Secondary Forces in Protein Folding

Robert W. Newberry[†,∥] and Ronald T. Raines[*,†,‡,§]

[†]Department of Chemistry and [‡]Department of Biochemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States

[§]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

ABSTRACT: A complete inventory of the forces governing protein folding is critical for productive protein modeling, including structure prediction and *de novo* design, as well as understanding protein misfolding diseases of clinical significance. The dominant contributors to protein folding include the hydrophobic effect and conventional hydrogen bonding, along with Coulombic and van der Waals interactions. Over the past few decades, important additional contributors have been identified, including C−H···O hydrogen bonding, $n{\rightarrow}\pi^*$ interactions, C5 hydrogen bonding, chalcogen bonding, and interactions involving aromatic rings (cation−π, X−H···π, π−π, anion−π, and sulfur−arene). These secondary contributions fall into two general classes: (1) weak but abundant interactions of the protein main chain and (2) strong but less frequent interactions involving protein side chains. Though interactions with high individual energies play important roles in specifying nonlocal molecular contacts and ligand binding, we estimate that weak but abundant interactions are likely to make greater overall contributions to protein folding, particularly at the level of secondary structure. Further research is likely to illuminate additional roles of these noncanonical interactions and could also reveal contributions yet unknown.
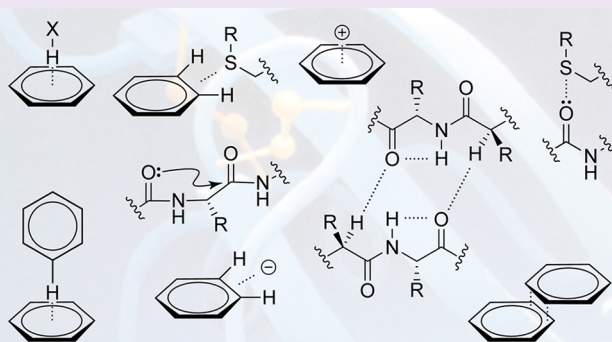
Proteins are the principal molecular machines of the cell, capable of myriad activities that enable life. Each individual protein derives its function from the unique three-dimensional arrangement of its chemical components.[1] Seminal experiments by Anfinsen[2] demonstrated that amino acid sequences can contain all of the chemical information necessary to specify a particular stable structure.[3] Decoding the chemical information present within the polypeptide chain should therefore allow one to predict the structure from its sequence alone. In view of the number of putative protein sequences generated from DNA sequencing data,[4] such technology would prove invaluable for addressing countless issues in biology. Moreover, the same insight could eventually allow physicians to predict the effects of particular mutations, empowering personalized medicine; this prospect could be especially important for the treatment of diseases caused by protein misfolding.[5] Finally, a complete understanding of the factors governing protein structure could be leveraged toward the design of new proteins with emergent functions,[6,7] the limits of which are hard to conceive. Understanding the molecular basis for protein structure has thus become one of the central scientific challenges of our age.

■ CANONICAL FORCES IN PROTEIN FOLDING

Under physiological conditions, the free energy of the folded state of a typical globular protein is 10−15 kcal/mol less than that of the unfolded state.[8,9] The enthalpic and entropic energy differences between the two states are, however, several-fold larger.[10,11] The ensuing dichotomy in the Gibbs equation—$\Delta G°$ is a small difference between large values of $\Delta H°$ and $T\Delta S°$—underlies the need for a comprehensive inventory of the forces that govern protein folding.

The interactions that stabilize protein structure also guide a polypeptide chain in attaining that structure.[12] The folding of polypeptide chains is well-known to be encouraged by a handful of non-covalent interactions: the hydrophobic effect, conventional hydrogen bonding, Coulombic interactions, and van der Waals interactions.[8,13] A detailed understanding of these canonical forces has led to the development of many important technologies, including force fields for molecular dynamics simulations[14−16] and automated methods for protein design,[17,18] which in turn have yielded exciting results.[19,20] To assess the limitations in these methods, the biophysics community has engaged in a recurring systematic evaluation called the Critical Assessment of Structure Prediction (CASP).[21] The premise of this biannual competition is simple: given only the target sequence, computational biophysicists attempt to predict the three-dimensional structures of proteins that have been determined recently by experimental structural biology. The contest has been held 12 times since 1994 and has led to many insights.[21] Though significant improvements were achieved during the early years of the CASP competition,
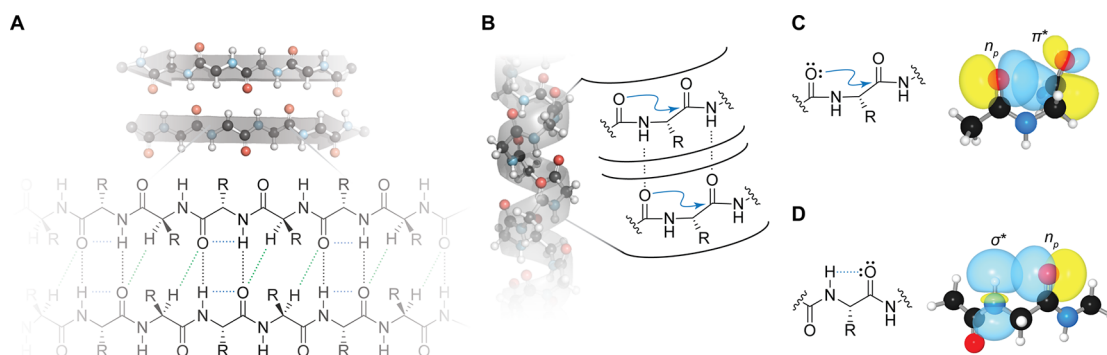
Figure 1. Secondary interactions involving the main chain. (A) Structural model of an idealized β-sheet, showing conventional main-chain hydrogen bonds (black dashes), C—H···O hydrogen bonds (green dashes), and C5 hydrogen bonds (blue dashes). (B) Structural model of an idealized α-helix, showing main-chain hydrogen bonds (black dashes) and n→π* interactions (blue arrows). (C, D) Orbital overlaps that underlie the formation of (C) n→π* interactions and (D) C5 hydrogen bonds.

progress of late has been slow.[22] In particular, the accuracy of models for sequences bearing low identity to proteins of known structure remains poor.[23] In the absence of homologous proteins or domains on which to base initial models, structure prediction relies increasingly on molecular mechanics approaches, which have been problematic.[24] Similar limitations have been noted in the prediction of organic crystal structures,[25,26] a problem that is conceptually similar to that of protein structure prediction in requiring an accurate inventory of relevant forces. (In addition, efficient packing is essential for both protein folding[27] and crystal growth,[28] as the interior of proteins has long been known to be closer to a solid than a liquid.[29])

The ongoing challenges in protein structure prediction and design are suggestive of an incomplete understanding of the forces that govern protein folding, structure, and stability.[24] Apparently, an understanding of the canonical forces alone is insufficient for a proper description of protein biophysics. To address this problem, researchers have identified a suite of additional interactions that also contribute to the enthalpy of protein folding. We review the manifestations and contributions of those secondary interactions herein.

