HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2017 April 17.

Published in final edited form as:

Nat Chem Biol. 2016 December ; 12(12): 1084–1088. doi:10.1038/nchembio.2206.

A prevalent intraresidue hydrogen bond stabilizes proteins

Robert W. Newberry[1] and **Ronald T. Raines**[1,2,*]

[1]Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

[2]Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

Abstract

Current limitations in *de novo* protein structure prediction and design suggest an incomplete understanding of the interactions that govern protein folding. Here we demonstrate that previously unappreciated hydrogen bonds occur within proteins between the amide proton and carbonyl oxygen of the same residue. Quantum calculations, infrared spectroscopy, and nuclear magnetic resonance spectroscopy show that these interactions share hallmark features of canonical hydrogen bonds. Biophysical analyses demonstrate that selective attenuation or enhancement of these C5 hydrogen bonds affects the stability of synthetic β-sheets. These interactions are common, affecting approximately 5% of all residues and 94% of proteins, and their cumulative impact provides several kcal/mol of conformational stability to a typical protein. C5 hydrogen bonds stabilize, especially, the flat β-sheets of the amyloid state, which is linked with Alzheimer's disease and other neurodegenerative disorders. Inclusion of these interactions in computational force fields would improve models of protein folding, function, and dysfunction.

Predicting the structure of a protein from its sequence remains challenging[1], motivating study of the noncovalent interactions that govern protein folding[2]. Hydrogen bonding has received particular attention due to its unique ability to specify the geometry by which chemical groups interact. In proteins, amide hydrogen-bond donors often approach carbonyl acceptors along the carbonyl bond axis, as in the α-helix and β-sheet[3–5]. This observation is consistent with a modern description of the carbonyl lone pairs, featuring a predominantly *s*-type orbital (n_s) along the carbonyl bond with a *p* orbital (n_p) in an orthogonal orientation (Fig. 1a,b)[6]. Whereas n_s is poised to engage in common hydrogen-bond patterns in proteins, the role of n_p is less clear. In an α-helix, the π^* orbital of an adjacent carbonyl group can accept lone pair electron density from n_p, forming a so-called $n \rightarrow \pi^*$ interaction[7]; however, adjacent carbonyl groups in β-sheets are too distil to accept electron density from n_p. Moreover, whereas most backbone carbonyl groups in proteins can form two noncovalent

*Correspondence and requests for materials should be addressed to R.T.R., rtraines@wisc.edu.

Author contributions

R.W.N. conceived of the project. R.W.N. and R.T.R. planned the experiments. R.W.N. carried out the experiments. R.W.N. and R.T.R. analysed the data and wrote the manuscript.

Supplementary Information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints.

Competing financial interests

The authors declare no competing financial interests.

interactions, one for each lone pair, the majority of backbone carbonyls that form only a single interaction are found in β-sheets[8], suggesting that we might not yet appreciate all of the interactions present in this latter secondary structure. We therefore inquired as to what other electron acceptors could potentially engage with the carbonyl *p*-type lone pair in β-sheets.

Upon inspection of the β-sheet, we noted close proximity of the carbonyl oxygen to the amide proton of the same residue, potentially creating a hydrogen bond through overlap of n_p with the σ^* orbital of the N–H bond (Fig. 1c). This geometry has been termed "C5" for the size of the ring enclosed by the putative hydrogen bond[9–11]. The C5 geometry has been observed for some amino acids[12,13] and small, charged proteins[14] in the gas phase, and calculations suggest it to be one of few minima on the potential energy surface[15,16]. The putative C5 hydrogen bond itself has, however, never been probed in the context of a peptide or protein in solution. If these interactions do indeed occur, their energies are likely to be weak, given that hydrogen-bond energies are maximized when the acceptor approaches along the donor bond axis[17]; indeed, algorithms for identifying hydrogen bonds in protein structures require this geometry for hydrogen-bond assignment[18]. Thus, it is unclear if C5 interactions are truly hydrogen bonds or if they contribute to protein stability.

Here we used a variety of experimental and computational methods to probe for the existence of C5 hydrogen bonds in proteins. In particular, we used quantum mechanical calculations and various types of spectroscopy on minimal peptide models to show that these interactions share features that are typical of hydrogen bonds. We then demonstrated that selective perturbation of the C5 hydrogen bond in backbone-modified peptides affects the conformational stability of a β-sheet. Bioinformatics analysis revealed that these interactions are prevalent in virtually all proteins, leading us to conclude that C5 hydrogen bonds contribute significantly to the conformational stability of proteins.

Results

Computational analysis of C5 hydrogen bonding

For an interaction to form, interpenetration of the donor and acceptor orbitals must occur. To evaluate the potential for orbital overlap in folded proteins, we measured the distance between putative donors and acceptors within β-sheet residues in sub-Å protein crystal structures. We found that many featured oxygen–hydrogen distances <2.61 Å, which is the sum of the van der Waals radii for hydrogen and oxygen[19] (Supplementary Results, Supplementary Fig. 1), suggesting that orbital interpenetration might occur. To investigate whether relevant geometries enable C5 hydrogen bonding, we used density functional theory calculations to generate a set of AcGlyNHMe conformations that sample the donor–acceptor distances observed in proteins (Fig. 2a and Supplementary Table 1). These conformers were then subjected to natural bond orbital (NBO) analysis to estimate the energy released by the mixing of the carbonyl *p*-type lone pair with the σ^* orbital of the hydrogen-bond acceptor (Supplementary Table 2)[20]. NBO analysis allows for calculation of the energy associated with the interactions of specific orbitals, which generally agree well with, for example, the energy of $n \rightarrow \pi^*$ interactions determined from torsion-balance experiments[21–23].

