i*/*i*+3 and/or *i*/*i*+4 hydrogen bonding patterns, while only 34% of those with Ala were. Among these hydrogen-bonded structures at SP, 43% are at the N-terminus of or within α-helices or 3_{10}-helices (22% of all Ser-*trans*-Pro sequences), with the other 49% representing β-turns (25% of all Ser-*trans*-Pro sequences) (Table 7). In contrast, no structures with *cis*-Pro in this register exhibited hydrogen-bonded β-turns. Collectively, these data suggest that Ser–Pro sequences with *trans*-Pro most frequently adopt β-turns with Ser as the *i* residue and with Pro–X as the central residues, and are significantly less likely to adopt β-turns with Ser-Pro as the central residues.

β-Turn types that can accommodate Pro as the *i*+1 residue include types I (AD), II (Pd or Pa), and VIII (AB1, AZ, AG). Further analysis of these data indicated that 86% of all β-turns in Ser-*trans*-Pro sequences with the Pro–X register were type I (AD) β-turns (Table 8). In type I β-turns, the Ser O_γ lone pair can hydrogen bond to the *i*+2 amide hydrogen (O_γ•••H–N_{i+2}), which is mediated by the *t* χ_1 rotamer.[25-28] This hydrogen-bonding pattern was observed frequently in the

PDB (Table 9), and results in intramolecular solvation of all amide hydrogens of the β-turn (Figure S7).

In SP structures in 3_{10}-helices and α-helices, the Ser residue commonly functions as the helix capping residue (Ncap). The Ncap residue (*i*) is not in the 3_{10}- or α-helical conformation, but provides the carbonyl that is hydrogen-bonded to the *i*+3 (3_{10}-helix) or *i*+4 (α-helix) amide N–H, which represents the first main-chain hydrogen bond of the helix (Figure S7).[67-69] 3_{10}-Helices and α-helices without Pro have 2 or 3, respectively, unsatisfied amide N–H hydrogen-bond donors at the N-terminus of the helix. The high frequency of Ser residues at the Ncap position (Table 7) is due to the ability of the Ser O_γ to function as a hydrogen-bond acceptor to the *i*+2 or *i*+3 amide N–H (Table 9).[63,70-73] Therefore, because Pro lacks an amide hydrogen, 3_{10}-helices and α-helices with an SP sequence at the N-terminus, and with the Ser O_γ functioning as a hydrogen-bonding amide capping group, only have 0 or 1 (respectively) solvent-exposed amide hydrogen at the helical N-termini. When functioning as a helix cap, O_γ•••H–N_{i+2} hydrogen bonds are associated with Ser in the *t* χ_1 rotamer, while O_γ•••H–N_{i+3} hydrogen bonds are associated with Ser in the g^+ χ_1 rotamer.

DFT calculations on combinations of conformations (conformational poses) observed at Ser–Pro sequences. In the bioinformatics work, we identified the most frequent sets of conformations observed at these residues, with a particular focus on structures that involve interactions with the Ser side-chain hydroxyl. Geometry optimization calculations were conducted using DFT methods on minimal peptides in order to better understand the roles of Ser noncovalent interactions in stabilizing specific conformations.

For structures with *cis*-Pro, C–H/O interactions were found to be local energy minima in all subtypes of type VI β-turns (Figure 5). For type VI β-turns with Ser in either the PPII conformation (type VIa1 PcisD, type VIb PcisP) or in the β conformation (type VIa2 BcisD, type VIb BcisP), C–H/O interactions were observed with a *t* χ_1 conformation at Ser. Notably, C–H/O interactions were energy minima with Pro in either the PPII or α_R/δ conformation, which are the two predominant energy minima of Pro in Ramachandran space. Overall, 4 different combinations of main-chain (Ser at PPII and β, Pro at PPII and α_R/δ) and side-chain (*t* χ_1) conformations exhibited close C–H/O interactions, consistent with bioinformatics data.

Hydrogen bonds with the Ser hydroxyl were observed in other structures, in both *trans*-Pro and *cis*-Pro (Figure S6). In the g^- and g^+ χ_1 conformations, the Ser hydroxyl can serve as a hydrogen bond donor with the carbonyl of the prior residue or of Ser, respectively. In the g^- or g^+ χ_1 conformations, Ser O_γ can also function as a hydrogen-bond acceptor to the Ser amide N–H. O–H/N interactions with the Ser amide N are also possible in the g^+ conformation. In structures with *trans*-Pro, hydrogen bonds of the Ser O–H with the Pro carbonyl were observed in the *t* χ_1 conformation.

Structures of SP sequences in type I β-turns (Ac-S–PA–NHMe), with Ser as the *i* residue, demonstrated the specific stabilization provided by Ser O_γ serving as a hydrogen-bond acceptor to the *i*+2 residue amide N–H (Figure 6). With Ser as the Ncap residue of a 3_{10}-helix, and with Ser in either the PPII or β conformation, a hydrogen bond between Ser O_γ and the *i*+2 residue amide N–H resulted in there being no solvent-exposed amide N–H hydrogens at the N-terminus of the 3_{10}-helix (Figure 7). This hydrogen bond is mediated via the *t* χ_1 rotamer. In contrast, in an

α-helix stabilized by a Ser_i O_γ•••H–N_{i+3} hydrogen bond, this capping interaction occurs via the g^+ χ_1 rotamer (Figure 8). Collectively, these computational results demonstrate a multitude of ways that the Ser hydroxyl can stabilize local structures in SP sequences.

Discussion

Herein, we conducted a comprehensive analysis of structures stabilized at Ser–Pro sequences in proteins, using a combination of small-molecule X-ray crystallography, solution NMR spectroscopy, bioinformatics analysis of proteins in the PDB, and DFT calculations. SP sequences are particularly prone to adopt local structures stabilized by hydrogen bonds or by C–H/O interactions.

Ser-trans-Pro sequences stabilize type I β-turns. 51% of SP sequences in the PDB with *trans*-Pro exhibit a hydrogen bond between the Ser (*i* residue) C=O and the amide N–H of the *i*+3 and/or *i*+4 residue. These hydrogen-bonding patterns are present in β-turns, 3_{10}-helices, and α-helices. β-turns were observed at 25% of all Ser-*trans*-Pro sequences with Ser as the *i* residue and Pro as the *i*+1 residue, with the vast majority type I β-turns. Notably, the first turn of 3_{10}-helices and α-helices has the same conformation at the *i*+1 and *i*+2 residues (α_R/δ) as a type I β-turn.

The key interaction of Ser that stabilizes a type I β-turn is a hydrogen bond between the Ser Oγ and the amide N–H of the *i*+2 residue (Figure 6). This interaction has previously been identified to stabilize type I β-turns.[25,27,28] More broadly, residues with the hydrogen bond acceptors Asp, Asn, and Ser are substantially overrepresented at the *i* position of type I β-turns, and Pro is by far the most frequent residue at the *i*+1 position of type I β-turns.[25,26] A hydrogen

bond between an oxygen of the *i* residue side chain and the amide N–H of the *i*+2 residue inherently stabilizes the type I β-turn conformation, via intra-turn solvation of that amide, which prevents competing hydrogen-bonding interactions at that amide that would otherwise lead to different structures. These data strongly suggest that type I β-turns should be a major conformation at SP sequences in proteins, including in intrinsically disordered proteins.

Ser-Pro sequences stabilize the N-termini of 3_{10}-helices and α-helices. Similarly, we found that 22% of all Ser-*trans*-Pro sequences in the PDB represent helical capping motifs. Ser is well-recognized to function as a common hydrogen-bond acceptor in α-helical N-terminal capping motifs.[69,71,72,74] Proline is also a common residue at the N-termini of α-helices, due to Pro not having an amide hydrogen. Amide hydrogens are solvent-exposed in the first 3 residues of the α-helix and thus do not participate in intrahelical hydrogen bonds.[2,68,74-76] The combination of Ser at the Ncap position of an α-helix, with the Ser O_γ hydrogen-bonded to the N3 amide N–H, and Pro at the N1 position, results in only one solvent-exposed amide N–H at the α-helical N-terminus (at residue N2) (Figure 8). Hydrogen bonding of a side-chain hydrogen-bond acceptor to the N3 amide N–H might be particularly favorable, due to the relatively weak nature of amide-water N–H•••OH_2 hydrogen bonds at the N3 amide.[72,77,78]

3_{10}-Helices are common in short helical structures, due to the *i*/*i*+3 register of hydrogen bonds yielding a larger total number of intrahelical hydrogen bonds in a short sequence versus those with an α-helix.[67,75,79] Mechanistic studies of protein folding also suggest that 3_{10}-helices are intermediates in the formation of α-helices, as they require the organization of fewer residues to form a helical turn.[80-82] The data herein suggest that SP sequences might promote the 3_{10}-helix to α-helix conversion in protein folding. 3_{10}-Helices are stabilized in SP sequences by a Ser

O_γ•••H–N_{i+2} hydrogen bond, which is mediated by Ser in the *t* χ_1 rotamer (Figure 7). Rotation of Ser to the g^+ rotamer promotes instead a hydrogen bond with the N3 amide N–H (Figure 8), as is observed in α-helical capping motifs. Thus, simple rotation about Ser χ_1 at SP sequences might favor a change in structure from 3_{10}-helix to α-helix, thereby promoting adoption of native protein structure at α-helices.