■ SECONDARY INTERACTIONS OF THE MAIN CHAIN

The most ubiquitous secondary interactions are those that occur between backbone atoms (Figure 1). These interactions enlist the lone pairs of the main-chain oxygens. In analogy to the hydrogen bonds donated by the main-chain nitrogen,[30] nearly all of the lone pairs of the main-chain oxygens are engaged in intramolecular interactions.[31]

C—H···O Hydrogen Bonds. The first secondary interactions identified were noncanonical hydrogen bonds involving carbon-based donors.[32] Though proteins typically feature weaker carbon acids than do some other organic molecules, there are some protons that are sufficiently acidic to engage in hydrogen bonding. For example, there is substantial evidence that histidine side chains can donate hydrogen bonds from C^ε.[33] By far, though, the most common C—H···O hydrogen bond donors are the C^α—H protons of the main chain.

C—H···O hydrogen bonds have been observed widely in crystal structures of small molecules and in the 1960s were proposed to contribute to protein stability.[34,35] Despite early debate, C—H···O hydrogen bonds are now well-accepted.[36,37] They share many properties with canonical hydrogen bonds, such as directionality and cooperativity,[38] though they notably

induce blue shifts in vibrational spectra.[39,40] Like canonical hydrogen bonds, they are predominantly electrostatic interactions, with smaller contributions from van der Waals attraction and charge transfer.[38] Experimental characterizations of their energy within peptides or proteins remain scarce[41,42] because of their small energies and the experimental challenge of probing the backbone. Calculations generally point to energies of 1−2 kcal/mol,[43,44] which is approximately half that of canonical hydrogen bonds.[45] Nevertheless, detailed analysis of the geometry of intermolecular contacts in proteins has shown a substantial propensity for carbon-based acids to engage with hydrogen-bond acceptors. Like other hydrogen bonds,[46] these interactions are identified by a short donor—acceptor distance (typically <2.5 Å) and relative linearity of the donor, acceptor, and their antecedents.

The most prevalent example of C—H···O hydrogen bonds in proteins is the interstrand C^α—H···O=C hydrogen bond in β-sheets (Figure 1A).[47,48] These C—H···O contacts occur at distances that are significantly shorter than expected for repulsive van der Waals interactions and correspond closely to those observed in small-molecule crystal structures with validated interactions. Moreover, the approach of the donor to the carbonyl acceptor occurs largely within the plane of the peptide bond, where the carbonyl electron density is maximal. These interactions could affect some 35% of residues in β-sheets.[31]

Other, albeit less frequent, examples of C—H···O hydrogen bonds occur in proteins. Contacts with carbonyl oxygens in the backbone of the α-helix have been noted,[49] though these interactions usually involve less acidic β-protons, so the energy contributed by such contacts is likely modest. Additionally, α-helices might benefit from C—H···O hydrogen bonds involving donation of a proline α-proton to carbonyl acceptors,[50] which has the potential to attenuate the strong helix-breaking tendencies of proline residues. Backbone C—H···O hydrogen bonds are also a feature of the collagen triple helix[51] and might contribute to binding energy and discrimination at protein—protein interfaces.[52] One notable example of the latter occurs between transmembrane helices. Transmembrane helices, whether within individual proteins or at interfaces within complexes of multiple proteins, often contact one another along ridges of small amino acids, typified by the GXXXG motif.[53] These contacts are mediated by multiple C—H···O hydrogen bonds between helices[54] and lead to a characteristic interaction geometry that has been termed the GAS-right motif.[55]

n→π* **Interactions.** A distinct interaction within the backbone has been posited to contribute to protein folding: the $n{\rightarrow}\pi^*$ interaction.[56,57] These weak interactions occur between adjacent carbonyl groups in the backbone due to donation of lone-pair (*n*) electron density from a carbonyl oxygen into the π^* orbital of another carbonyl group (Figure 1B,C). Originally invoked to explain the correlation of pyrrolidine ring pucker with the *cis*−*trans* conformation of prolyl peptide bonds in collagen,[58] this interaction has now been recognized in a variety of systems.[59−61] A signature of an $n{\rightarrow}\pi^*$ interaction is a sub-van der Waals contact of the donor oxygen on the acceptor carbon along the Bürgi−Dunitz trajectory.[62,63] Though these interactions were posited to be a particular example of a dipolar interaction,[64−66] extensive evidence[67−69] indicates that these interactions harbor distinct charge-transfer character, a view that has gained acceptance.[70] The ensuing electronic donation can pyramidalize the acceptor carbonyl group, as has been observed in high-resolution protein crystal structures.[71,72] Computational[69] and experimental[56] studies on small molecules have estimated the energy of a typical $n{\rightarrow}\pi^*$ interaction to be between 0.3 and 0.7 kcal/mol; experimental measurements of the energy of an $n{\rightarrow}\pi^*$ interaction have yet to be achieved in a folded protein. Nevertheless, prototypical $n{\rightarrow}\pi^*$ interactions are strong enough to compete with canonical hydrogen bonds.[73] Moreover, polarizing an acceptor carbonyl group with a hydrogen-bond acceptor increases the interaction energy of an $n{\rightarrow}\pi^*$ interaction.[74,75] Despite their modest energy, $n{\rightarrow}\pi^*$ interactions are predicted to contribute significantly to protein stability because of their frequency: a third of all residues in folded proteins are poised to engage in $n{\rightarrow}\pi^*$ interactions.[57] Moreover, they contribute differentially to secondary structure formation: >70% of residues in α-helices are predicted to engage in an $n{\rightarrow}\pi^*$ interaction, but <10% of β-sheet residues are predicted to do so.[57] This interaction is implicated in stabilizing not only α-helices[76] but also other helical conformations such as 3_{10}[57] and PPII geometries.[77,78] Additional interactions are possible in amino acid side chains.[79−81]

C5 Hydrogen Bonds. An analogous interaction has been identified in β-sheets.[82,83] Specifically, amide protons in β-strands can donate an intraresidue hydrogen bond to its own carbonyl oxygen, forming a C5 hydrogen bond (Figure 1A,D). These interactions become significant at donor−acceptor distances below 2.5 Å. Despite their nonlinear geometry, they bear the hallmarks of traditional hydrogen bonding, and their perturbation also causes predictable changes in the stability of β-sheets. Though calculations suggest that these interactions are significantly weaker than traditional hydrogen bonds, often affording only around 0.25 kcal/mol, nearly 5% of residues in folded proteins are affected by such interactions, making the C5 hydrogen bond a key contributor to protein structure and stability. Moreover, bioinformatic[83] and crystallographic[84] analyses indicate that C5 hydrogen bonds likely contribute to amyloid formation, which is implicated in many neurodegenerative diseases.[5]

■ SECONDARY INTERACTIONS INVOLVING SIDE CHAINS

Many secondary interactions engage side-chain atoms (Figure 2). Among such secondary interactions, those involving the aromatic rings of phenylalanine, tyrosine, and tryptophan residues are most important.[85] The unique electron distribu-
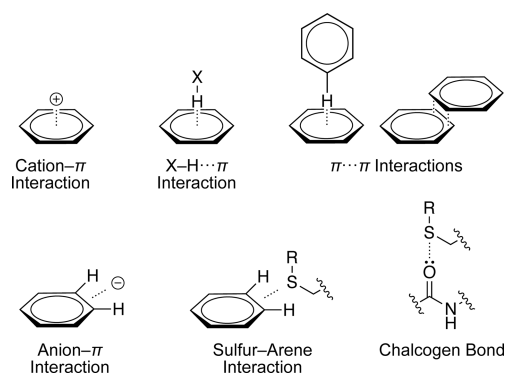

Figure 2. Secondary interactions involving side chains.

tion in these side chains enables a number of possible interactions. The facial π cloud bears significant partial negative charge and is nucleophilic, whereas the ring edge bears a partial positive charge and is electrophilic. This charge deposition also creates a permanent electric quadrupole that forms strong electrostatic interactions with both cations and anions. Indeed, perhaps the most important single example of noncanonical forces in protein folding is the cation−π interaction.