Our calculations predict that C5 interactions can release significant energy, especially at donor–acceptor distances <2.5 Å. There, energy-release can exceed 0.25 kcal/mol (Supplementary Fig. 2b), which is similar to that expected from $n \rightarrow \pi^*$ interactions[7,23]. In addition, the overall change in energy along this coordinate is similar to the hydrogen-bond energy estimated by NBO analysis (Supplementary Fig. 2c), suggesting that the interaction itself could make a significant contribution to the stability of these conformations. Moreover, as the donor–acceptor distance decreases, we calculated a concomitant lengthening of the N–H donor bond, increase in the partial positive charge on the donor hydrogen, red shift of the donor stretching frequency, and downfield chemical shift of the donor proton (Supplementary Fig. 2d–g), each consistent with these interactions having the properties of typical hydrogen bonds[17].

C5 Hydrogen bonding in a preorganized amino acid

Encouraged by these computational predictions, we set out to probe a single putative C5 hydrogen bond experimentally. To preorganize the putative donor and acceptor for interaction, we employed derivatives of diethylglycine (Deg) (Fig. 2), which has been shown by computation[24] as well as NMR and vibrational spectroscopy[25] to populate the C5 geometry in solution. To probe the interaction, we compared diethylglycines bearing either an amide or an ester as the putative hydrogen-bond acceptor, as *bona fide* hydrogen bonds are attenuated by replacing an amide acceptor with an ester[26]. Conformational analysis demonstrates that these compounds adopt the C5 geometry in solution (Supplementary Fig. 3), allowing us to isolate the effect of the putative C5 hydrogen bond.

Frequency analysis of the optimized C5 geometries of AcDegNHMe and AcDegOMe predicts that AcDegNHMe has a 30 cm^{-1} lower N–H stretching frequency than does AcDegOMe, consistent with the stronger C5 hydrogen bond predicted by NBO analysis (2.84 kcal/mol versus 1.73 kcal/mol). At a concentration of 10 mM, where similar compounds have been shown to be monomeric in solution[25], infrared spectroscopy revealed that the N–H stretching mode of the putative C5 hydrogen-bond donor is indeed red-shifted —by 33 cm^{-1}—in AcDegNHMe relative to AcDegOMe (Fig. 2a); for comparison, there is only a 6 cm^{-1} difference in the analogous donor stretching frequencies of AcGlyNHMe and AcGlyOMe in dichloromethane[27]. The donor stretching mode of AcDegNHMe is also red-shifted by 3.2 cm^{-1} relative to that in $AcDegNH_2$ (**6**) (Supplementary Fig. 4), consistent with a primary amide being a less effective hydrogen-bond acceptor than is a secondary amide.

We then examined the effect of this interaction on the NMR properties of the donor proton. Replacing the acceptor amide with an ester creates an upfield chemical shift of the donor proton, despite the ester being more electron-withdrawing (Supplementary Fig. 5). These chemical shifts were also insensitive to ten-fold dilution, indicating the absence of intermolecular hydrogen bonding at these concentrations. Switching from a solvent that does not engage in hydrogen bonding to one that does causes a downfield shift of the donor proton, and this effect is attenuated by the presence of an internal hydrogen bond. In accord with previous studies[25], we found that the internal amide proton of AcDegNHMe experiences a much smaller chemical shift upon changing the solvent from $CDCl_3$ to DMSO-d_6 than do the amide protons of AcDegOMe or 3-acetamido-3-methylpentane,

which is a diethylglycine derivative lacking a putative C5 hydrogen-bond acceptor (Fig. 2b). Moreover, in deuterated methanol, the C5 hydrogen-bonded N–H proton of $AcDegNH_2$ experiences smaller temperature-induced chemical shifts (5.4 ppb/°C) than does either of the terminal N–H protons (5.8 and 6.5 ppb/°C), suggesting that C5 hydrogen bonding persists even when exposed completely to a protic solvent.

We next assessed the ability of this interaction to slow the exchange of the donor proton. Following addition of D_2O to a DMSO-d_6 solution, the putative C5 hydrogen-bond donor in AcDegNHMe exchanges much more slowly than does the corresponding proton in AcDegOMe (Fig. 2c, Supplementary Figs. 6 and 7), consistent with the stronger protective effect of the amide acceptor. This difference in exchange rate is contingent upon adoption of the C5 geometry, as the exchange rates of AcGlyNHMe and AcGlyOMe are indistinguishable (Fig. 2d). These data establish that these interactions are consistent with hydrogen bonding. They might also explain the anomalously slow exchange rates of solvent-exposed β-sheet residues in proteins (*e.g.*, Staphylococcal nuclease[28]) that are not engaged in canonical hydrogen bonds.

C5 hydrogen bonding in a β-sheet

Confident that C5 interactions constitute hydrogen bonding and that they significantly affect the chemical properties of the backbone, we sought to characterize their contributions to the stability of an actual β-sheet. To do so, we employed a tryptophan zipper (TrpZip2) β-hairpin as a model (Fig. 3a; Supplementary Table 3)[29]. This small, designed antiparallel β-hairpin achieves robust secondary structure despite its short length, providing a convenient model system. To discern whether C5 hydrogen bonding does occur in β-sheet structure, we compared the ^{1}H NMR chemical shifts of TrpZip2-A and TrpZip2-B (Fig. 3b; Supplementary Tables 4–6); the latter features an ester linkage between Thr3 and Trp4, where it can accept a canonical hydrogen bond from Thr10-NH and a C5 hydrogen bond from Thr3-NH. As expected, Thr10-NH experiences a large upfield shift in the presence of a weaker hydrogen-bond acceptor. Importantly, Thr3-NH likewise experiences a large upfield shift, despite the greater inductive effect of the ester group, which, for comparison, leads to a downfield shift of Glu5-NH. The sensitivity of an amide proton to the identity of the C5 hydrogen-bond acceptor strongly suggests that C5 hydrogen bonds operate in folded proteins.