Both in β-turns and in helical capping motifs, Ser O_γ can stabilize structures via hydrogen bonds, to either the *i*+2 or *i*+3 amide hydrogen, when Ser is in either the PPII or β main-chain conformation. The ability to stabilize secondary structures via either PPII or β is consistent with the high frequencies of SP sequences with Ser in either the PPII or β conformation and with Pro in the α_R or δ conformation (Table 4). Even without an intramolecular hydrogen bond, the crystal structure of Ac-hyp(4-I-Ph)-OMe (Figure 3) exhibited the same combinations of conformations (Ser in PPII or β, Pro in α_R/δ) as are present in β-turns and helical capping motifs, consistent with these structures being strongly favored in SP sequences.

Changes in structure at SP sequences due to proline cis-trans isomerism. The vast majority of SP sequences with *cis*-Pro adopt a type VI β-turn structure. β-Turns at Ser-*cis*-Pro are centered on SP (Ser at the *i*+1 residue, Pro at the *i*+2 residue). In contrast, β-turns at Ser-*trans*-Pro are predominantly centered on Pro-X (Ser at the *i* residue, Pro at the *i*+1 residue). Proline *cis-trans* isomerization inherently generates large changes in protein structure.[3,4,83-85] At SP sequences, the Ser-*trans*-Pro to Ser-*cis*-Pro amide conformational change both induces a β-turn and switches the register of any β-turn present, and thus can change protein structure even if a β-turn is present with *trans*-Pro. In addition, we recently proposed that pSer-Pro sequences can promote β-turns with pSer the *i*+1 residue and Pro the *i*+2 residue,[16] and thus Ser

phosphorylation at SP sequences could also change structure in part via changing both the presence and the register of β-turns in proteins.

The RNA Polymerase II C-terminal domain in eukaryotes consists of consensus repeats of the YSPTSPS sequence, with the number of repeats and identities of non-consensus residues differing between organisms.[7,86] Humans have 52 such repeats, with each consensus repeat containing two SP sequences (one SPT and one SPS). Bioinformatics analysis of proteins (by others and by us) and DFT calculations found that the SPS and SPT sequences are particularly prone to stabilization of type I β-turns, with the Ser O_γ•••H–N_{i+2} hydrogen bond further stabilized by a hydrogen bond between Ser_i O_γ–H and Ser/Thr_{i+2} O_γ (Figure 9, Figure S8).[6,25-28,87-89] The Pol II CTD has previously been identified to be conformationally heterogeneous, with a significant propensity to adopt β-turn conformations.[9,87,90-93] The data herein further confirm the favorability of type I β-turns within the Pol II CTD. Moreover, phosphorylation of Ser2, Thr4, and/or Ser5 of the Pol II CTD is central to changes in transcriptional elongation and termination, providing a "phosphorylation code" for transcription.[7,8] The data herein provide further context for considering the nature of the transcriptional code of structural changes in the Pol II CTD.

The cis-Proline conformation is stabilized by C–H/O interactions in Ser-cis-Pro sequences. Herein, we identified that *cis*-Pro can be stabilized at SP sequences via a C–H/O interaction between Ser Oγ and Pro C–H_α. We observed evidence of this interaction in the solid state by small-molecule X-ray crystallography; in solution via an increase in the population of *cis*-Pro in SP dipeptides versus AP dipeptides and via a downfield change in the Pro Hα chemical shift; via bioinformatics analysis of Ser-*cis*-Pro structures in the PDB; and via DFT calculations. C–H/O interactions are inherently favorable at the polarized C–H bonds of

proline.[17] We observed that C–H/O interactions can stabilize all type VI β-turn subtypes. Notably, the crystal structure of the marine natural product stylopeptide 1 (CSD ZOHMIS) exhibits Ser-*cis*-Pro in a type VIa1 β-turn (PcisD) conformation, with a close (2.57 Å) Pro C–H_α•••O_γ Ser C–H/O interaction that appears to stabilize the *cis*-proline conformation, which is also present in solution.[94]

Thus, while C–H/O interactions mediated by a neutral Ser O_γ lone pair are inherently weaker[20,95] than those of anionic Glu-*cis*-Pro and pSer-*cis*-Pro, the data indicate that they are still significant in peptides and proteins in water. While C–H/O interactions are typically described in electrostatic terms (interaction of the δ^+ on a polarized C–H bond with the δ^- on O), they are predominantly stereoelectronic (molecular orbital-based) in nature, and exhibit only a weak dependence of interaction strength on solvent dielectric constant.[17] The stereoelectronic nature of C–H/O interactions can explain the common observation of C–H/O interactions stabilizing diverse protein structures.[22-24,96]

C–H/O interactions can stabilize cis-non-proline amide bonds. Herein, we identified that C–H/O interactions between a Ser side-chain O and Pro C–H_α are commonly observed and stabilizing to the *cis*-Pro conformation. *cis* amide bonds are also observed at 0.03% of non-proline residues (versus at 5% of proline residues).[13,15,83,97] While *cis* amide bonds are much less likely at non-proline residues than at Pro for any given amide bond, over 95% of all amide bonds in proteins are at non-proline residues. Thus, while *cis*-Pro represents approximately 0.25% of all amide bonds in proteins (5% × 5%), *cis*-non-Pro represents approximately 0.03% of all amide bonds in proteins (95% × 0.03%), or 1 out of every ~3000 amide bonds. As such, the overall frequency of *cis*-non-Pro amide bonds is substantial, representing ~10% of all *cis* amide bonds in

proteins. Because *cis*-non-Pro amide bonds are substantially higher in energy than *trans*-non-Pro amide bonds (the 0.03% frequency suggests that *cis*-non-Pro is approximately 5 kcal mol^{-1} higher in energy than *trans*-non-Pro), any observed *cis*-non-Pro amide bond needs to be substantially stabilized by other interactions within the protein.[98,99]

In addition, *cis-trans* isomerization from *trans*-non-Pro to *cis*-non-Pro is characterized by extremely slow kinetics ($t_{1/2}$ ~ 1000 s), with rapid reversion from *cis*-non-Pro to *trans*-non-Pro ($t_{1/2}$ < 1 s) in simple peptides.[100,101] These kinetics and thermodynamics suggest a significant role for local structures in stabilizing observed *cis*-non-Pro amide bonds.[101-104] The rate of *cis-trans* amide interconversion at non-*cis*-Pro amide bonds is accelerated by molecular chaperones such as the *E. coli* Hsp70 DnaK, which functions as a non-prolyl amide isomerase.[105] Notably, *cis*-non-Pro amide bonds are structurally highly conserved, and are usually present at or near functional sites in proteins.[97,106-111] Therefore, we examined protein structures to see whether C–H/O interactions between a Ser side-chain O and C–H_α of the subsequent residue might stabilize the *cis* amide bond at Ser-*cis*-non-Pro structures.

Xanthine-guanine phosphoribosyl transferase (XGPRT) is a critical enzyme for numerous bacteria, functioning in purine salvage. X-ray crystallography has demonstrated that a conserved Ser-*cis*-Arg in XGPRT[112,113] is present (or implied by sequence conservation) in divergent bacteria, including *E. coli*, *Salmonella*, *Shigella*, and *Yersinia pestis*. This non-prolyl *cis* amide bond amide bond is part of the phosphoribose binding site, with the *cis*-Arg amide N–H hydrogen bonded to the ribosyl phosphate (Figure 10). Structures from *E. coli* and *Yersinia pestis* exhibit a close C–H/O interaction between the Ser O_γ and the Arg C–H_α, which helps stabilize the *cis*-Arg amide bond. Geometry optimization of the local structures confirms the presence of a C–H/O interaction that appears to stabilize the *cis*-Arg conformation (Figure 10b). These

crystallographic data suggest that C–H/O interactions might be a potentially general method to stabilize *cis* amide bonds in proteins.

Methods

Peptide synthesis. Dipeptides were synthesized by solution-phase synthesis (Schemes S1-S2). Synthesis and characterization details are in the Supporting Information.

NMR spectroscopy. Compounds were characterized at 298 K in $CDCl_3$ or CD_3OD using either a Bruker 400 MHz NMR spectrometer equipped with a cryogenic QNP probe or a Bruker 600 MHz spectrometer equipped with a 5-mm Bruker SMART probe (Table S1). 1-D NMR spectra were obtained using an excitation sculpting pulse sequence. Peaks were assigned using COSY and ^{1}H-^{13}C HSQC spectra. NMR spectra and additional experimental details are in the Supporting Information.

X-ray crystallography. Boc-Ser-hyp(4-I-Ph)-OMe and Ac-Ser-hyp(4-I-Ph)-OMe were allowed to dissolve in acetone and subjected to crystallization via evaporation at 25 °C. Crystallization of both compounds occurred over one week. Additional details are in the Supporting Information (Figure S18-S19, Table S12-S26).

Computational chemistry. Calculations were conducted with Gaussian 09.[114] For small-molecule crystal structures, hydrogen positions were determined using a restrained geometry optimization, in which the positions of the heavy atoms were fixed based on those observed crystallographically, while the positions of the hydrogens were optimized using the M06-2X DFT functional and the Def2TZVP basis set in implicit water (IEFPCM).[115-117] Full, unrestrained geometry optimization of molecules derived from crystal structures were determined using the same methods (Figure S1). Structures of all other molecules were determined via iterative

geometry optimization, with final geometry optimization using the M06-2X functional with the 6-311++G(d,p) basis set in implicit water.[118] Additional details of calculations and coordinates of all geometry-optimized structures are in the Supporting Information (Figures S2, S6).