Cation−π Interactions. The significance of the cation−π interaction derives not only from its specific roles but also from its energy, which is distinctly larger than that of other secondary interactions in proteins: individual cation−π interactions can contribute 2−5 kcal/mol to the binding of ligands to their receptors.[86] Originally articulated in the supramolecular chemistry of organic cations,[87] these interactions are largely electrostatic attractions between electric monopoles and the electronically negative surfaces of aromatic rings, along with their corresponding quadrupole moments.[88] Other contributions to the cation−π interaction, such as dispersion and charge transfer, exist and can be important, but predictions of binding affinity based on electrostatics alone are usually quite successful.[89,90] Importantly, this energy is often sufficient to overcome the desolvation penalty for the binding of ions to protein pockets or cavities.[91] Indeed, binding sites for a variety of organic cations feature an abundance of aromatic residues, making the cation−π interaction key for the recognition of small-molecule ligands[86] or post-translational modifications on proteins such as histones.[92]

In addition to these critical functional roles, the abundance of aromatic and cationic residues in protein side chains presents an important opportunity for their contribution to protein folding. Cation−π interactions have been observed to perturb the pK_a values of functional groups in proteins,[93] demonstrating their influence unambiguously. Even under stringent criteria for identification, cation−π interactions affect approximately one in every 77 amino acid residues; using similar criteria, canonical salt bridges are 2-fold more common.[94] As interactions between side chains, both cation−π interactions and salt bridges are much less abundant than are secondary interactions of the main chain. For example, whereas ~26% of tryptophan residues are engaged in a cation−π interaction,[94] the frequency of tryptophan in eukaryotic proteins is only $(1.2 \pm 0.2)\%$,[95] diminishing the overall impact.

Arginine forms cation−π interactions more often than does lysine.[94] This preference is not due to intrinsic interaction energies but likely arises from the ability of arginine to participate in additional intermolecular interactions while

proximal to an aromatic ring. Aromatic residues in proteins form cation−π interactions with a relative frequency of tryptophan more often than tyrosine more often than phenylalanine, which parallels the intrinsic interaction energies of the side chains. Analysis of cation−π distributions across structural motifs is less advanced, but studies of designed peptides indicate that cation−π interactions can make significant contributions to secondary structure.[96,97] In addition, cation−π interactions are a common feature of protein−protein interfaces.[98] Though the contributions of individual cation−π interactions to ligand binding can lead to stabilization by several kilocalories per mole,[86] experimental measurements of contributions to protein stability provide somewhat lower values, suggesting that the energies of single cation−π interactions in peptides and proteins are generally in the range of 0.5−1.0 kcal/mol.[85,99] Nonetheless, this energy has been sufficient for the design of model proteins stabilized by cation−π interactions.[100]

X−H⋯π Interactions. Unsurprisingly, given their affinity for cations, aromatic rings are additionally capable of accepting hydrogen bonds[101,102] even from weaker carbon-based acids.[103−105] Like cation−π interactions, these X−H⋯π interactions appear to be especially relevant for ligand binding, where they contribute especially to carbohydrate recognition;[106−108] the binding sites of lectins are often enriched in aromatic residues, particularly tryptophan,[109] which can direct the binding mode of the carbohydrate with exquisite specificity via C−H⋯π interactions.[107,108] Within protein structure, hydrogen bonds to aromatic acceptors are identified by short contact of the donor heavy atom with the center of the aromatic ring at a steep angle of elevation to the plane of the ring.[102] By these criteria, they appear sufficiently common to contribute to protein folding; surveys estimate that approximately 10% of aromatic residues accept hydrogen bonds from nitrogen, oxygen, or sulfur donors.[102] In addition, interactions of aromatic rings with backbone donors, although relatively infrequent,[102] could demarcate changes in the secondary structure pattern and stabilize structural termini. Nonetheless, these interactions are uniformly weaker than cation−π interactions, commensurate with the reduction in electrostatic attraction. Gas-phase studies indicate that energies of these interactions are approximately 5−10 times weaker than for analogous cations; for example, the interaction energy for ammonia and benzene in the gas phase is 1.8 kcal/mol,[85] whereas that for ammonium and benzene is 18 kcal/mol.[88] Nevertheless, tuning the strength of an individual C−H⋯π interaction is sufficient to modulate the thermostability of designed miniproteins.[110]

π−π Interactions. The unique electron distribution in aromatic rings also allows them to interact favorably with one another. Inspired by the high aromatic contact of protein interiors, an early survey of aromatic−aromatic interactions in proteins found not only that such pairs were more common than expected by chance but also that particular short contact distances were favored strongly, suggesting an attraction.[111] Specific attraction between aromatic rings was first suggested by the dominance of enthalpy in the interaction, ruling out the previously hypothesized solvophobic nature.[85] Extensive characterization has shown that aromatic rings interact primarily in two geometries: T-shaped (or edge-to-face) and displaced-stacked (or offset-stacked).[85,111,112] Both arrangements are well-described by a balance between contributions from dispersion[85] and electrostatics.[113] Geometries observed

in crystalline or gas-phase arenes are recapitulated in proteins, though the exact preference of aromatic residues for different geometries differs among estimates,[114] possibly because of the expanding number of known protein structures.[115] Regardless, these contacts can clearly offer stability to proteins. Interestingly, thermophilic proteins have significantly more aromatic−aromatic contacts than do homologues from mesophilic organisms, consistent with a contribution to thermostability.[116] Experimental characterizations of individual aromatic−aromatic interactions in peptide and protein model systems have estimated the energy of a single interaction to be approximately 0.5−1.5 kcal/mol, with smaller values being observed for solvent-exposed residues in peptides[117,118] and larger values for residues in proteins.[119] These experimental energies generally agree well with those from calculations.[120,121] Importantly, over half of aromatic residues in proteins have been predicted to engage in attractive interactions on the basis of the simple criterion of an intercentroid distance of less than 7 Å.[111]

Anion−π Interactions. Aromatic rings can also interact with anions at their edges, which bear partial positive charge.[122] Indeed, initial analyses revealed that carboxylates contact aromatic rings in proteins more frequently than would be expected by chance, and the approach is predominantly edge-to-edge.[123] Energy deconvolution indicates that these interactions are likely dominated by electrostatics rather than van der Waals interactions. Experimental measurements of the energy of a single anion−π interaction suggest that such interactions contribute approximately 0.5 kcal/mol.[124] Detection of anion−π interactions in proteins is complicated by the number of degrees of freedom between the interacting partners. Indeed, both energy calculations and Boltzmann statistics find a differential potential of ring atoms to engage in anion−π interactions.[125] Using an energy-based criterion for identification, anion−π interactions involving phenylalanine have been observed in approximately 70% of proteins.[126] Using geometric identification criteria, tryptophan was found to have a higher propensity than phenylalanine or tyrosine to engage in anion−π interactions, possibly because of its size or dipole moment. Most anion−π interactions are distant in sequence, though local contacts in both α-helices and β-sheets are known.[126] Preliminary analyses have also catalogued the coincidence and cooperativity of anion−π interactions with hydrogen bonds,[125] π−π interactions,[127] and cation−π interactions.[127] Anion−π interactions might contribute to the formation of protein−protein interfaces[125] and are likely to be especially important in protein−DNA interactions[127] because DNA features both additional anionic phosphoryl groups and electron-deficient π systems that encourage the approach of anionic amino acid residues. An especially strong anion−π interaction has been reported to stabilize a β-hairpin in the WW domain.[128]