To evaluate the importance of the C5 interaction specifically, we eliminated the canonical hydrogen-bond donor by replacing it with an ester (TrpZip2-C). In the absence of the canonical donor, replacement of an amide acceptor with an ester attenuates the C5 hydrogen bond of interest selectively (TrpZip2-D). Circular dichroism spectra indicated this attenuation reduces structural content (Supplementary Fig. 8). Thermal denaturation demonstrated that TrpZip2-D also loses all measurable structure by 50 °C, while TrpZip2-C retains some residual structure at temperatures up to ~75 °C (Fig. 3c). Because introduction of the second ester attenuates the interstrand oxygen–oxygen repulsion created by hydrogen bond deletion[30], the observed decrease in the thermal stability of TrpZip2-D relative to TrpZip2-C might underestimate the contributions of C5 hydrogen bonding.

Though suggestive, results from analysis of TrpZip2-C and TrpZip2-D are clouded by the low stability of these model proteins. Moreover, attenuation of the C5 hydrogen-bond acceptor also necessarily perturbs the proposed interstrand C^{α}–H···O=C hydrogen bond in β-sheets[31], potentially confounding our results. To enhance the stability of backbone-modified TrpZip2 and to isolate C5 interactions from other noncovalent interactions, we probed these interactions at the termini of the hairpin rather than in its center. Terminal modifications have been found to be less disruptive[32], and the C-terminal carbonyl group cannot accept a C^{α}–H···hydrogen bond. Elimination of the N-terminal serine residue leaves the C-terminal carbonyl group without a canonical hydrogen-bond donor and isolates the C5 hydrogen bond of interest (TrpZip2-E). Thermal denaturation revealed that attenuation of a terminal C5 hydrogen bond with an ester acceptor (TrpZip2-F) lowers the value of T_m (which is the temperature at the midpoint of the thermal transition between the folded and unfolded states) by 2 °C (Fig. 3d, Supplementary Table 3).

Finally, we attempted to enhance the strength of the C5 hydrogen bond by replacing the N-terminal acetyl group with a trifluoroacetyl group, which increases the acidity of the N-terminal hydrogen-bond donor. We did so in a variant in which the C5 hydrogen bond is isolated by removing the C-terminal lysine residue (TrpZip2-G and TrpZip2-H). We observed that the T_m value of TrpZip2 capped with a trifluoroacetyl group is 4 °C higher than when capped with an acetyl group (Fig. 3d, Supplementary Table 3), indicating that increasing the strength of only a single C5 hydrogen bond can increase the global stability of a β-sheet.

C5 hydrogen bonding in proteins

Given the prevalence of β-sheets in proteins, C5 hydrogen bonds en masse could contribute significantly to protein stability. To evaluate their potential contributions, we analyzed the energy of C5 hydrogen bonds by conducting NBO analyses on a set of AcGlyNHMe conformers that sample β-sheet geometry (Fig. 4a and Supplementary Table 7). We found a peak C5 hydrogen bond energy of 1.44 kcal/mol, with energies of at least 0.25 kcal/mol for residues with absolute backbone dihedral angles >140°; this region also corresponds to residues in sub-Å crystal structures with donor–acceptor distances <2.5 Å (Fig. 4b), which in turn correlated with several properties of hydrogen bonding in our computational analysis of AcGlyNHMe (Supplementary Fig. 2). Residues with absolute backbone dihedral angles >140° should therefore be expected to experience the effects of hydrogen bonding due to a C5 interaction. Considering only residues with absolute backbone dihedral angles >140°, we found that ~5% of all residues engage in C5 hydrogen bonding, and 94% of the proteins we examined contain at least one C5 hydrogen bond. Most of these residues (62%) were assigned to β-sheet secondary structure, and 13% of all β-sheet residues engage in C5 hydrogen bonds. Antiparallel β-sheets have a higher frequency (14%) than do parallel β-sheets (9%), consistent with differences in donor–acceptor distances (Supplementary Fig. 9a), suggesting that C5 hydrogen bonds might contribute to differences in the stability of these two architectures. The difference in C5 hydrogen-bond frequency between parallel and antiparallel β-sheets could be due to the influence of canonical hydrogen bonds, which pull the C5 donor and acceptor away from each in parallel β-sheets (Supplementary Fig. 9b). β-Bulges and bends each have C5 hydrogen-bond frequencies of ~10% and together make up

10% of residues with C5 interactions. The remaining interactions (27%) were not assigned to a secondary structure, suggesting that these interactions impart stability to irregular loops or turns. Residues engaged in C5 hydrogen bonds are also less solvent-exposed than are other residues (Supplementary Fig. 10), suggesting that C5 hydrogen bonds, like their canonical counterparts, must compete with water for the hydrogen-bonding potential of the protein backbone. We expect the effects of C5 hydrogen bonding to be most significant in the interiors of proteins.