Bioinformatics. On June 14, 2023, a search of the Protein Data Bank (PDB) was conducted for structures containing Ser-Pro sequences with a resolution ≤ 2.0 Å and a sequence similarity less than 30%. A search of the PDB for structures containing Ala-Pro sequences was conducted on July 17, 2023, using the same search parameters. Perl scripts were written to extract and calculate information used in analyses such as residue numbers, residue identities, dihedral angles, Ramachandran plots, interatomic distances, and amide hydrogen position estimations. Structures used in analyses were manually filtered to remove structures with nearby broken backbone amide bonds. The final data sets contained 2045 structures with SP sequences (1929 structures with *trans*-proline and 116 structures with *cis*-proline) and 1832 structures with AP sequences (1729 with *trans*-proline and 103 with *cis*-proline). Additional details are in the Supporting Information (Figures S9-S17, Tables S2-S11).

Acknowledgements

We thank Himal Ganguly for preliminary analysis of C–H/O interactions in Ser-Pro sequences. We thank NSF (CHE-2004110) for funding. Instrumentation support was provided by NIH (GM110758, S10 OD026896A) and NSF (CHE-1229234).

Conflict of interest statement

The authors declare no conflicts of interest.

Data availability statement

NMR spectra used for compound characterization are in the Supporting Information. Coordinates of crystal structures are deposited with the CCDC. Coordinates of results from DFT calculations are in the Supporting Information. All other data involved in this work are available from the authors upon request.

Author contributions

HCO conducted all synthesis, characterization, and analysis of molecules; conducted all bioinformatics analysis; conducted DFT calculations; and wrote the manuscript. GPAY conducted X-ray crystallography and solved the crystal structures; NJZ conducted DFT calculations, wrote the manuscript, obtained funding for the work, and supervised the project. All authors contributed to editing of the manuscript.

Supporting Information Available

Details of peptide synthesis and characterization; additional analysis of structures determined by DFT calculations; additional analysis of bioinformatics data; details of structures solved by X-ray crystallography; and coordinates of all geometry-optimized structures. This material is available free of charge via the journal web site.

Figures

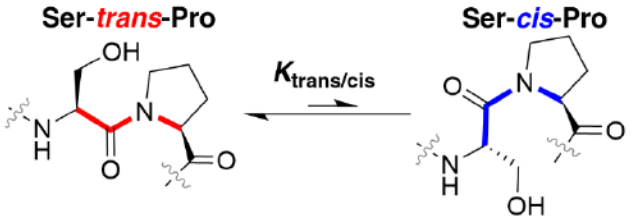


Figure 1. ***trans*-Proline and *cis*-proline in Ser-Pro sequences**.

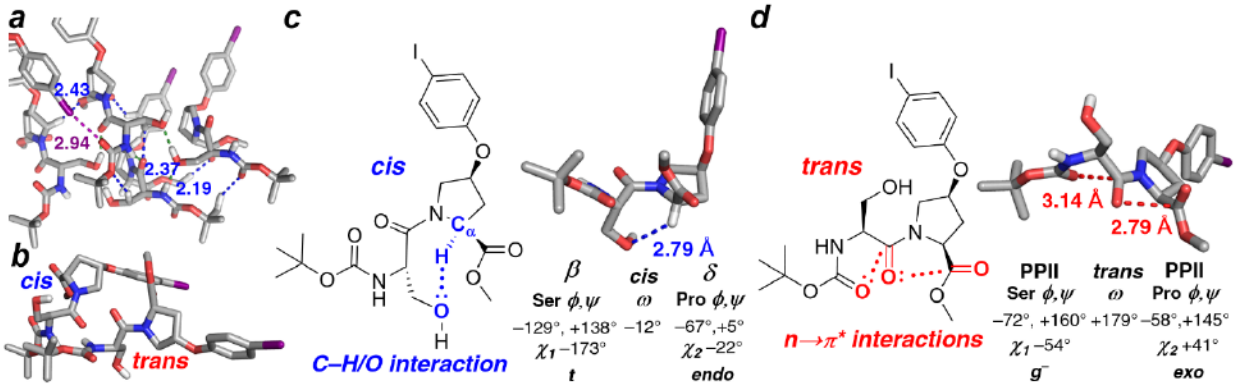


Figure 2. X-ray crystal structure of Boc-Ser-hyp(4-I-Ph)-OMe with two molecules in the unit cell, one with *trans*-proline and one with *cis*-proline. (a) Crystal assembly is mediated by intermolecular interactions including C–H/O interactions, blue; N–H•••O and O–H•••O hydrogen bonds, green; and a C–I•••O halogen bond, purple. Interatomic distances of the halogen bond and of close C–H/O interactions are indicated (Å). (b) The unit cell contains one molecule with the *cis*-proline conformation and one molecule with the *trans*-proline conformation. (c) The molecule with *cis*-proline exhibited an *endo* ring pucker, the β conformation at Ser, and the δ conformation at Pro. This conformation represents a type VIa2 (BcisD) β-turn. The O_{Boc}•••O_{OMe} distance, which is equivalent to a $C_{\alpha,i}$•••$C_{\alpha,i+3}$ distance in a protein context, is 4.67 Å. This molecule has a C–H/O interaction between Ser O_γ and Pro H_α (C_α–H•••O distance 2.79 Å, blue). (d) The molecule with *trans*-proline exhibited an *exo* ring pucker and the PPII conformation at both Ser and Pro, with PPII stabilized by consecutive n→π* interactions. The positions of the hydrogens in both molecules [(c) and (d)], which are not definitively determined in the crystal structure, were optimized using the M06-2X DFT functional with the Def2TZVP basis set and implicit H_2O solvation, while keeping the crystallographically determined positions of the heavy atoms fixed.

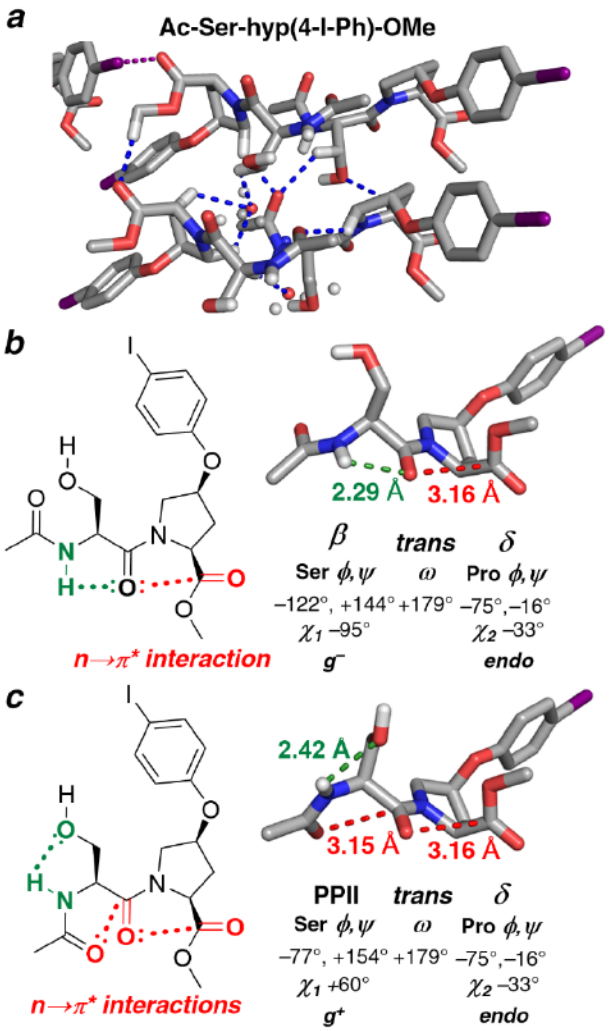


Figure 3. Crystal structure of Ac-Ser-hyp(4-I-Ph)-OMe. (a). Crystal packing, with intermolecular C–H/O interactions (blue) and C–**I•••**O halogen bonds (purple) highlighted. In the unit cell, two distinct conformations were observed at Ser (β or PPII). Pro was in the δ conformation with a *trans*-Pro amide for both conformations at Ser. (b) In one conformation, Ser is in the β conformation, with the amide hydrogen engaged in an intraresidue backbone C5 hydrogen bond (N–H•••O distance 2.29 Å, green). The Ser side chain is not involved in any intramolecular hydrogen bonds. (c) In the other conformation, Ser is in the PPII conformation. The Ser side chain O_γ is engaged in an intraresidue hydrogen bond with its amide hydrogen (N–H•••O distance 2.42 Å, green).