Sulfur−Arene Interactions. Aromatic rings are additionally capable of interacting with lone pairs, though these contacts generally involve electron-deficient rings, unlike those in proteins.[129] Nevertheless, reports have documented an enrichment of sulfur atoms near aromatic rings in proteins[130,131] and protein−protein interfaces,[131] leading to postulation of a so-called sulfur−arene interaction.[85] Early reports estimate that half of sulfur-containing residues form short contacts with aromatic rings.[130] As in small-molecule crystal structures,[132] sulfur atoms in proteins approach aromatic rings along the ring edge, though detailed geometries

have not been catalogued for each type of residue (cysteine, cystine, and methionine). Experimental perturbations of this interaction in peptides have found stabilizing energies on the order of 0.5 kcal/mol.[133,134] Although results from peptide studies suggest that these interactions are dominated by the hydrophobic effect rather than by a specific attraction,[134] experiments in proteins suggest that sulfur−arene interactions cannot be replaced by purely hydrophobic interactions.[131] Further research is needed to clarify the thermodynamic contributions of sulfur−arene interactions to protein folding.

Chalcogen Bonding. Sulfur atoms can also participate in stereoelectronic interactions with electron-pair donors in a paradigm termed chalcogen bonding.[135] In these interactions, electron density from a donor, often a carbonyl oxygen, is transferred into the σ^* orbital of one of the bonds to the sulfur atom.[136] These contacts were originally observed in surveys of small-molecule crystal structures,[137] which first indicated the characteristic interaction geometry. The interaction has since been implicated in many examples of small-molecule structure and reactivity.[138] Only in the early 2000s, however, was chalcogen bonding implicated in protein structure.[139,140] Interaction geometries of methionine and disulfide sulfur atoms are broadly consistent with those observed in small-molecule crystal structures and occur most frequently with main-chain oxygens in α-helices.[140] Calculations on a prototypical interaction indicated an energy of 0.64 kcal/mol due to charge transfer;[140] interactions with cystine disulfides are modestly stronger than those with methionine thioethers. Surveys of protein structures find that 13% of cystine residues and 7% of methionine residues engage in sub-van der Waals contacts with oxygen atoms, which could allow for energetically significant interactions. Chalcogen bonding has also been implicated in ligand binding,[141] especially for heterocyclic ligands.[142]

■ RELATIVE CONTRIBUTIONS OF SECONDARY FORCES

This complex suite of interactions can be divided largely into two groups. The first is the set of strong interactions that are relatively uncommon in proteins because of either geometric constraints or low amino acid frequency, typified by the cation−π interaction. These interactions can contribute significant energy to the overall energy of folding, but more importantly, they direct the formation of specific contacts, particularly at positions remote in sequence. Moreover, as these interactions pertain largely to side-chain functionalities, their appreciation is likely to improve methods for predicting protein structure from sequence.

These interactions can be contrasted with the weaker yet more abundant interactions, such as C−H···O hydrogen bonds and $n \rightarrow \pi^*$ interactions. There, individual interactions are likely to be of little importance since their energies fall below the thermal energy at ambient temperatures; however, their cumulative effects over a large number of residues can make a substantial contribution to protein stability. Most are highly local interactions, occurring within a single residue or between adjacent residues, and could thereby guide the earliest events in the protein folding process. In addition, invoking the specific geometric preferences of these interactions might improve model accuracy and refinement. Finally, even crude estimates of the total contributions of pervasive, weak interactions suggest that they play critical roles in stabilizing the overall fold

of proteins (Table 1 and Figure 3), perhaps comparable to some canonical interactions.

Table 1. Estimated Frequencies and Energies of Secondary Forces in Protein Folding

interaction	approximate frequency per 100 residues	approximate energy (kcal/mol)[a]
$n \rightarrow \pi^*$ interactions	33[57]	0.25[69]
C−H···O hydrogen bonds	10[31]	0−1[41,42]
π−π interactions[b]	5[111]	0.5−1.5[117−119]
C5 hydrogen bonds	5[83]	0.25−1.5[83]
cation−π interactions	1−2[94]	0.5−2[85]
sulfur−arene interactions[b]	2−3[130]	0.3−0.5[133,134]
anion−π interactions	1−2[125]	0.5[124]
chalcogen bonds	<1[140]	0.64[140]
X−H···π interactions	1[102]	0.35[143]

[a]Preference is given to experimental measurements in proteins and peptides. Computational values are used in the absence of experimental data. [b]The frequency per 100 residues was estimated by multiplying the frequency of relevant residues[95] by the fraction of those residues that engage in the interaction.
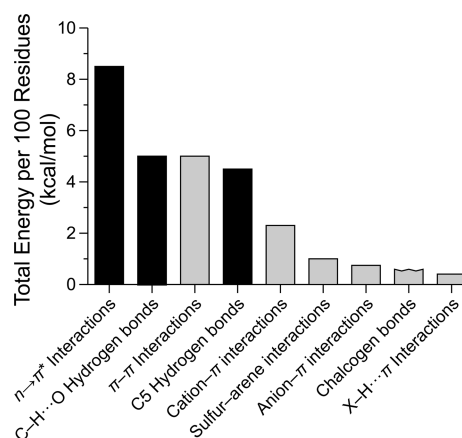

Figure 3. Bar graph of the estimated enthalpic contributions of secondary interactions to the conformational stability of globular proteins. Black bars, interactions of the main chain (Figure 1); gray bars, interactions involving side chains (Figure 2). Data are from Table 1. The sum of the energies is ∼27 kcal/mol per 100 residues.

■ OUTLOOK

A comprehensive understanding of secondary contributions to protein structure should benefit computational force fields. In view of the intimacy of these interactions, secondary interactions might encourage the dense packing commonly observed in folded proteins. In addition, because some of these interactions, such as $n \rightarrow \pi^*$ interactions and C5 hydrogen bonds, correlate with secondary structure, including these parameters could improve secondary structure prediction or refinement. Importantly, partitioning of energetic contributions between individual interactions would allow them to be scrutinized independently. As is, force fields subsume a variety of interactions into relatively few terms. Consider, for example, the hydrogen bonds in an α-helix. There is significant evidence that the α-helix is stabilized by both hydrogen bonds and $n \rightarrow \pi^*$ interactions; however, force fields account only for the hydrogen bonds. Hence, in order to achieve agreement with experimental results, hydrogen-bonding potentials have, in

effect, absorbed the computational energy that should be attributed to the $n \rightarrow \pi^*$ interaction. This approach might be sufficient for modeling an α-helix, but it distorts hydrogen-bonding energies in other regions. Likewise, the empirical optimization of electrostatic parameters might be distorted by absorbing contributions from cation–π interactions. Such canopies reside at the core of computational models, and success might be enhanced by dissecting contributions from secondary interactions and treating those contributions independently.