To estimate the total contributions from C5 hydrogen bonds, we binned residues from a non-redundant set of high-resolution protein crystal structures by their backbone dihedral angles. We assigned an energy from our NBO calculations to each bin, and summed the energetic contributions, again only considering residues with absolute backbone dihedral angles >140°. Notably, the TrpZip2 model protein feature dihedral angles outside this range[29], so our criteria likely provide a lower bound to the expected contributions of C5 hydrogen bonds. We found that C5 hydrogen bonds contribute an average total stabilizing energy of ~4.5 kcal/mol per 100 residues. This energy is similar to contributions estimated for $n \rightarrow \pi^*$ interactions, which have energies generally greater than 0.27 kcal/mol each[23] and affect approximately one-third of residues in folded proteins[5], as well as for cation–π interactions which have energies generally greater than 2 kcal/mol each and affect at least one in every 77 residues[33]. Moreover, β-rich proteins could benefit from significantly higher contributions (Supplementary Fig. 11). Although these estimates do not account for differences in individual microenvironments, they do suggest a role for C5 hydrogen bonds in stabilizing folded proteins.

Discussion

We speculate that local C5 hydrogen bonds contribute to protein folding by directing the polyproline II secondary structure, which is believed to be common in the unfolded states of proteins[34], toward β-strand geometry. These local contacts could prepare strands for association into sheets, which are otherwise stabilized largely by nonlocal contacts. An overabundance of such interactions could, however, contribute to the formation of amyloids, as C5 hydrogen bonds are stronger when backbone dihedral angles mirror one another (Fig. 4a). Proteins tend to avoid this flat conformation[35,36]. Indeed, amyloid structures show highly symmetric backbone dihedral angles (Supplementary Fig. 12)[37], suggesting a significant contribution from C5 hydrogen bonds for this feature of protein folding landscapes[38]. Moreover, residues engaged in C5 hydrogen bonds are significantly enriched in small amino acids, most notably glycine (Supplementary Fig. 13a,b), polymers of which have a high propensity for aggregation[39]. Conversely, β-branched residues, which are abundant in natural β-sheets[40], are unlikely to engage in C5 hydrogen bonds. For example, whereas threonine is significantly more common in β-sheets than is serine, serine is much more likely to engage in C5 hydrogen bonds because the methyl group of threonine can clash with the main chain in conformations containing a C5 hydrogen bond (Supplementary Fig. 13c). Hence, the replacement of small amino acids with bulky ones during the evolution of β-sheets might, in part, serve to mitigate the propensity of C5 hydrogen bonds to encourage amyloid formation.

Our data indicate that C5 hydrogen bonds are common in folded proteins and affect their conformational stability. They complement other noncanonical interactions in protein structures, such as C^{α}–H···O hydrogen bonds[31], cation–π[33], and $n \rightarrow \pi^*$ interactions[5]. We conclude that C5 hydrogen bonds merit inclusion in computational force fields, which do not consider such interactions explicitly[41], despite their sharing many properties with traditional hydrogen bonds. Their accurate accounting could improve models of protein structure, stability, and folding.

Online Methods

Computational chemistry

All computations were performed with Gaussian 09 software from Gaussian (Wallingford, CT) and the B3LYP/6-311G+(2d,p) level of theory. A focused library of AcGlyNHMe conformations was constructed by optimizing the geometry of the compound in the gas phase for fixed values of the putative hydrogen-bond donor–acceptor distance (NH···O), ranging from d = 1.974 to 2.974 Å. This range was selected to sample distances observed in folded proteins. Optimized geometries were subjected to frequency analysis, gauge-independent atomic orbital NMR calculations, and analysis with NBO 5.9 software, which was from the Theoretical Chemistry Institute of the University of Wisconsin–Madison (Madison, WI) as implemented in Gaussian 09. All energies were corrected by the zero-point vibrational energy. Similarly, optimized gas-phase conformations of AcDegNHMe and AcDegOMe were obtained at the same level of theory. Frequency calculations yielded no imaginary frequencies, indicating a true stationary point on the potential energy surface. Energies were corrected by the zero-point vibrational energy. Atomic charges were determined by natural population analysis[42]. Calculated Cartesian coordinates are listed in Supplementary Note 1.

Diethylglycine derivatives and peptides

AcGlyNHMe was from Chem-Impex International (Wood Dale, IL). AcGlyOMe was from Sigma–Aldrich (St. Louis, MO). Derivatives of diethylglycine and TrpZip peptides were synthesized and characterized as described in Supplementary Note 2.

Fourier transform infrared spectroscopy

Fourier transfer infrared (FTIR) spectra were acquired with a Nicolet iS10 spectrometer from Thermo Fisher Scientific (Waltham, MA). Diethylglycine derivatives were dissolved in $CDCl_3$ to a final concentration of 10 mM. Following a purge with N_2(g), 256 scans between 400 and 4000 cm^{-1} were acquired and averaged, and the absorbance was calculated relative to background from neat $CDCl_3$.

Temperature-induced chemical shift

$AcDegNH_2$ was dissolved in methanol-d_4 to a final concentration of 10 mM, and ^{1}H NMR spectra were recorded over a temperature range of 20–40 °C. The per-degree change in chemical shift was then determined by linear regression of the observed chemical shifts as a function of temperature.

Hydrogen–deuterium exchange

Spectra for H/D exchange experiments were acquired with a DMX 400 MHz spectrometer from Bruker (Billerica, MA) in the National Magnetic Resonance Facility at Madison (NMRFAM). H/D exchange experiments were performed by co-dissolving AcDegNHMe and AcDegOMe, or AcGlyNHMe and AcGlyOMe in DMSO-d_6 to a final concentration of 50 mM each. Aliquots (0.50 mL) of the resulting solutions were transferred to NMR tubes. Exchange was initiated by the addition of 10 μL of D_2O at time $t = 0$. Samples were mixed thoroughly by repeated inversion for 30 s before collection of the first spectrum. ^{1}H NMR spectra were collected by averaging 16 individual scans to provide adequate signal to noise. Integrations were determined from the area of the calculated fit of the amide region (7.2–8.0 ppm for diethylglycines or 7.6–8.4 ppm for glycines), as determined with the program MestReNova 9.0 from MestreLab Research (Escondido, CA). Experiments were performed in triplicate.