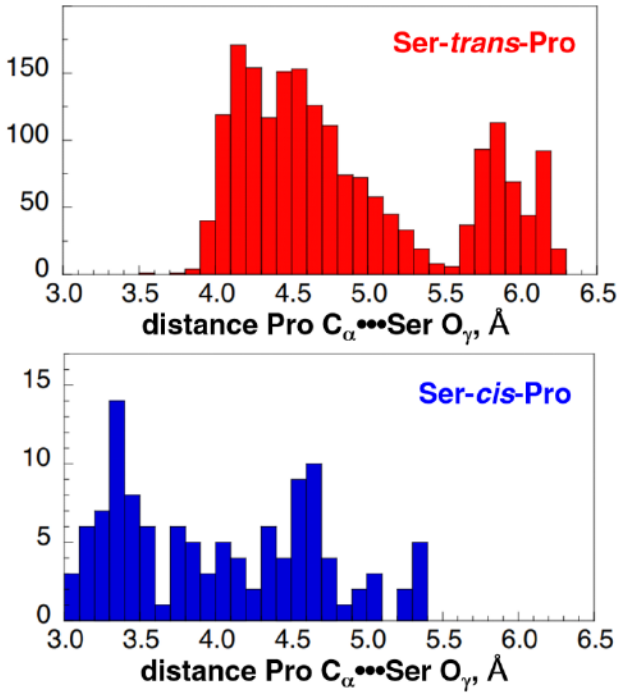


Figure 4. Potential C–H/O interactions in Ser-Pro sequences in the PDB. Ser-Pro sequences were analyzed for the distance between Pro C_α and Ser O_γ, in both Ser-*trans*-Pro (top) and Ser-*cis*-Pro (bottom). The Pro C_α•••Ser O_γ distance was used to identify C_α–H/O interactions, as few PDB structures have accurate hydrogen positions. A Pro C_α•••Ser O_γ distance ≤ 3.8 Å is consistent with a C_α–H/O interaction. Using these parameters, 45% of structures containing *cis*-proline have a C_α–H/O interaction (52 total structures). Only 2 structures containing *trans*-proline were observed with Pro C_α•••Ser O_γ distances < 3.8 Å; both had Ser O_γ–H•••Pro O hydrogen bonds. The data set includes all crystal structures containing a SP sequence solved with a resolution ≤ 2.0 Å and sequence similarity < 30%. This search resulted in a total of 2045 structures: 1929 structures with *trans*-proline and 116 structures with *cis*-proline.

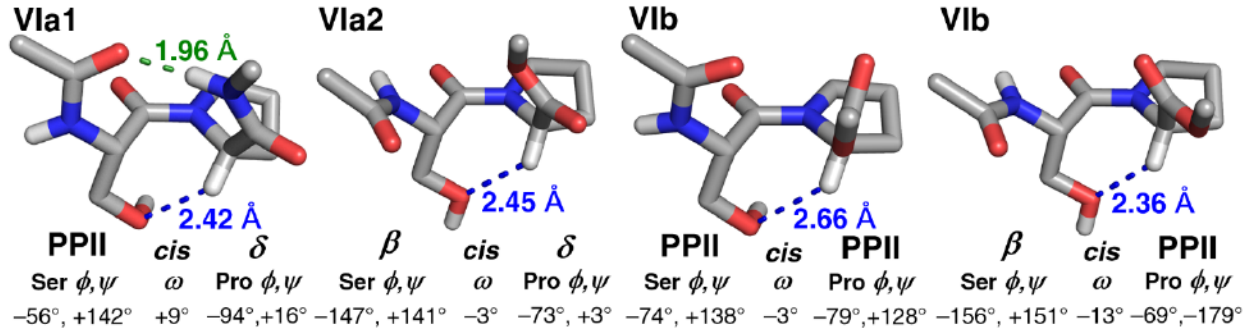


Figure 5. Structures of type VI β-turns stabilized via C–H/O interactions. Structures of (type VIa1) Ac-Ser-Pro-NHMe and (others) Ac-Ser-Pro-OMe type VI β-turns with C–H/O interactions (blue) between Ser O_γ and Pro C–H_α that stabilize the *cis*-Pro conformation in Ser-*cis*-Pro. These structures were derived from those identified via bioinformatics, then subjected geometry optimization by DFT methods. All C–H/O interactions are mediated via the Ser *t* χ_1 rotamer. Type VIa1 β-turns are mediated via *i*/*i*+3 main-chain O_i•••H–N_{i+3} hydrogen bonds (green), while other type VI β-turn subtypes are not.

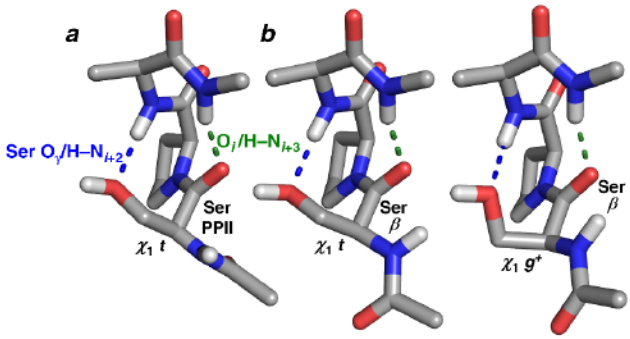


Figure 6. Stabilization of type I β-turns by Ser$_i$ O_γ•••H–N_{i+2} hydrogen bonds. (a,b) Geometry-optimized structures of Ac-Ser_i–Pro_{i+1}-Ala_{i+2}–N_{i+3}HMe type I β-turns (mediated by a main-chain *i*/*i*+3 C_i=O••••H–N_{i+3} hydrogen bond, green) in which the Ser O_γ engages in a hydrogen bond (blue) with the amide N–H of the *i*+2 residue of the turn (here, Ala). The combination of Pro as the *i*+1 residue and the Ser hydrogen bond to the *i*+2 residue amide results in no solvent-exposed amide hydrogens within the β-turn. Ser is the *i* residue of the turn, with Ser in the (a) PPII or (b) β conformation. The side chain-main chain hydrogen bond is mediated (a) via the *t* χ_1 rotamer when Ser adopts the PPII conformation or (b) via either the (left) *t* or (right) g^+ χ_1 rotamer when Ser is in the β conformation. The type I β-turn requires a *trans*-proline amide (Ser-*trans*-Pro).

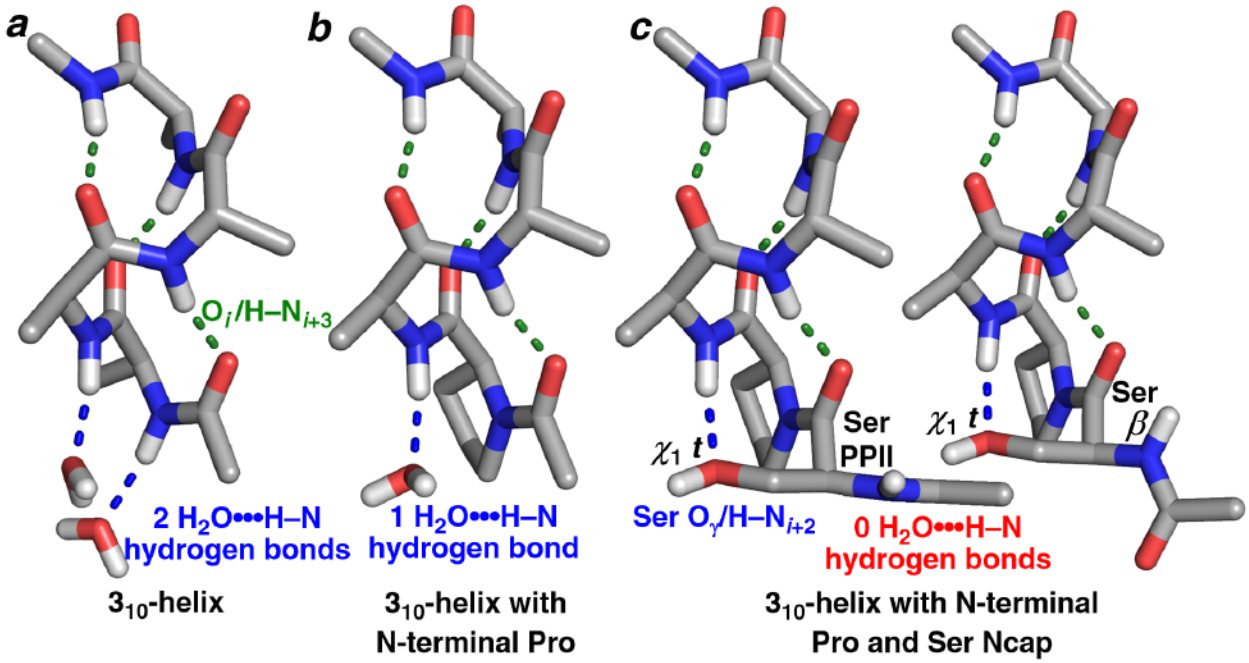


Figure 7. Solvation at the N-termini of 3_{10}-helices, including 3_{10}-helix capping by a Ser O_γ•••H–N_{i+2} hydrogen bond. Geometry-optimized structures of (a) Ac-Ala_4-NHMe, (b) Ac-Pro-Ala_3-NHMe, and (c) Ac-Ser-Pro-Ala_3-NHMe, with Pro and Ala residues in the 3_{10}-helix conformation. In (c), Ser functions as the Ncap residue (*i*) of the 3_{10}-helix, with a Ser O_γ hydrogen bond with the amide N–H of the *i*+2 residue, and with Ser in either the (left) PPII or (right) β conformation. This Ser O_γ•••H–N_{i+2} side chain-main chain hydrogen bond is mediated via Ser in the *t* χ_1 rotamer. In (a), the N-terminus of the 3_{10}-helix has 2 unsatisfied amide hydrogen-bond donors which must be solvated. When (b) the N-terminal residue of the 3_{10}-helix is Pro, only one unsatisfied amide hydrogen-bond donor is present. In 3_{10}-helices with Ser as the Ncap (*i*) residue and Pro as the first (*i*+1) residue, there are no N-terminal amide hydrogens that must be solvated, as a result of the helix-capping Ser O_γ•••H–N_{i+2} hydrogen bond. A similar example of a capped 3_{10}-helix stabilized via a SP sequence is in Figure S7. Geometry optimization was conducted using the M06-2X DFT functional, the 6-311++G(d,p) basis set, and implicit water solvation.