Significant progress has been made in inventorying the non-covalent interactions available to proteins. Still, additional interactions undoubtedly lack recognition, much less curation. In addition, many known interactions (Table 1) remain poorly characterized in terms of their nature, precise energetic contributions, and frequency. Probing the contributions of backbone interactions is particularly challenging given the lack of genetic approaches to perturbation. Data on the distribution of secondary interactions across secondary and tertiary structural motifs are limited, as are data on the interplay of these interactions with canonical hydrogen-bonding or Coulombic interactions (as well as with one another). In addition, relatively little work has been done to characterize secondary interactions involving post-translational modifications.[143,144] For example, whereas the strong electrostatic consequences of phosphorylation are well-described, the effects of acylation or oxidation remain largely opaque. Moreover, the nature, strength, and roles of many of these interactions in cellular environments remain poorly characterized, as most of the relevant studies to date have been performed *in vitro*—an important consideration given that many secondary interactions are sensitive to solvent and other environmental conditions. Finally, as protein design and engineering efforts advance, consideration of interactions not possible in natural, proteinogenic amino acids (e.g., halogen bonding[145−147]) could also warrant attention.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: rtraines@mit.edu.

ORCID ⊙
Robert W. Newberry: 0000-0002-2020-2641
Ronald T. Raines: 0000-0001-7164-1719

Present Address
∥R.W.N.: Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94158, United States.

Notes
The authors declare no competing financial interest.

■ KEYWORDS

protein folding: process by which a linear polypeptide adopts its three-dimensional conformation
C−H···O hydrogen bond: hydrogen bond between a carbon-based acid and an oxygen lone pair

$n \rightarrow \pi^*$ interaction: stereoelectronic interaction between a lone pair and a π antibonding orbital, especially that between two carbonyl groups
C5 hydrogen bond: intraresidue hydrogen bond between backbone N−H and C=O groups in a β-strand
cation−π interaction: interaction of a positive charge with the face of an aromatic ring
X−H···π interaction: hydrogen bond donated to the face of an aromatic ring
π−π interaction: interaction between two aromatic rings in either an edge-to-face or offset-stacked geometry
anion−π interaction: interaction of a negative charge with the edge of an aromatic ring
sulfur−arene interaction: short contact between a sulfur atom and an aromatic ring
chalcogen bonding: stereoelectronic interaction between a lone pair and σ antibonding orbital of a C−S or S−S bond

■ REFERENCES

(1) Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on. *Science 338*, 1042−1046.

(2) Anfinsen, C. B., Haber, E., Sela, M., and White, F. H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A. 47*, 1309−1314.

(3) Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science 181*, 223−230.

(4) Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Glucksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D.,

Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) The sequence of the human genome. *Science 291*, 1304−1351.

(5) Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem. 75*, 333−366.

(6) Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) Protein design: A hierarchic approach. *Science 270*, 935−941.

(7) Huang, P.-S., Boyken, S. E., and Baker, D. (2016) The coming of age of *de novo* protein design. *Nature 537*, 320−327.

(8) Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry 29*, 7133−7155.

(9) Richards, F. M. (1997) Protein stability: Still an unsolved problem. *Cell. Mol. Life Sci. 53*, 790−802.

(10) Privalov, P. L., and Gill, S. J. (1988) Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem. 39*, 191−234.

(11) Makhatadze, G. I., and Privalov, P. I. (1995) Energetics of protein structure. *Adv. Protein Chem. 47*, 307−425.

(12) Weissman, J. S., and Kim, P. S. (1991) Reexamination of the folding of BPTI: Predominance of native intermediates. *Science 253*, 1386−1393.

(13) Pace, C. N., Scholtz, J. M., and Grimsley, G. R. (2014) Forces stabilizing proteins. *FEBS Lett. 588*, 2177−2184.

(14) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187−217.

(15) Mayo, S. L., Olafson, B. D., and Goddard, W. A., III (1990) DREIDING: A generic force field for molecular simulations. *J. Phys. Chem. 94*, 8897−8909.

(16) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. 117*, 5179−5197.

(17) Dahiyat, B. I., and Mayo, S. L. (1997) *De novo* protein design: Fully automated sequence selection. *Science 278*, 82−87.

(18) Das, R., and Baker, D. (2008) Macromolecular modeling with Rosetta. *Annu. Rev. Biochem. 77*, 363−382.

(19) Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364−1368.

(20) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., III, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) *De novo* computational design of retro-aldol enzymes. *Science 319*, 1387−1391.

(21) Moult, J. (2005) A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol. 15*, 285−289.

(22) Kryshtafovych, A., Fidelis, K., and Moult, J. (2011) CASP9 results compared to those of previous CASP experiments. *Proteins: Struct., Funct., Genet. 79*, 196−207.

(23) Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2014) Critical assessment of methods of protein structure prediction (CASP)—round X. *Proteins: Struct., Funct., Genet. 82*, 1−6.

(24) Koehl, P., and Levitt, M. (1999) A brighter future for protein structure prediction. *Nat. Struct. Biol. 6*, 108−111.

(25) Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P., and Williams, D. E. (2000) A test of crystal structure prediction of small organic molecules. *Acta Crystallogr., Sect. B: Struct. Sci. 56*, 697−714.

(26) Bardwell, D. A., Adjiman, C. S., Arnautova, Y. A., Bartashevich, E., Boerrigter, S. X., Braun, D. E., Cruz-Cabeza, A. J., Day, G. M., Della Valle, R. G., Desiraju, G. R., van Eijck, B. P., Facelli, J. C., Ferraro, M. B., Grillo, D., Habgood, M., Hofmann, D. W. M., Hofmann, F., Jose, K. V., Karamertzanis, P. G., Kazantsev, A. V., Kendrick, J., Kuleshova, L. N., Leusen, F. J. J., Maleev, A. V., Misquitta, A. J., Mohamed, S., Needs, R. J., Neumann, M. A., Nikylov, D., Orendt, A. M., Pal, R., Pantelides, C. C., Pickard, C. J., Price, L. S., Price, S. L., Scheraga, H. A., van de Streek, J., Thakur, T. S., Tiwari, S., Venuti, E., and Zhitkov, I. K. (2011) Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. *Acta Crystallogr., Sect. B: Struct. Sci. B67*, 535−551.

(27) Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255*, 178−183.

(28) Brock, C. P., and Dunitz, J. D. (1994) Towards a grammar of crystal packing. *Chem. Mater. 6*, 1118−1127.

(29) Klapper, M. H. (1971) On the nature of the protein interior. *Biochim. Biophys. Acta, Protein Struct. 229*, 557−566.

(30) Fleming, P. J., and Rose, G. D. (2005) Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci. 14*, 1911−1917.

(31) Bartlett, G. J., and Woolfson, D. N. (2016) On the satisfaction of backbone-carbonyl lone pairs of electrons in protein structures. *Protein Sci. 25*, 887−897.

(32) Horowitz, S., and Trievel, R. C. (2012) Carbon−oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem. 287*, 41576−41582.

(33) Derewenda, Z. S., Derewenda, U., and Kobos, P. M. (1994) (His)C^ε−H···O=C< hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol. 241*, 83−93.