NMR spectroscopy of peptides

NMR spectra of peptides were acquired with an Avance III 600 MHz spectrometer equipped with a 1.7-mm cryogenic probe from Bruker in the NMRFAM. As described previously[29], TrpZip peptides were dissolved to a final concentration of ~1 mM in 40 μL of 20 mM potassium phosphate buffer, pH 7.0, containing 10% v/v D_2O. Homonuclear ROESY, COSY, and TOCSY spectra were collected at 25 °C with water suppression, and resonances were assigned by standard methods. Specifically, sequential H^{α}_{i}–NH_{i+1} ROESY correlations provided connectivities and confirmed the β-like backbone geometry.

Circular dichroism (CD) spectroscopy

CD spectra were acquired with a Model 420 spectrometer from Aviv Biomedical (Lakewood, NJ) in the Biophysics Instrumentation Facility at the University of Wisconsin–Madison. As described previously[29], far-UV CD spectra of TrpZip peptides were obtained at 20 μM in 20 mM potassium phosphate buffer, pH 7.0. Ellipticity was recorded in 1-nm increments in the 200–300 nm range with 1-nm bandwidth and 3 s averaging. Thermal denaturation was performed at a concentration of 20 μM, except for TrpZip2-C and TrpZip2-D, which were denatured at 100 μM; for comparison, tryptophan zippers have been demonstrated previously to exist as monomers in solution at concentrations into the millimolar range[29]. Ellipticity at 228 nm was recorded by averaging for 15 s with a 1-nm bandwidth. Temperature was increased from 4 to 96 °C in 1-°C steps with a 5-min equilibration between steps. The overall rate of temperature increase was 6–10 °C/h. Data for peptides that were well folded at 4 °C were converted to percent folded. Values of T_m, which is the temperature at the midpoint of the thermal transition between the folded and unfolded states, were determined as described previously[43]. All ellipticity data fitted well to a two-state model ($R^2 > 0.999$), and the ensuing T_m values were independent of the experimentally determined concentration.

Bioinformatics

A nonredundant set (<25% pairwise sequence identity) of 192 protein crystal structures (>40 residues, R <20%) with a resolution of 1.0 Å or better was culled from the PDB of 28

November 2012 using the PISCES server[44], as was a similar set of 1,884 crystal structures determined to a resolution of 1.6 Å or better. Secondary structure assignments were made using Kabsch and Sander criteria as implemented with the program PROMOTIF[45]. Relative backbone solvent-accessible surface areas were calculated with the program NACCESS 2.1.1 from the Department of Biochemistry and Molecular Biology of the University College London (London, UK). Residues with backbone atoms modeled in multiple conformations were excluded from analysis. Parallel and antiparallel β-sheets were identified as those forming only parallel or antiparallel contacts, respectively. Strands engaged in mixed contacts were excluded from analysis for the purposes of comparing C5 frequencies between parallel and antiparallel β-sheets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank G.J. Bartlett, I.C. Tanrikulu, and L.L. Kiessling for discussions, and W.M. Westler, T. Zhang, and M.T. Zanni for assistance with spectroscopy. This work was supported by grants R01 AR044276 (NIH), R01 GM044783 (NIH), and CHE-1124944 (NSF). R.W.N. was supported by Biotechnology Training Grant T32 GM008349 (NIH) and by an ACS Division of Organic Chemistry Graduate Fellowship. The National Magnetic Resonance Facility at Madison is supported by grant P41 GM103399 (NIH). High-performance computing is supported by grant CHE-0840494 (NSF). The Biophysics Instrumentation Facility at the University of Wisconsin–Madison was established with grants BIR-9512577 (NSF) and S10 RR013790 (NIH).

References

1. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science. 2012; 338:1042–1046. [PubMed: 23180855]
2. Dill KA. Dominant forces in protein folding. Biochemistry. 1990; 29:7133–7155. [PubMed: 2207096]
3. Pauling L, Corey RB, Branson HR. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA. 1951; 37:205–211. [PubMed: 14816373]
4. Pauling L, Corey RB. The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci USA. 1951; 37:251–256. [PubMed: 14834147]
5. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
6. Clauss AD, et al. Rabbit-ears hybrids, VSEPR sterics, and other orbital anachronisms. Chem Educ Res Pract. 2014; 15:417–434.
7. Bartlett GJ, Choudhary A, Raines RT, Woolfson DN. $n \rightarrow \pi^*$ Interactions in proteins. Nat Chem Biol. 2010; 6:615–620. [PubMed: 20622857]
8. Bartlett GJ, Woolfson DN. On the satisfaction of backbone-carbonyl lone pairs of electrons in protein structures. Protein Sci. 2016; 25:887–897. [PubMed: 26833776]
9. Avignon M, Huong PV, Lascombe J, Marraud M, Neel J. Etude, par spectroscopie infra-rouge, de la confromation de quelques composés peptidiques modèles. Biopolymers. 1969; 8:69–89.
10. Burgess AW, Scheraga HA. Stable conformations of dipeptides. Biopolymers. 1973; 12:2177–2183. [PubMed: 4744757]
11. Toniolo C. Intramolecularly hydrogen-bonded peptide conformations. Crit Rev Biochem. 1980; 9:1–44. [PubMed: 6254725]