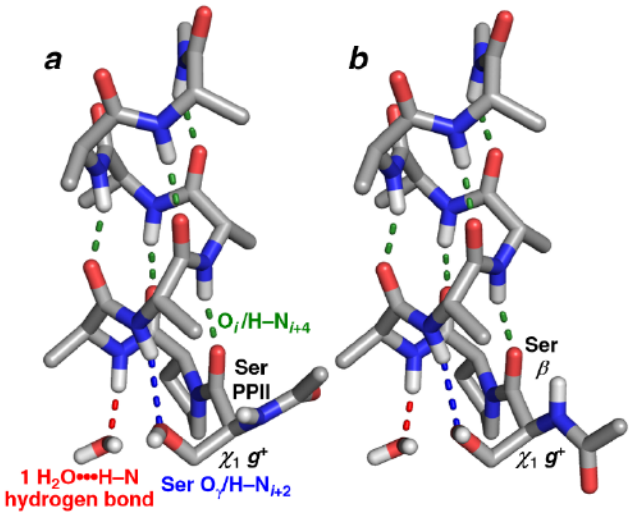


Figure 8. N-Capping of α-helices in Ser$_i$-*trans*-Pro$_{i+1}$ sequences via a Ser O_γ•••H–N_{i+3} hydrogen bond. Geometry-optimized structures of Ac-Ser-*trans*-Pro-Ala_6-NHMe, with Pro and Ala residues in the α-helix conformation, with the N-terminal Ser in (a) the PPII conformation or (b) the β conformation. In both structures, a Ser O_γ•••H–N_{i+3} hydrogen bond, mediated via the Ser g^+ χ_1 rotamer, results in only one unsatisfied N-terminal amide hydrogen-bond donor (at N2). In contrast, an uncapped α-helix without an N-terminal Pro has three unsatisfied amide N–H hydrogen-bond donors.

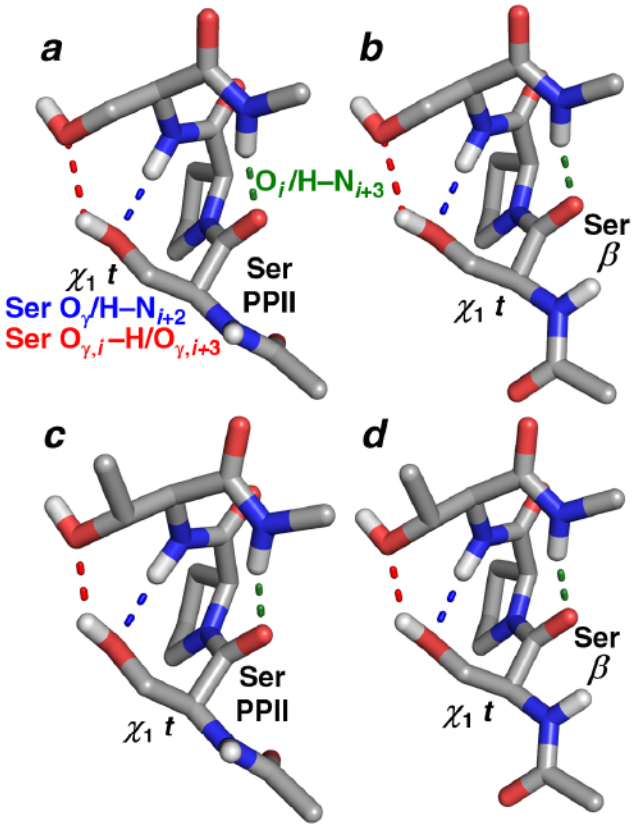


Figure 9. Type I β-turns in SPS and SPT sequences stabilized via simultaneous Ser_i O_γ•••H–N_{i+2} and Ser_i O_γ–H•••O_γ Ser_{i+2} hydrogen bonds. (a–d) Optimized structures of (a,b) Ac-Ser_i–Pro_{i+1}-Ser_{i+2}–N_{i+3}HMe or (c,d) Ac-Ser_i–Pro_{i+1}-Thr_{i+2}–N_{i+3}HMe in a type I β-turn mediated by a main-chain i/i+3 hydrogen bond (green). Ser_i O_γ simultaneously acts as a hydrogen-bond acceptor to the main-chain amide H–N_{i+2} (blue) and as a hydrogen-bond donor to the Ser_{i+2} or Thr_{i+2} side chain O_γ (red). This dual hydrogen-bonding interaction of the Ser_i side chain is mediated by Ser in the t χ_1 rotamer, with Ser in either (a,c) the PPII conformation or (b,d) the β conformation.

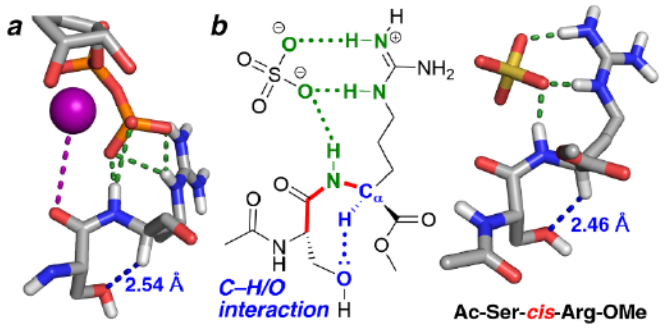
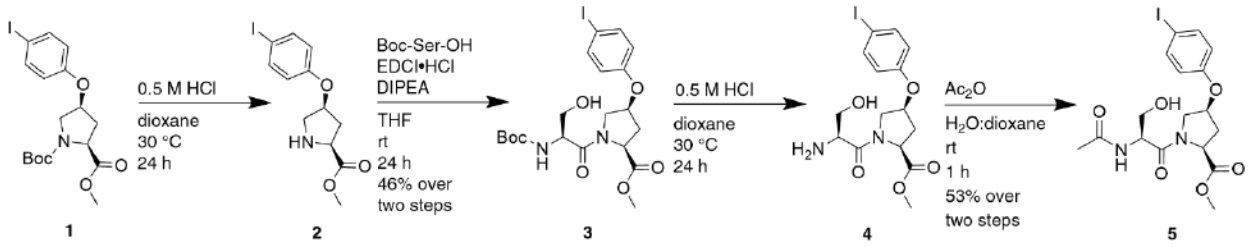


Figure 10. Ser-*cis*-nonPro structures with C–H/O interactions. (a) X-ray crystal structure of *E. coli* Xanthine-guanine phosphoribosyltransferase (XGPRT) bound to a 5-phospho-alpha-D-ribose 1-diphosphate (PRPP) analog (PDB 1a95, 2.00 Å resolution). The Ser-*cis*-Arg motif binds the anionic phosphate via N–H•••O hydrogen bonds (green) as well as the magnesium cofactor (Mg^{2+}•••O, purple). Ser stabilizes the *cis*-Arg conformation via a C–H/O interaction (C_α–H•••O distance 2.54 Å, blue). The Ser-*cis*-Arg motif is conserved in purine scavenging enzymes across multiple pathogenic bacteria. (b) ChemDraw representation of the structure of XGPRT from *Yersinia pestis* (PDB 6kp5, 1.10 Å resolution), with a sulfate anion in the active site and an equivalent C–H/O interaction (C_α–H•••O distance 2.57 Å). The core Ser-*cis*-Arg•SO_4^{2-} structure from this protein was truncated to Ac-Ser-*cis*-Arg-OMe•SO_4^{2-} and subjected to full geometry optimization (right), which demonstrated a close C–H/O interaction (C_α–H•••O distance 2.46 Å, blue) stabilizing the *cis*-Arg conformation.

Schemes



Scheme 1. Synthesis of Ac-Ser-hyp(4-I-Ph)-OMe.

Tables

Table 1. Summary of ^{1}H NMR data on Ser-hyp(4-I-Ph) and Ala-hyp(4-I-Ph) dipeptides in d_4-MeOH.[a,b]

		δ Pro H_α, ppm		
	$K_{trans/cis}$	δ_{cis}	δ_{trans}	$\Delta\delta_{cis-trans}$
Boc-Ser-hyp(4-I-Ph)-OMe	2.6	5.02	4.76	+0.26
Boc-Ala-hyp(4-I-Ph)-OMe	5.9	4.78	4.74	+0.04
Ac-Ser-hyp(4-I-Ph)-OMe	2.3	5.03	4.74	+0.29
Ac-Ala-hyp(4-I-Ph)-OMe	4.2	4.78	4.74	+0.04

[a] $K_{trans/cis}$ is the equilibrium constant between *trans*-Pro and *cis*-Pro, calculated as the ratio of the volumes of peaks with *trans*-Pro versus *cis*-Pro. For Ac-Ser-Pro-OMe, $K_{trans/cis}$ = 6.8.[11] The lower $K_{trans/cis}$ of Ac-Ser-hyp(4-I-Ph)-OMe is consistent with previously observed effects of the 4*S*-aryloxy group to increase the population of *cis*-proline.[29,30] In general, higher frequencies of *cis*-proline are observed in peptides than in globular proteins, due to the absence of tertiary structure stabilizing the *trans*-proline conformation in peptides.[12,13,15,61]

[b] $\Delta\delta_{cis-trans} = \delta_{cis}$ (Pro H_α) – δ_{trans} (Pro H_α)

Table 2. Ser χ_1 rotamer populations at Ser-*trans*-Pro and Ser-*cis*-Pro sequences in the PDB. Populations of each Ser χ_1 rotamer are represented as a percent of structures in each Pro amide conformation.

	rotamer	χ_1	$\%_{total}$
trans	g^-	–60	24.0
	g^+	+60	39.0
	t	±180	37.0
cis	g^-	–60	28
	g^+	+60	12
	t	±180	60

Table 3. Conformations of individual residues in Ser-Pro and Ala-Pro sequences in the PDB. Populations of conformations observed at individual residues in Ser-Pro and Ala-Pro sequences in the PDB. Populations are represented as a percent of structures in each Pro amide isomer. The torsion angle ranges used to define the regions of the Ramachandran plot are in Figure S10 and Table S2.