(34) Ramachandran, G. N., and Sasisekharan, V. (1965) Refinement of the structure of collagen. *Biochim. Biophys. Acta, Biophys. Incl. Photosynth. 109*, 314−316.

(35) Ramachandran, G. N., and Venkatachalam, C. M. (1966) The stability of the two-bonded collagen triple helix. *Biochim. Biophys. Acta, Biophys. Incl. Photosynth. 120*, 457−458.

(36) Taylor, R., and Kennard, O. (1982) Crystallographic evidence for the existence of C−H···O, C−H···N, and C−H···Cl hydrogen bonds. *J. Am. Chem. Soc. 104*, 5063−5070.

(37) Gu, Y., Kar, T., and Scheiner, S. (1999) Fundamental properties of the CH···O Interaction: Is it a true hydrogen bond? *J. Am. Chem. Soc. 121*, 9411−6422.

(38) Steiner, T. (1997) Unrolling the hydrogen bond properties of C−H···O interactions. *Chem. Commun. 8*, 727−723.

(39) Hobza, P., and Havlas, Z. (2000) Blue-shifting hydrogen bonds. *Chem. Rev. 100*, 4253−4264.

(40) Qian, W., and Krimm, S. (2002) Vibrational spectroscopy of hydrogen bonding: Origin of the different behavoir of the C−H···O hydrogen bond. *J. Phys. Chem. A 106*, 6628−6636.

(41) Yohannan, S., Faham, S., Yang, D., Grosfeld, D., Chamberlain, A. K., and Bowie, J. U. (2004) A C^α−H···O hydrogen bond in a membrane protein is not stabilizing. *J. Am. Chem. Soc. 126*, 2284−2285.

(42) Arbely, E., and Arkin, I. T. (2004) Experimental measurement of the strength of a C^α−H···O bond in a lipid bilayer. *J. Am. Chem. Soc. 126*, 5362−5363.

(43) Scheiner, S., Kar, T., and Gu, Y. (2001) Strength of the C^αH···O hydrogen bond of amino acid residues. *J. Biol. Chem. 276*, 9832−9837.

(44) Scheiner, S. (2005) Relative strengths of NH···O and CH···O hydrogen bonds between polypeptide chain segments. *J. Phys. Chem. B 109*, 16132−16141.

(45) Vargas, R., Garza, J., Dixon, D. A., and Hay, B. P. (2000) How strong is the C$^\alpha$–H···O═C hydrogen bond? *J. Am. Chem. Soc. 122*, 4750−4755.

(46) McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol. 238*, 777−793.

(47) Derewenda, Z. S., Lee, L., and Derewenda, U. (1995) The occurrence of C–H···O hydrogen bonds in proteins. *J. Mol. Biol. 252*, 248−262.

(48) Scheiner, S. (2006) Contributions of NH···O and CH···O hydrogen bonds to the stability of β-sheets in proteins. *J. Phys. Chem. B 110*, 18670−18679.

(49) Manikandan, K., and Ramakumar, S. (2004) The occurrence of C−H···O hydrogen bonds in α-helices and helix termini in globular proteins. *Proteins: Struct., Funct., Genet. 56*, 768−781.

(50) Chakrabarti, P., and Chakrabarti, S. (1998) C−H···O Hydrogen bond involving proline residues in α-helices. *J. Mol. Biol. 284*, 867−873.

(51) Bella, J., and Berman, H. M. (1996) Crystallographic evidence for C$^\alpha$−H···O═C hydrogen bonds in a collagen triple helix. *J. Mol. Biol. 264*, 734−742.

(52) Jiang, L., and Lai, L. (2002) CH···O hydrogen bonds at protein−protein interfaces. *J. Biol. Chem. 277*, 37732−37740.

(53) Senes, A., Gerstein, M., and Engelman, D. M. (2000) Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. *J. Mol. Biol. 296*, 921−936.

(54) Senes, A., Ubarretxena-Belandia, I., and Engelman, D. M. (2001) The C$^\alpha$−H···O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. U. S. A. 98*, 9056−9061.

(55) Mueller, B. K., Subramaniam, S., and Senes, A. (2014) 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. 111*, E888−E895.

(56) Hinderaker, M. P., and Raines, R. T. (2003) An electronic effect on protein structure. *Protein Sci. 12*, 1188−1194.

(57) Bartlett, G. J., Choudhary, A., Raines, R. T., and Woolfson, D. N. (2010) n→π* Interactions in proteins. *Nat. Chem. Biol. 6*, 615−620.

(58) Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001) Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc. 123*, 777−778.

(59) Singh, S. K., and Das, A. (2015) The n→π* interaction: A rapidly emerging non-covalent interaction. *Phys. Chem. Chem. Phys. 17*, 9596−9612.

(60) Newberry, R. W., and Raines, R. T. (2017) The n→π* interaction. *Acc. Chem. Res. 50*, 1838−1846.

(61) Sahariah, B., and Sarma, B. K. (2019) Relative orientation of the carbonyl groups determines the nature of orbital interactions in carbonyl−carbonyl short contacts. *Chem. Sci. 10*, 909−917.

(62) Bürgi, H. B., Dunitz, J. D., and Shefter, E. (1973) Geometric reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc. 95*, 5065−5067.

(63) Bürgi, H. B., Dunitz, J. D., and Shefter, E. (1974) Chemical reaction paths. IV. Aspects of O···C═O interactions in crystals. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B30*, 1517−1527.

(64) Fischer, F. R., Wood, P. A., Allen, F. H., and Diederich, F. (2008) Orthogonal dipolar interactions between amide carbonyl groups. *Proc. Natl. Acad. Sci. U. S. A. 105*, 17290−17294.

(65) Fäh, C., Hardegger, L. A., Ebert, M. O., Schweizer, W. B., and Diederich, F. (2010) Self-association based on orthogonal C═O··· C═O interactions in the solid and liquid state. *Chem. Commun. 46*, 67−69.

(66) Worley, B., Richard, G., Harbison, G. S., and Powers, R. (2012) ^{13}C NMR reveals no evidence of n→π* interactions in proteins. *PLoS One 7*, No. e42075.

(67) Choudhary, A., Gandla, D., Krow, G. R., and Raines, R. T. (2009) Nature of amide carbonyl−carbonyl interactions in proteins. *J. Am. Chem. Soc. 131*, 7244−7246.

(68) Kamer, K. J., Choudhary, A., and Raines, R. T. (2013) Intimate interactions with carbonyl groups: Dipole−dipole or n→π*? *J. Org. Chem. 78*, 2099−2103.

(69) Newberry, R. W., VanVeller, B., Guzei, I. A., and Raines, R. T. (2013) n→π* Interactions of amides and thioamides: Implications for protein stability. *J. Am. Chem. Soc. 135*, 7843−7846.

(70) Persch, E., Dumele, O., and Diederich, F. (2015) Molecular recognition in chemical and biological systems. *Angew. Chem., Int. Ed. 54*, 3290−3327.

(71) Choudhary, A., and Raines, R. T. (2011) Signature of n→π* interactions in α-helices. *Protein Sci. 20*, 1077−1081.

(72) Newberry, R. W., Bartlett, G. J., Vanveller, B., Woolfson, D. N., and Raines, R. T. (2014) Signatures of n→π* interactions in proteins. *Protein Sci. 23*, 284−288.