12. Dian BC, et al. The infrared and ultraviolet spectra of single conformations of methyl-capped dipeptides: *N*-Acetyl tryptophan amide and *N*-acetyl tryptophan methyl amide. J Chem Phys. 2002; 117:10688–10702.
13. Blanco S, Lesarri A, Lopez JC, Alonso JL. The gas-phase structure of alanine. J Am Chem Soc. 2004; 126:11675–11683. [PubMed: 15366915]
14. González Flórez AI, et al. Charge-induced unzipping of isolated proteins to a defined secondary structure. Angew Chem Int Ed. 2016; 55:3295–9299.
15. Zimmerman SS, Pottle MS, Nemethy G, Scheraga HS. Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. Macromolecules. 1977; 10:1–9. [PubMed: 839855]
16. Scheiner S. Relative strengths of NH···O and CH···O hydrogen bonds between polypeptide chain segments. J Phys Chem B. 2005; 109:16132–16141. [PubMed: 16853050]
17. Steiner T. The hydrogen bond in the solid state. Angew Chem Int Ed. 2002; 41:48–76.
18. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol. 1994; 238:777–793. [PubMed: 8182748]
19. Rowland RS, Taylor R. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii. J Phys Chem. 1996; 100:7384–7391.
20. Reed AE, Curtiss LA, Weinhold F. Intermolecular interactions from a natural bond orbital, donor–acceptor viewpoint. Chem Rev. 1988; 88:899–926.
21. Hinderaker MP, Raines RT. An electronic effect on protein structure. Protein Sci. 2003; 12:1188–1194. [PubMed: 12761389]
22. Choudhary A, Gandla A, Krow GR, Raines RT. Nature of amide carbonyl–carbonyl interactions in proteins. J Am Chem Soc. 2009; 131:7244–7246. [PubMed: 19469574]
23. Newberry RW, VanVeller B, Guzei IA, Raines RT. $n \rightarrow \pi^*$ Interactions of amides and thioamides: Implications for protein stability. J Am Chem Soc. 2013; 135:7843–7846. [PubMed: 23663100]
24. Benedetti E, et al. Structural versatility of peptides from $C^{\alpha,\alpha}$-dialkylated glycines. I A conformational energy computation and X-ray diffraction study of homo-peptides from $C^{\alpha,\alpha}$-diethylglycine. Biopolymers. 1988; 27:357–371.
25. Toniolo C, et al. Structural versatility of peptides from $C^{\alpha,\alpha}$-dialkylated glycines. II An IR absorption and ^{1}H-NMR study of homo-oligopeptides from $C^{\alpha,\alpha}$-diethylglycine. Biopolymers. 1988; 27:373–379.
26. Deechongkit S, et al. Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics. Nature. 2004; 430:101–105. [PubMed: 15229605]
27. Dado GP, Gellman SH. Structural and thermodynamic characterization of temperature-dependent changes in the folding pattern of a synthetic triamide. J Am Chem Soc. 1993; 115:4228–4245.
28. Skinner JJ, Lim WK, Bedard S, Black BE, Englander SW. Protein dynamics viewed by hydrogen exchange. Protein Sci. 2012; 21:996–1005. [PubMed: 22544544]
29. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: Stable, monomeric β-hairpins. Proc Natl Acad Sci USA. 2001; 98:5578–5583. [PubMed: 11331745]
30. Fu Y, Gao J, Bieschke J, Dendle MA, Kelly JW. Amide-to-*E*-olefin versus amide-to-ester backbone H-hond perturbations: Evaluating the O–O repulsion for extracting H-bond energies. J Am Chem Soc. 2006; 128:15948–15949. [PubMed: 17165703]
31. Derewenda ZS, Lee L, Derewenda U. The occurrence of C–H···O hydrogen bonds in proteins. J Mol Biol. 1995; 252:248–262. [PubMed: 7674305]
32. Culik RM, Jo H, DeGrado WF, Gai F. Using thioamides to site-specifically interrogate the dynamics of hydrogen bond formation in β-sheet folding. J Am Chem Soc. 2012; 134:8026–8029. [PubMed: 22540162]
33. Gallivan JP, Dougherty DA. Cation–π interactions in structural biology. Proc Natl Acad Sci USA. 1999; 98:9459–9464.
34. Shi Z, Chen K, Liu Z, Kallenbach NR. Conformation of the backbone in unfolded proteins. Chem Rev. 2006; 106:1877–1897. [PubMed: 16683759]
35. Richardson JS, Richardson DC. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA. 2002; 99:2754–2759. [PubMed: 11880627]

36. Cheng PN, Pham JD, Nowick JS. The supramolecular chemistry of β-sheets. J Am Chem Soc. 2013; 135:5477–5492. [PubMed: 23548073]
37. Sawaya MR, et al. Atomic structures of amyloid cross-β spines reveal varied steric zippers. Nature. 2007; 447:453–457. [PubMed: 17468747]
38. Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci. 1999; 24:239–332.
39. Ohnishi S, Kamikubo H, Onitsuka M, Kataoka M, Shortle D. Conformational preference of polglycine in solution to elongated structure. J Am Chem Soc. 2006; 128:16338–16344. [PubMed: 17165789]
40. Minor DL Jr, Kim PS. Measurement of the β-sheet-forming propensities of amino acids. Nature. 1994; 367:660–663. [PubMed: 8107853]
41. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol. 2003; 326:1239–1259. [PubMed: 12589766]
42. Reed AE, Weinstock RB, Weinhold F. Natural population analysis. J Chem Phys. 1985; 83:735–746.
43. Huang R, et al. Cross-strand coupling and site-specific unfolding thermodynamics of a Trpzip β-hairpin peptide using ^{13}C isotope labeling and IR spectroscopy. J Phys Chem B. 2009; 113:5661–5674. [PubMed: 19326892]
44. Wang G, Dunbrack RL. PISCES: A protein sequence culling server. Bioinformatics. 2003; 19:1589–1591. [PubMed: 12912846]
45. Hutchinson EG, Thornton JM. PROMOTIF—A program to identify and analyze structural motifs in proteins. Protein Sci. 1996; 5:212–220. [PubMed: 8745398]