		% **X** at X–Pro		% **Pro**	
	conformation	**Ser**	**Ala**	**at SP**	**at AP**
trans	α_R	4.9	14.1	52.3	32.1
	δ	0.1	0.1	12.3	5.1
	β	31.7	23.4	0.1	0.2
	PPII	52.6	49.6	34.1	60.0
	ζ	8.1	9.8	0.0	0.0
	α_L	1.1	1.4	n.a.[a]	n.a.
	undefined	1.4	1.5	1.2	2.6
cis	α_R	0	1	1	2
	δ	0	1	27	36
	β	66	33	1	2
	PPII	30	60	67	55
	ζ	3	4	1	5
	α_L	1	0	n.a.	n.a.
	undefined	0	1	3	0

[a] n.a. = not applicable. Pro is covalently restrained to negative values of ϕ and therefore cannot adopt the α_L conformation.

Table 4. Combinations of conformations observed at Ser-Pro and Ala-Pro sequences.[a] Populations of the most frequent combinations of conformations observed at Ser-Pro and Ala-Pro sequences in the PDB. Populations are represented as a percent of structures in each Pro amide conformation. See Table S4 for complete analysis of all combinations of conformations.

	conformation		SP	AP
	X	Pro	$\%_{total}$	$\%_{total}$
trans	α_R	α_R	4.2	12.8
	β	α_R	13.8	3.2
	β	δ	4.8	1.0
	β	PPII	12.2	18.0
	PPII	α_R	30.2	11.5
	PPII	δ	6.1	2.2
	PPII	PPII	15.9	34.3
cis	β	δ	9	3
	β	PPII	55	28
	PPII	δ	17	32
	PPII	PPII	9	23

[a] Populations of combinations of conformations that are either present at > 5% of the total population or that are different between SP and AP by greater than a factor of 1.5 are shown.

Table 5. Frequencies of C–H/O interactions at Ser-*cis*-Pro conformations in different type VI β-turns. Frequencies of C–H/O interactions observed at the most common conformations observed at Ser-*cis*-Pro structures. See Table S5 for complete analysis.

	conformation		β-turn type			
cis	Ser	Pro	classical	Dunbrack	$\%_{total}$	% C_α–H/O
	PPII	δ	VIa1	PcisD	17	65
	β	δ	VIa2	BcisD	9	70
	PPII	PPII	VIb	PcisP	9	55
	β	PPII	VIb	BcisP	55	31

Table 6. Frequencies of main-chain *i*/*i*+3 C=O_i•••H–N_{i+3} and *i*/*i*+4 C=O_i•••H–N_{i+4} hydrogen bonds at Ser-Pro and Ala-Pro sequences, with Ser/Ala as the *i*+1 residue (X_i–(Ser/Ala)-Pro–X_{i+3} register) or as the *i* residue (Ser$_i$/Ala$_i$–Pro-X–X_{i+3} register). Frequencies of C=O_i•••H–N_{i+3} and C=O_i•••H–N_{i+4} backbone hydrogen bonds in each register for Ser and Ala.

register	%$_{total}$ C=O_i ···H–N_{i+3}	%$_{total}$ C=O_i ···H–N_{i+4}	%$_{total}$ both[a]	%$_{total}$ any[b]
X–S-*trans*-P–X	3.6	3.1	2.0	4.7
X–A-*trans*-P–X	10.5	9.4	6.1	13.8
X–S-*cis*-P–X	18	8	2	24
X–A-*cis*-P–X	27	16	6	37
S–*trans*-P-X–X	42.9	25.2	16.9	51.2
A–*trans*-P-X–X	27.1	14.3	7.5	33.9
S–*cis*-P-X–X	0	0	0	0
A–*cis*-P-X–X	0	0	0	0

[a] Percent of structures with both C=O_i•••H–N_{i+3} and C=O_i•••H–N_{i+4} hydrogen bonds.
[b] Percent of structures with any C=O_i•••H–N_{i+3} or C=O_i•••H–N_{i+4} hydrogen bond, including structures with both hydrogen bonds.

Table 7. Most frequent hydrogen-bonded secondary structures present at Ser-*trans*-Pro and Ala-*trans*-Pro sequences in the PDB. Helix initiating structures were identified by a sequence of five or more consecutive residues adopting the α_R/δ conformation following the first residue of the sequence.

secondary structure	%$_{total}$ S–*trans*-P-X–X	%$_{total}$ A–*trans*-P-X–X
β-turn	25.2	18.0
3_{10}-helix or α-helix initiation	22.1	13.6
α-turn	3.9	2.3

Table 8. Most frequent β-turn types at S–*trans*-P-X–X and A–*trans*-P-X–X sequences. 36.8% of S–*trans*-P-X–X and 18.2% of A–*trans*-P-X–X sequences in the PDB were found in to be in β-turns (Table 7). Among the structures in the β-turn conformation, populations of the most common types of β-turn observed at P-*trans*-X sequences are indicated. See Table S10 for full analysis.

conformation		β-turn type		S–*trans*-P-X–X	A–*trans*-P-X–X
Pro	X	classical	Dunbrack	$\%_{total}$	$\%_{total}$
α_R	α_R/δ	I	AA/AD	85.5	60.2
PPII	α_L	II	Pa	4.9	27.1

Table 9. Frequencies of Ser$_i$ O$_\gamma$•••H–N$_{i+2}$ and Ser$_i$ O$_\gamma$•••H–N$_{i+3}$ side chain-main chain hydrogen bonds with Ser as the *i* residue in type I β-turns and as the N–capping residue of 3$_{10}$-helices and α-helices. Main-chain hydrogen bonded secondary structures in the PDB with Ser as the *i* residue and Pro as the *i*+1 residue that have Ser$_i$ O$_\gamma$•••H–N$_{i+2}$ and/or Ser$_i$ O$_\gamma$•••H–N$_{i+3}$ hydrogen bonds.

	Ser$_i$ O$_\gamma$ hydrogen bonds			
S–*trans*-P-X–X	% Ser$_i$ O$_\gamma$•••H–N$_{i+2}$	% Ser$_i$ O$_\gamma$•••H–N$_{i+3}$	% both	% any
type I β-turn	50.7	6.3	2.6	54.4
3$_{10}$-helix Ncap	39.8	31.4	3.4	67.8
α-helix Ncap	9.7	76.5	6.7	79.5

References

(1) Pundir, S.; Martin, M. J.; O'Donovan, C. UniProt Protein Knowledgebase. *Methods Mol. Biol.* **2017**, *1558*, 41-55.

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

(3) Reiersen, H.; Rees, A. R. The hunchback and its neighbours: proline as an environmental modulator. *Trends Biochem. Sci.* **2001**, *26*, 679-684.

(4) Mateos, B.; Conrad-Billroth, C.; Schiavina, M.; Beier, A.; Kontaxis, G.; Konrat, R.; Felli, I. C.; Pierattelli, R. The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators. *J. Mol. Biol.* **2020**, *432*, 3093-3111.

(5) Steinhilb, M. L.; Dias-Santagata, D.; Mulkearns, E. E.; Shulman, J. M.; Biernat, J.; Mandelkow, E. M.; Feany, M. B. S/P and T/P phosphorylation is critical for tau neurotoxicity in Drosophila. *J. Neurosci. Res.* **2007**, *85*, 1271-1278.

(6) Suzuki, M. SPXX, a Frequent Sequence Motif in Gene Regulatory Proteins. *J. Mol. Biol.* **1989**, *207*, 61-84.

(7) Eick, D.; Geyer, M. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem. Rev.* **2013**, *113*, 8456-8490.

(8) Jeronimo, C.; Bataille, A. R.; Robert, F. The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem. Rev.* **2013**, *113*, 8491-8522.

(9) Portz, B.; Lu, F. Y.; Gibbs, E. B.; Mayfield, J. E.; Mehaffey, M. R.; Zhang, Y. J.; Brodbelt, J. S.; Showalter, S. A.; Gilmour, D. S. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nature Commun.* **2017**, *8*, article 15231.

(10) Ubersax, J. A.; Ferrell, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 530-541.

(11) Pandey, A. K.; Ganguly, H. K.; Sinha, S. K.; Daniels, K. E.; Yap, G. P. A.; Patel, S.; Zondlo, N. J. An Inherent Structural Difference Between Serine and Threonine Phosphorylation: Phosphothreonine Prefers an Ordered, Compact, Cyclic Conformation. *ACS Chem. Biol.* **2023**, *18*, 1938-1958.

(12) Grathwohl, C.; Wüthrich, K. X-Pro Peptide Bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers* **1976**, *15*, 2025-2041.