(73) Newberry, R. W., Orke, S. J., and Raines, R. T. (2016) n→π* Interactions are competitive with hydrogen bonds. *Org. Lett. 18*, 3614−3617.

(74) Kuemin, M., Nagel, Y. A., Schweizer, S., Monnard, F. W., Ochsenfeld, C., and Wennemers, H. (2010) Tuning the *cis/trans* conformer ratio of Xaa−Pro amide bonds by intramolecular hydrogen bonds: The effect on PPII helix stability. *Angew. Chem., Int. Ed. 49*, 6324−6327.

(75) Shoulders, M. D., Kotch, F. W., Choudhary, A., Guzei, I. A., and Raines, R. T. (2010) The aberrance of the 4S diastereomer of 4-hydroxyproline. *J. Am. Chem. Soc. 132*, 10857−10865.

(76) Wenzell, N. A., Ganguly, H. K., Pandey, A. K., Bhatt, M. R., Yap, G. P. A., and Zondlo, N. J. (2019) Electronic and steric control of n→π* interactions: Stabilization of the α-helix conformation without a hydrogen bond. *ChemBioChem 20*, 963−967.

(77) Horng, J. C., and Raines, R. T. (2006) Stereoelectronic effects on polyproline conformation. *Protein Sci. 15*, 74−83.

(78) Wilhelm, P., Lewandowski, B., Trapp, N., and Wennemers, H. (2014) A crystal structure of an oligoproline PPII-helix, at last. *J. Am. Chem. Soc. 136*, 15829−15832.

(79) Pal, T. K., and Sankararamakrishnan, R. (2010) Quantum chemical investigations on intraresidue carbonyl−carbonyl contacts in aspartates of high-resolution protein structures. *J. Phys. Chem. B 114*, 1038−1049.

(80) Bartlett, G. J., Newberry, R. W., Vanveller, B., Raines, R. T., and Woolfson, D. N. (2013) Interplay of hydrogen bonds and n→π* interactions in proteins. *J. Am. Chem. Soc. 135*, 18682−18688.

(81) Kilgore, H. R., and Raines, R. T. (2018) n→π* Interactions modulate the properties of cysteine residues and disulfide bonds in proteins. *J. Am. Chem. Soc. 140*, 17606−17611.

(82) Toniolo, C., and Benedetti, E. (1980) Intramolecularly hydrogen-bonded peptide conformations. *CRC Crit. Rev. Biochem. 9*, 1−44.

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

(84) Gallagher-Jones, M., Glynn, C., Boyer, D. R., Martynowycz, M. W., Hernandez, E., Miao, J., Zee, C. T., Novikova, I. V., Goldschmidt, L., McFarlane, H. T., Helguera, G. F., Evans, J. E., Sawaya, M. R., Cascio, D., Eisenberg, D. S., Gonen, T., and Rodriguez, J. A. (2018) Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat. Struct. Mol. Biol. 25*, 131−134.

(85) Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed. 42*, 1210−1250.

(86) Dougherty, D. A. (1996) Cation−π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science 271*, 163−168.

(87) Dougherty, D. A. (2013) The cation−π interaction. *Acc. Chem. Res. 46*, 885−893.

(88) Ma, J. C., and Dougherty, D. A. (1997) The cation−π interaction. *Chem. Rev. 97*, 1303−1324.

(89) Mecozzi, S., West, A. P., Jr., and Dougherty, D. A. (1996) Cation−π interactions in simple aromatics: Electrostatics provide a predictive tool. *J. Am. Chem. Soc. 118*, 2307−2308.

(90) Mecozzi, S., West, A. P., Jr., and Dougherty, D. A. (1996) Cation−π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. U. S. A. 93*, 10566−10571.

(91) Kumpf, R. A., and Dougherty, D. A. (1993) A mechanism for ion selectivity in potassium channels: Computational studies of cation−π interactions. *Science 261*, 1708−1710.

(92) Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nat. Struct. Mol. Biol. 14*, 1025−1040.

(93) Loewenthal, R., Sancho, J., and Fersht, A. R. (1992) Histidine−aromatic interactions in barnase: Elevation of histidine pK_a and contribution to protein stability. *J. Mol. Biol. 224*, 759−770.

(94) Gallivan, J. P., and Dougherty, D. A. (1999) Cation−π interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A. 96*, 9459−9464.

(95) Echols, N., Harrison, P., Balasubramanian, S., Luscombe, N. M., Bertone, P., Zhang, Z., and Gerstein, M. (2002) Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res. 30*, 2515−2523.

(96) Tsou, L. K., Tatko, C. D., and Waters, M. L. (2002) Simple cation−π interaction between a phenyl ring and a protonated amine stabilizes an α-helix in water. *J. Am. Chem. Soc. 124*, 14917−14921.

(97) Waters, M. L. (2004) Aromatic interactions in peptides: Impact on structure and function. *Biopolymers 76*, 435−445.

(98) Crowley, P. B., and Golovin, A. (2005) Cation−π interactions in protein−protein interfaces. *Proteins: Struct., Funct., Genet. 59*, 231−239.

(99) Salonen, L. M., Ellermann, M., and Diederich, F. (2011) Aromatic rings in chemical and biological recognition: Energetics and structures. *Angew. Chem., Int. Ed. 50*, 4808−4842.

(100) Craven, T. W., Cho, M. K., Traaseth, N. J., Bonneau, R., and Kirshenbaum, K. (2016) A miniature protein stabilized by a cation−π interaction network. *J. Am. Chem. Soc. 138*, 1543−50.

(101) Levitt, M., and Perutz, M. F. (1988) Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol. 201*, 751−754.

(102) Steiner, T., and Koellner, G. (2001) Hydrogen bonds with π-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J. Mol. Biol. 305*, 535−557.

(103) Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J., and Hilgenfeld, R. (2001) CH···π-Interactions in proteins. *J. Mol. Biol. 307*, 357−377.

(104) Plevin, M. J., Bryce, D. L., and Boisbouvier, J. (2010) Direct detection of CH/π interactions in proteins. *Nat. Chem. 2*, 466−471.

(105) Perras, F. A., Marion, D., Boisbouvier, J., Bryce, D. L., and Plevin, M. J. (2017) Observation of CH···π interactions between methyl and carbonyl groups in proteins. *Angew. Chem., Int. Ed. 56*, 7564−7567.

(106) Chen, W., Enck, S., Price, J. L., Powers, D. L., Powers, E. T., Wong, C. H., Dyson, H. J., and Kelly, J. W. (2013) Structural and energetic basis of carbohydrate−aromatic packing interactions in proteins. *J. Am. Chem. Soc. 135*, 9877−9884.

(107) Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate−aromatic interactions in proteins. *J. Am. Chem. Soc. 137*, 15152−15160.

(108) Kiessling, L. L. (2018) Chemistry-driven glycoscience. *Bioorg. Med. Chem. 26*, 5229−5238.

(109) Vyas, N. K. (1991) Atomistic features of protein−carbohydrate interactions. *Curr. Opin. Struct. Biol. 1*, 732−740.

(110) Baker, E. G., Williams, C., Hudson, K. L., Bartlett, G. J., Heal, J. W., Porter Goff, K. L., Sessions, R. B., Crump, M. P., and Woolfson, D. N. (2017) Engineering protein stability with atomic precision in a monomeric miniprotein. *Nat. Chem. Biol. 13*, 764−770.