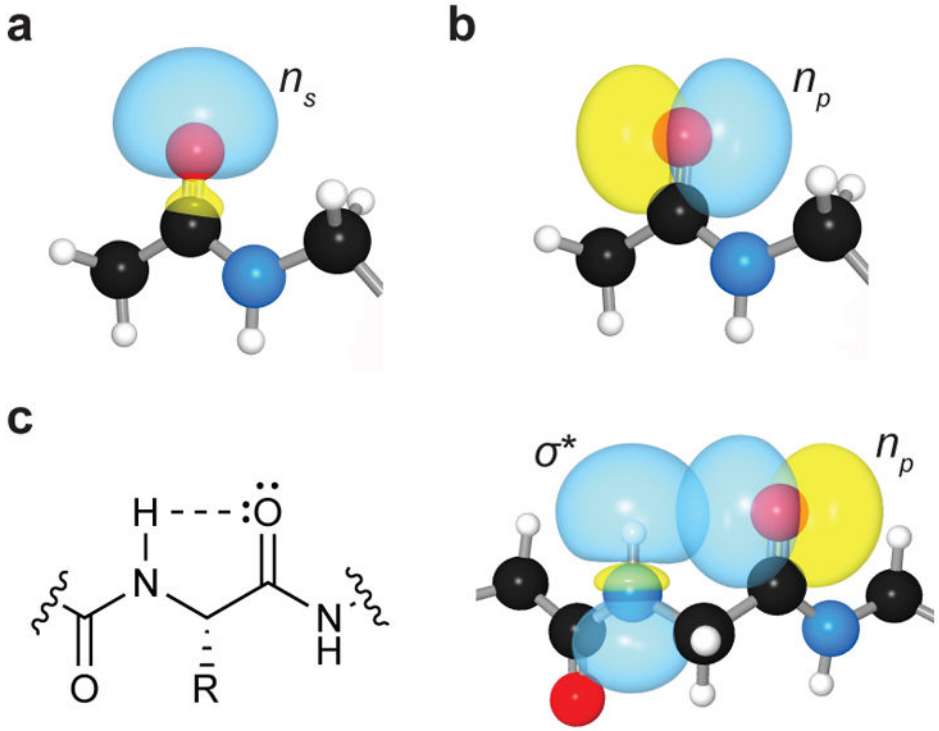

Figure 1. Amide carbonyl lone pairs allow for C5 hydrogen bonding

a, **b**, *s*-Type (**a**) and *p*-type (**b**) carbonyl lone pairs. **c**, Putative C5 hydrogen bond, characterized by overlap of the *p*-type carbonyl lone pair and N–H σ^* orbital.

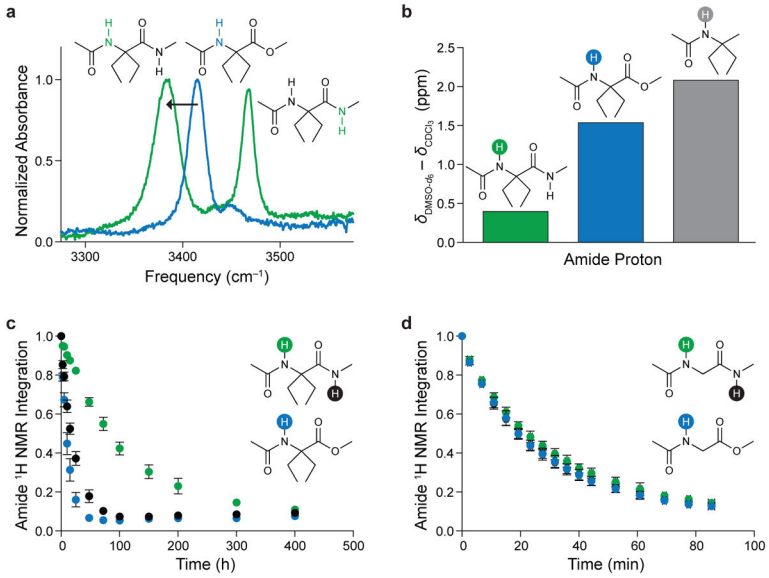

Figure 2. C5 Interactions display properties typical of hydrogen bonds

a, N–H stretching region of FTIR spectra of AcDegNHMe (**1**) and AcDegOMe (**2**) in $CDCl_3$. **b**, Change in chemical shift of amide protons of AcDegNHMe, AcDegOMe, and 3-acetamido-3-methylpentane (**3**) between $CDCl_3$ and DMSO-d_6 solutions. **c**, **d**, Integration of donor amide ^{1}H NMR signals of AcDegNHMe and AcDegOMe (**c**), or AcGlyNHMe (**4**) and AcGlyOMe (**5**) (**d**) in DMSO-d_6 over time following the addition of D_2O. Error bars in panels **c** and **d** are the standard deviation of experiments performed in triplicate.