(13) 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.

(14) 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.

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

(16) Ganguly, H. K.; Elbaum, M. B.; Zondlo, N. J. Serine-404 Phosphorylation and the R406W Modification in Tau Stabilize the cis-Proline Amide Bond, via Phosphoserine-Proline C–H/O and Proline-Aromatic C–H/π Interactions. *chemRxiv* **2023**, *DOI: 10.26434/chemrxiv-2023-w152b*.

(17) 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.

(18) Desiraju, G. R. The C–H...O Hydrogen Bond: Structural Implications and Supramolecular Design. *Acc. Chem. Res.* **1996**, *29*, 441-449.

(19) Gu, Y. L.; Kar, T.; Scheiner, S. Fundamental properties of the CH•••O interaction: Is it a true hydrogen bond? *J. Am. Chem. Soc.* **1999**, *121*, 9411-9422.

(20) Jones, C. R.; Baruah, P. K.; Thompson, A. L.; Scheiner, S.; Smith, M. D. Can a C–H...O Interaction Be a Determinant of Conformation? *J. Am. Chem. Soc.* **2012**, *134*, 12064-12071.

(21) Adhikari, U.; Scheiner, S. Magnitude and Mechanism of Charge Enhancement of CH center dot center dot O Hydrogen Bonds. *J. Phys. Chem. A* **2013**, *117*, 10551-10562.

(22) Chakrabarti, P.; Chakrabarti, S. C–H•••O hydrogen bond involving proline residues in alpha-helices. *J. Mol. Biol.* **1998**, *284*, 867-873.

(23) Horowitz, S.; Trievel, R. C. Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function. *J. Biol. Chem.* **2012**, *287*, 41576-41582.

(24) Derewenda, Z. S.; Lee, L.; Derewenda, U. The Occurence of C–H…O Hydrogen Bonds in Proteins. *J. Mol. Biol.* **1995**, *252*, 248-262.

(25) 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.

(26) Hutchinson, E. G.; Thornton, J. M. A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **1994**, *3*, 2207-2216.

(27) Suzuki, M.; Yagi, N. Structure of the Spxx Motif. *Proc. Royal Soc. B Biol. Sci.* **1991**, *246*, 231-235.

(28) Song, B. B.; Bomar, M. G.; Kibler, P.; Kodukula, K.; Galande, A. K. The Serine-Proline Turn: A Novel Hydrogen-Bonded Template for Designing Peptidomimetics. *Org. Lett.* **2012**, *14*, 732-735.

(29) Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J. Proline Editing: A General and Practical Approach to the Synthesis of Functionally and Structurally Diverse Peptides. Analysis of Steric versus Stereoelectronic Effects of 4-Substituted Prolines on Conformation within Peptides. *J. Am. Chem. Soc.* **2013**, *135*, 4333-4363.

(30) Forbes, C. R.; Pandey, A. K.; Ganguly, H. K.; Yap, G. P. A.; Zondlo, N. J. 4R- and 4S-Iodophenyl Hydroxyproline, 4R-Pentynoyl Hydroxyproline, and S-Propargyl-4-Thiolphenylalanine: Conformationally Biased and Tunable Amino Acids for Bioorthogonal Reactions. *Org. Biomol. Chem.* **2016**, *14*, 2327-2346.

(31) Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. Pyrrolidine ring puckering in cis and trans-proline residues in proteins and polypeptides: Different puckers are favoured in certain situations. *J. Mol. Biol.* **1992**, *228*, 725-734.

(32) Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Sci.* **2001**, *10*, 2627-2632.

(33) Kang, Y. K.; Choi, H. Y. Cis-trans isomerization and puckering of proline residue. *Biophys. Chem.* **2004**, *111*, 135-142.

(34) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56*, 1363-1388.

(35) 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.

(36) Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981**, *34*, 167-339.

(37) Wilmot, C. M.; Thornton, J. M. Beta-Turns and Their Distortions - a Proposed New Nomenclature. *Protein Eng.* **1990**, *3*, 479-493.

(38) Rose, G. D.; Gierasch, L. M.; Smith, J. A. Turns in peptide and proteins. *Adv. Protein Chem.* **1985**, *37*, 1-109.

(39) Kleiger, G.; Grothe, R.; Mallick, P.; Eisenberg, D. GXXXG and AXXXA: Common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* **2002**, *41*, 5990-5997.

(40) Senes, A.; Engel, D. E.; DeGrado, W. F. Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr. Opin. Struct. Biol.* **2004**, *14*, 465-479.

(41) Mueller, B. K.; Subramaniam, S.; Senes, A. A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical C alpha-H hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E888-E895.

(42) Teese, M. G.; Langosch, D. Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **2015**, *54*, 5125-5135.

(43) Yesselman, J. D.; Horowitz, S.; Brooks, C. L.; Trievel, R. C. Frequent side chain methyl carbon-oxygen hydrogen bonding in proteins revealed by computational and stereochemical analysis of neutron structures. *Proteins Struct. Funct. Bioinform.* **2015**, *83*, 403-410.

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

(45) Karplus, P. A. Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Sci.* **1996**, *5*, 1406-1420.

(46) Ho, B. K.; Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **2005**, *5*.

(47) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. n->pi* interactions in proteins. *Nat. Chem. Biol.* **2010**, *6*, 615-620.

(48) Newberry, R. W.; Raines, R. T. The n->pi* interaction. *Acc. Chem. Res.* **2017**, *50*, 1838-1846.

(49) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. Host-Guest Scale of Left-Handed Polyproline II Helix Formation. *Proteins* **2003**, *53*, 68-75.

(50) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with b-sheet scales. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17964-17968.

(51) 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.

(52) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J. OGlcNAcylation and Phosphorylation Have Opposing Structural Effects in tau: Phosphothreonine Induces Particular Conformational Order. *J. Am. Chem. Soc.* **2014**, *136*, 3803-3816.

(53) Elbaum, M. B.; Zondlo, N. J. OGlcNAcylation and Phosphorylation Have Similar Structural Effects in α-Helices: Post-Translational Modifications as Inducible Start and Stop Signals in α-Helices, with Greater Structural Effects on Threonine Modification. *Biochemistry* **2014**, *53*, 2242-2260.

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

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

(56) Newberry, R. W.; Raines, R. T. A prevalent intraresidue hydrogen bond stabilizes proteins. *Nat. Chem. Biol.* **2016**, *12*, 1084-1088.

(57) Newberry, R. W.; Orke, S. J.; Raines, R. T. n ->pi* Interactions Are Competitive with Hydrogen Bonds. *Org. Lett.* **2016**, *18*, 3614-3617.

(58) Zondlo, N. J. Solvation stabilizes intercarbonyl n→π* interactions and polyproline II helix. *Phys. Chem. Chem. Phys.* **2022**, *24*, 13571-13586.

(59) Adhikary, R.; Zimmermann, J.; Liu, J.; Forrest, R. P.; Janicki, T. D.; Dawson, P. E.; Corcelli, S. A.; Romesberg, F. E. Evidence of an Unusual N-H center dot center dot center dot N Hydrogen Bond in Proteins. *J. Am. Chem. Soc.* **2014**, *136*, 13474-13477.

(60) Deepak, R. N. V. K.; Sankaramakrishnan, R. Unconventional N-H•••N Hydrogen Bonds Involving Proline Backbone Nitrogen in Protein Structures. *Biophys. J.* **2016**, *110*, 1967-1979.

(61) Alderson, T. R.; Lee, J. H.; Charlier, C.; Ying, J. F.; Bax, A. Propensity for cis-Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, *19*, 37-42.

(62) Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. Amino Acids with Hydrogen-Bonding Side Chains have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution. *Chem. Eur. J.* **2011**, *17*, 6789-6797.

(63) Vijayakumar, M.; Qian, H.; Zhou, H. X. Hydrogen bonds between short polar side chains and peptide backbone: Prevalence in proteins and effects on helix-forming propensities. *Proteins* **1999**, *34*, 497-507.

(64) Eswar, N.; Ramakrishnan, C. Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Eng.* **2000**, *13*, 227-238.

(65) O'Neil, K. T.; Degrado, W. F. A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino-Acids. *Science* **1990**, *250*, 646-651.

(66) Pace, C. N.; Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **1998**, *75*, 422-427.

(67) Richardson, J. S.; Richardson, D. C. Amino-Acid Preferences for Specific Locations at the Ends of Alpha-Helices. *Science* **1988**, *240*, 1648-1652.

(68) Doig, A. J.; MacArthur, M. W.; Stapley, B. J.; Thornton, J. M. Structures of N-termini of helices in proteins. *Protein Sci.* **1997**, *6*, 147-155.

(69) Aurora, R.; Rose, G. D. Helix capping. *Protein Sci.* **1998**, *7*, 21-38.

(70) Serrano, L.; Fersht, A. R. Capping and Alpha-Helix Stability. *Nature* **1989**, *342*, 296-299.

(71) Lyu, P. C.; Wemmer, D. E.; Zhou, H. X.; Pinker, R. J.; Kallenbach, N. R. Capping Interactions in Isolated Alpha-Helices - Position-Dependent Substitution Effects and Structure of a Serine-Capped Peptide Helix. *Biochemistry* **1993**, *32*, 421-425.

(72) Doig, A. J.; Baldwin, R. L. N- and C-Capping Preferences for All 20 Amino-Acids in Alpha-Helical Peptides. *Protein Sci.* **1995**, *4*, 1325-1336.