(111) Burley, S. K., and Petsko, G. A. (1985) Aromatic−aromatic interaction: A mechanism of protein structure stabilization. *Science 229*, 23−28.

(112) Blundell, T., Singh, J., Thornton, J., Burley, S. K., and Petsko, G. A. (1986) Aromatic interactions. *Science 234*, 1005.

(113) Hunter, C. A., and Sanders, J. K. M. (1990) The nature of π−π interactions. *J. Am. Chem. Soc. 112*, 5525−5534.

(114) Hunter, C. A., Singh, J., and Thornton, J. M. (1991) π−π Interactions: The geometry and energetics of phenylalanine−phenylalanine interactions in proteins. *J. Mol. Biol. 218*, 837−846.

(115) McGaughey, G. B., Gagné, M., and Rappé, A. K. (1998) π-Stacking interactions: Alive and well in proteins. *J. Biol. Chem. 273*, 15458−15463.

(116) Kannan, N., and Vishveshwara, S. (2000) Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Eng., Des. Sel. 13*, 753−761.

(117) Tatko, C. D., and Waters, M. L. (2002) Selective aromatic interactions in β-hairpin peptides. *J. Am. Chem. Soc. 124*, 9372−9373.

(118) Butterfield, S. M., Patel, P. R., and Waters, M. L. (2002) Contribution of aromatic interactions to α-helix stability. *J. Am. Chem. Soc. 124*, 9751−9755.

(119) Serrano, L., Bycroft, M., and Fersht, A. R. (1991) Aromatic−aromatic interactions and protein stability: Investigation by double-mutant cycles. *J. Mol. Biol. 218*, 465−475.

(120) Burley, S. K., and Petsko, G. A. (1986) Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc. 108*, 7995−8001.

(121) Chelli, R., Gervasio, F. L., Procacci, P., and Schettino, V. (2002) Stacking and T-shape competition in aromatic−aromatic amino acid interactions. *J. Am. Chem. Soc. 124*, 6133−6143.

(122) Frontera, A., Gamez, P., Mascal, M., Mooibroek, T. J., and Reedijk, J. (2011) Putting anion−π interactions into perspective. *Angew. Chem., Int. Ed. 50*, 9564−9583.

(123) Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H.-N., Philip, V. M., Hinde, R. J., and Howell, E. E. (2007) A preference for edgewise interactions between aromatic rings and carboxylate anions: The biological relevance of anion−quadrupole interactions. *J. Phys. Chem. B 111*, 8242−8249.

(124) Shi, Z., Olson, C. A., Bell, A. J., Jr., and Kallenbach, N. R. (2002) Non-classical helix-stabilizing interactions: C−H···O H-bonding between Phe and Glu side chains in α-helical peptides. *Biophys. Chem. 101*, 267−279.

(125) Chakravarty, S., Ung, A. R., Moore, B., Shore, J., and Alshamrani, M. (2018) A comprehensive analysis of anion−quadrupole interactions in protein structures. *Biochemistry 57*, 1852−1867.

(126) Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E., and Hinde, R. J. (2011) A survey of aspartate−phenylalanine and glutamate−phenylalanine interactions in the protein data bank: Searching for anion−π pairs. *Biochemistry 50*, 2939−2950.

(127) Lucas, X., Bauzá, A., Frontera, A., and Quiñonero, D. (2016) A thorough anion−π interaction study in biomolecules: On the importance of cooperativity effects. *Chem. Sci. 7*, 1038−1050.

(128) Smith, M. S., Lawrence, E. E. K., Billings, W. M., Larsen, K. S., Bécar, N. A., and Price, J. L. (2017) An anion−π interaction strongly stabilizes the β-sheet protein WW. *ACS Chem. Biol. 12*, 2535−2537.

(129) Egli, M., and Sarkhel, S. (2007) Lone pair−aromatic interactions: To stabilize or not to stabilize. *Acc. Chem. Res. 40*, 197−205.

(130) Reid, K. S. C., Lindley, P. F., and Thornton, J. M. (1985) Sulphur−aromatic interactions in proteins. *FEBS Lett. 190*, 209−213.

(131) Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., and Sachs, J. N. (2012) The methionine−aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem. 287*, 34979−34991.

(132) Zauhar, R. J., Colbert, C. L., Morgan, R. S., and Welsh, W. J. (2000) Evidence for a strong sulfur−aromatic interactions derived from crystallographic data. *Biopolymers 53*, 233−248.

(133) Viguera, A. R., and Serrano, L. (1995) Side-chain interactions between sulfur-containing amino acids and phenylalanine in α-helices. *Biochemistry 34*, 8771−8779.

(134) Tatko, C. D., and Waters, M. L. (2004) Investigation of the nature of the methionine−π interaction in β-hairpin peptide model systems. *Protein Sci. 13*, 2515−2522.

(135) Iwaoka, M. Chalcogen bonds in protein architecture. In *Noncovalent Forces*; Scheiner, S., Ed.; Challenges and Advances in Computational Chemistry and Physics, Vol. *19*; Springer: Cham, Switzerland, 2015; pp 265−289.

(136) Pascoe, D. J., Ling, K. B., and Cockroft, S. L. (2017) The origin of chalcogen-bonding interactions. *J. Am. Chem. Soc. 139*, 15160−15167.

(137) Rosenfield, R. E., Jr., Parthasarathy, R., and Dunitz, J. D. (1977) Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles. *J. Am. Chem. Soc. 99*, 4860−4862.

(138) Vogel, L., Wonner, P., and Huber, S. M. (2019) Chalcogen bonding: An overview. *Angew. Chem., Int. Ed. 58*, 1880−1891.

(139) Pal, D., and Chakrabarti, P. (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. *J. Biomol. Struct. Dyn. 19*, 115−128.

(140) Iwaoka, M., Takemoto, S., Okada, M., and Tomoda, S. (2002) Weak nonbonded S···X (X = O, N, and S) interactions in proteins. *Bull. Chem. Soc. Jpn. 75*, 1611−1625.

(141) Kříž, K., Fanfrlík, J., and Lepšík, M. (2018) Chalcogen bonding in protein−ligand complexes: PDB survey and quantum mechanical calculations. *ChemPhysChem 19*, 2540−2548.

(142) Mitchell, M. O. (2017) Discovering protein−ligand chalcogen bonding in the protein data bank using endocyclic sulfur-containing heterocycles as ligand search subsets. *J. Mol. Model. 23*, 287.

(143) Hughes, R. M., and Waters, M. L. (2006) Effects of lysine acetylation in a β-hairpin peptide: Comparison of an amide−π and a cation−π interaction. *J. Am. Chem. Soc. 128*, 13586−13591.

(144) Beaver, J. E., and Waters, M. L. (2016) Molecular recognition of Lys and Arg methylation. *ACS Chem. Biol. 11*, 643−653.

(145) Auffinger, P., Hays, F. A., Westhof, E., and Ho, P. S. (2004) Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A. 101*, 16789−16794.

(146) Metrangolo, P., Neukirch, H., Pilati, T., and Resnati, G. (2005) Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Acc. Chem. Res. 38*, 386−395.

(147) Voth, A. R., Khuu, P., Oishi, K., and Ho, P. S. (2009) Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat. Chem. 1*, 74−79.