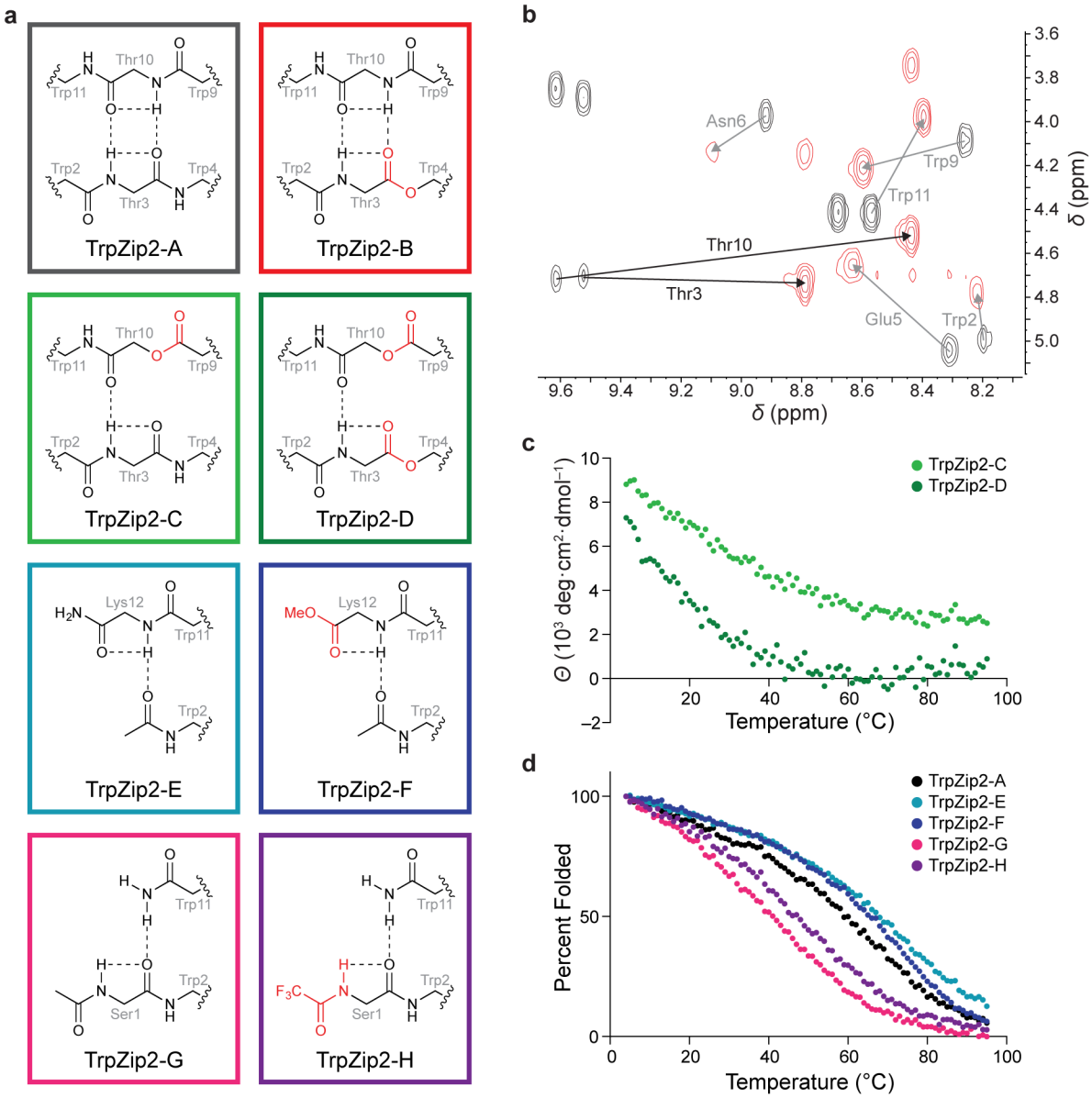

Figure 3. Perturbation of C5 hydrogen bonds impacts β-sheet stability

a, TrpZip2 peptides. **b**, NH–H$^{\alpha}$ region of ^{1}H–^{1}H TOCSY spectra of TrpZip2-A and TrpZip2-B. **c**, **d**, Thermal denaturation of TrpZip2 peptides based on measurements of ellipticity at 228 nm.

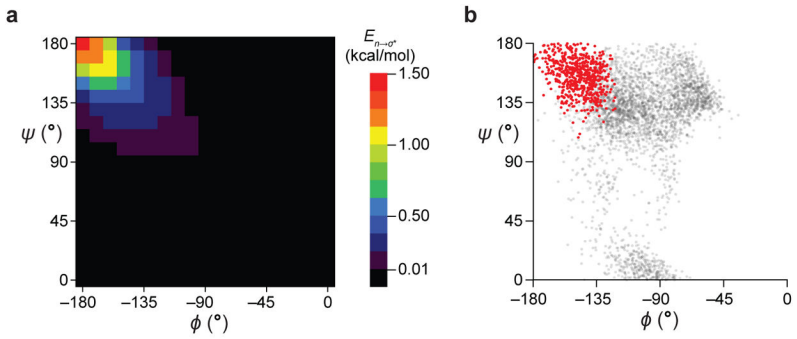

Figure 4. Energy and frequency of C5 hydrogen bonds in proteins

a, Ramachandran plot of the energy of C5 hydrogen bonds. **b**, Ramachandran plot of residues in sub-Å protein crystal structures with reported hydrogen coordinates. Residues with donor–acceptor distances <2.5 Å are shown in red.