(73) Leader, D. P.; Milner-White, E. J. The structure of the ends of α-helices in globular proteins: Effect of additional hydrogen bonds and implications for helix formation. *Proteins Struct. Funct. Bioinform.* **2011**, *79*, 1010-1019.

(74) Presta, L. G.; Rose, G. D. Helix Signals in Proteins. *Science* **1988**, *240*, 1632-1641.

(75) Barlow, D. J.; Thornton, J. M. Helix Geometry in Proteins. *J. Mol. Biol.* **1988**, *201*, 601-619.

(76) Cochran, D. A. E.; Penel, S.; Doig, A. J. Effect of the N1 residue on the stability of the alpha-helix for all 20 amino acids. *Protein Sci.* **2001**, *10*, 463-470.

(77) Serrano, L.; Neira, J. L.; Sancho, J.; Fersht, A. R. Effect of Alanine Versus Glycine in Alpha-Helices on Protein Stability. *Nature* **1992**, *356*, 453-455.

(78) Bhatt, M. R.; Ganguly, H. K.; Zondlo, N. J. Acyl capping group identity effects on α-helicity: on the importance of amide•water hydrogen bonds to α-helix stability. *Biochemistry* **2024**, *63, DOI: 10.1021/acs.biochem.3c00646* accepted.

(79) Karle, I. L.; Balaram, P. Structural characteristics of alpha-helical peptides containing Aib residues. *Biochemistry* **1990**, *29*, 6747-6756.

(80) Millhauser, G. L. Views of helical peptides: a proposal for the position of 3(10) helix along the thermodynamic folding pathway. *Biochemistry* **1995**, *34*, 3873-3877.

(81) Armen, R.; Alonso, D. O. V.; Daggett, V. The role of alpha-, 310-, and pi-helix in helix->coil transitions. *Protein Sci.* **2003**, *12*, 1145-1157.

(82) Tirado-Rives, J.; Jorgensen, W. L. Molecular dynamics simulations of the unfolding of an a-helical analog of ribonuclease-A S-peptide in water. *Biochemistry* **1991**, *30*, 3864-3871.

(83) Fischer, G. Chemical aspects of peptide bond isomerisation. *Chem. Soc. Rev.* **2000**, *29*, 119-127.

(84) Andreotti, A. H. Native State Proline Isomerization: An Intrinsic Molecular Switch. *Biochemistry* **2003**, *42*, 9515-9524.

(85) Schmidpeter, P. A. M.; Koch, J. R.; Schmid, F. X. Control of protein function by prolyl isomerization. *Biochim. Biophys. Acta Gen. Subj.* **2015**, *1850*, 1973-1982.

(86) Corden, J. L. RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. *Chem. Rev.* **2013**, *113*, 8423-8455.

(87) Kumaki, Y.; Matsushima, N.; Yoshida, H.; Nitta, K.; Hikichi, K. Structure of the YSPTSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2001**, *1548*, 81-93.

(88) Bomar, M. G.; Raghavender, U. S.; Spindel, A. W. I.; Kodukula, K.; Galande, A. K. The ST pinch: A side chain-to-side chain hydrogen-bonded motif. *Proteins Struct. Funct. Bioinform.* **2012**, *80*, 1259-1263.

(89) Aubry, A.; Marraud, M. Serine and Beta-Turns. *Biopolymers* **1983**, *22*, 341-345.

(90) Werner-Allen, J. W.; Lee, C. J.; Liu, P.; Nicely, N. I.; Wang, S.; Greenleaf, A. L.; Zhou, P. cis-Proline-mediated Ser(P)(5) Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72. *J. Biol. Chem.* **2011**, *286*, 5717-5726.

(91) 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*, article 15233.

(92) Amith, W. D.; Dutagaci, B. Complex Conformational Space of the RNA Polymerase II C-Terminal Domain upon Phosphorylation. *J. Phys. Chem. B* **2023**, *127*, 9223-9235.

(93) Lushpinskaia, I. P.; Flores-Solis, D.; Zweckstetter, M. Structure and phase separation of the C-terminal domain of RNA polymerase II. *Biol. Chem.* **2023**, *404*, 839-844.

(94) Pettit, G. R.; Srirangam, J. K.; Herald, D. L.; Xu, J. P.; Boyd, M. R.; Cichacz, Z.; Kamano, Y.; Schmidt, J. M.; Erickson, K. L. Isolation and crystal structure of stylopeptide 1, a new marine Porifera cycloheptapeptide. *J. Org. Chem.* **1995**, *60*, 8257-8261.

(95) Nepal, B.; Scheiner, S. Effect of Ionic Charge on the CH center dot center dot center dot pi Hydrogen Bond. *J. Phys. Chem. A* **2014**, *118*, 9575-9587.

(96) Bulusu, G.; Desiraju, G. R. Strong and Weak Hydrogen Bonds in Protein-Ligand Recognition. *J. Indian Inst. Sci.* **2020**, *100*, 31-41.

(97) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. Non-proline cis peptide bonds in proteins. *J. Mol. Biol.* **1999**, *286*, 291-304.

(98) Loos, P. F.; Assfeld, X.; Rivail, J. L. Intramolecular interactions and cis peptidic bonds. *Theo. Chem. Acc.* **2007**, *118*, 165-171.

(99) Mathieu, S.; Poteau, R.; Trinquier, G. Estimating the "Steric clash" at cis peptide bonds. *J. Phys. Chem. B* **2008**, *112*, 7894-7902.

(100) Scherer, G.; Kramer, M. L.; Schutkowski, M.; Reimer, U.; Fischer, G. Barriers to rotation of secondary amide peptide bonds. *J. Am. Chem. Soc.* **1998**, *120*, 5568-5574.

(101) Nguyen, K.; Iskandar, M.; Rabenstein, D. L. Kinetics and Equilibria of Cis/Trans Isomerization of Secondary Amide Peptide Bonds in Linear and Cyclic Peptides. *J. Phys. Chem. B* **2010**, *114*, 3387-3392.

(102) Schiene-Fischer, C.; Fischer, G. Direct measurement indicates a slow cis/trans isomerization at the secondary amide peptide bond of glycylglycine. *J. Am. Chem. Soc.* **2001**, *123*, 6227-6231.

(103) Pappenberger, G.; Aygun, H.; Engels, J. W.; Reimer, U.; Fischer, G.; Kiefhaber, T. Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat. Struct. Biol.* **2001**, *8*, 452-458.

(104) Zhang, J.; Germann, M. W. Characterization of Secondary Amide Peptide Bond Isomerization: Thermodynamics and Kinetics from 2D NMR Spectroscopy. *Biopolymers* **2011**, *95*, 755-762.

(105) Schiene-Fischer, C.; Habazettl, J.; Schmid, F. X.; Fischer, G. The hsp70 chaperone DnaK is a secondary amide peptide bond cis-trans isomerase. *Nat. Struct. Biol.* **2002**, *9*, 419-424.

(106) Svensson, A. K. E.; O'Neill, J. C.; Matthews, C. R. The coordination of the isomerization of a conserved non-prolyl cis peptide bond with the rate-limiting steps in the folding of dihydrofolate reductase. *J. Mol. Biol.* **2003**, *326*, 569-583.

(107) Guan, R. J.; Xiang, Y.; He, X. L.; Wang, C. G.; Wang, M.; Zhang, Y.; Sundberg, E. J.; Wang, D. C. Structural mechanism governing Cis and Trans isomeric states and an intramolecular switch for Cis/Trans isomerization of a non-proline peptide bond observed in crystal structures of scorpion toxins. *J. Mol. Biol.* **2004**, *341*, 1189-1204.

(108) He, X. L.; Guan, R. J.; Zhang, Y.; Wang, D. C. A dimeric structure of the scorpion toxin BmK M1 at 1.4 Å resolution:: Non-proline cis peptide bond and its inntrinsic structural elements. *Prog. Biochem. Biophys.* **2006**, *33*, 231-240.

(109) Mineev, K. S.; Chernykh, M. A.; Motov, V. V.; Prudnikova, D. A.; Pavlenko, D. M.; Kuzmenkov, A. I.; Peigneur, S.; Tytgat, J.; Vassilevski, A. A. A scorpion toxin affecting sodium channels shows double cis-trans isomerism. *FEBS Lett.* **2023**, *597*, 2358-2368.

(110) Herzberg, O.; Moult, J. Analysis of the Steric Strain in the Polypeptide Backbone of Protein Molecules. *Proteins* **1991**, *11*, 223-229.

(111) Exarchos, K. P.; Exarchos, T. P.; Rigas, G.; Papaloukas, C.; Fotiadis, D. I. Extraction of consensus protein patterns in regions containing non-proline cis peptide bonds and their functional assessment. *BMC Bioinform.* **2011**, *12*, article 142.

(112) Vos, S.; deJersey, J.; Martin, J. L. Crystal structure of Escherichia coli xanthine phosphoribosyltransferase. *Biochemistry* **1997**, *36*, 4125-4134.

(113) Vos, S.; Parry, R. J.; Burns, M. R.; de Jersey, J.; Martin, J. L. Structures of free and complexed forms of Escherichia coli xanthine-guanine phosphoribosyltransferase. *J. Mol. Biol.* **1998**, *282*, 875-889.

(114) 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.

(115) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(116) Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations. 1. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.

(117) 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.

(118) Raghavachari, K.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 20. Basis sets for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650-654.

Graphical Table of Contents