Topological Evidence of Previously Overlooked N_{i+1}-H---N_i H-Bonds and their Contribution to Protein Structure and Stability

Matthew Holcomb, Ramkrishna Adhikary, Jörg Zimmermann, and Floyd E. Romesberg*

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA

ABSTRACT: Hydrogen-bonds (H-bonds) between backbone N-H donors and CO acceptors are central to our understanding of protein structure and stability. However, while interactions between backbone N atoms and the N-H of the following residue are also common, they have been ignored as potential H-bonds due to their bent geometry and the assumption that the amide N is a poor H-bond acceptor. Recently, we reported indirect experimental evidence that these interactions constitute functional H-bonds. We now report a combined AIM and NCI theoretical analysis of electron density that unambiguously supports the characterization of these interactions as H-bonds. The calculations further suggest that the N_{i+1}-H---N_i H-bonds are largely electrostatic in nature, and importantly, that they make a significant contribution to stability. Thus, given their apparently frequent occurrence, N_{i+1}-H---N_i H-bonds likely make critical, but previously unrecognized, contributions to protein structure and function.

INTRODUCTION

Since Pauling's elucidation of the α-helix and β-sheet structures, hydrogen-bonds (H-bonds) between backbone N-H donors and CO acceptors have been understood as major determinants of protein secondary structure. In general, secondary structure is commonly represented with a Ramachandran plot, which plots the observed torsion angle about the C_α-N bond (ϕ) against that about the C_α-C bond (ψ) for each residue. While regions of the plot that are most populated correspond to α-helices and β-sheets, as well as to turn structural elements, residues are also commonly found to lie within the "bridge region," where ϕ and ψ are around ±90° and 0°, respectively. The bridge region is traditionally considered unfavorable due to a steric clash between the amide nitrogen, N_i, and the nitrogen of the following amino acid, N_{i+1};[1,2] however, the distances between N_i and the N_{i+1} hydrogen (N_{i+1}-H) are within those normally considered stabilizing,[1] and the proton appears positioned to interact with the N_i electron density. These interactions were first noted by Pohl in 1971,[3] and in 1980, Gieren and coworkers noted that the ϕ, ψ and N-C_α-C angles appeared to be correlated in a manner that preserves the interaction.[4] This conclusion was confirmed through the analysis of a larger number of proteins by Scarsdale[5] and again by Karplus.[6] Soon afterwards, evidence that the same N_{i+1}-H---N_i interactions contribute to the catalysis of proline isomerization was presented.[7,8] However, after these reports, these interactions received virtually no additional attention.

We recently reported IR data and Natural Bond Order (NBO) calculations suggesting that these N_{i+1}-H---N_i interactions constitute stabilizing H-bonds.[9] Since our original report, additional bioinformatics studies and NBO calculations have supported the categorization of these interactions as H-bonds.[10,11] Nonetheless, the 92 to 107° angle of the N_{i+1}-H---N_i interactions (as well as the assumption that amide nitrogens are poor H-bond acceptors) might appear to call into question this characterization. At least near-linearity is commonly used as a criterion for identifying H-bonds in proteins, based on the geometrical requirements of the forces assumed to underlie them.[12] For example, some degree of covalency is often ascribed to H-bonds due to charge transfer from the acceptor orbital to the σ* orbital of the donor bond, and overlap of these orbitals is maximal with a linear geometry. In addition, electrostatics, which is commonly thought to make the most important contribution to H-bond formation, is also thought to favor linearity because it allows for optimal dipole alignment. However, focusing on dipole-dipole interactions is only an approximation of the electrostatic interaction, because it ignores the contribution of higher order multipoles, such as dipole-quadrupole and quadrupole-quadrupole interactions, which are also known to contribute to H-bond formation, and which favor more bent geometries.[13–15] Moreover, due to geometric constraints, intramolecular H-bonds often assume similarly bent structures and have been referred to as "weak" H-bonds.[7,16] Despite their characterization as weak, these bent H-bonds can contribute up to 4 kcal/mol to stability.[16] Indeed, bent backbone N-H---O H-bonds are known to influence peptide conformational preferences,[17] as are amide N-H---π bonds.[18,19] Thus, given the common occurrence of N_{i+1}-H---N_i interactions in proteins, and the fact that the native fold of a protein is typically only stabilized ~5 to 15 kcal/mol relative to misfolded or unfolded states,[20] if they are indeed stabilizing H-bonds, then they could have a profound effect on stability, structure, and function.

RESULTS AND DISCUSSION

To further explore whether the N_{i+1}-H---N_i interactions constitute H-bonds, we first employed Bader's Atoms in Molecules (AIM) theory.[21] This method is rooted in a topological analysis of electron density (ρ) and is well validated for the detection of H-bonds and estimation of their strength and degree of covalency. We began with the N_{i+1}-H---N_i interactions at residues Pro165 and Pro185 of the N-terminal Src homology 3 domain from the human CrkII adaptor protein (nSH3). The structure of nSH3[22] clearly reveals that the amide N of each Pro residue acts as the N_i of an

N_{i+1}-H---N_i interaction (Figure 1) and these residues were the focus of our original experimental studies.[9] The coordinates of the heavy atoms of each Pro residue and its N_{i+1} amide nitrogen were taken from the structure coordinates (PDB ID 1CKA), the N_i was capped with an acetyl group, the N_{i+1} was capped with a methyl group, and hydrogen atoms were then added to the resulting 1-acetyl-*N*-methylpyrrolidine-2-carboxamide structure. The structures were subjected to a constrained optimization at the MP2/aug-cc-pVDZ level using Gaussian09,[23] with the angles between heavy atoms fixed, and the resultant wavefunctions were subjected to topological analysis using Multiwfn.[24] Interestingly, in each case a bond critical point (BCP), which is a necessary and sufficient condition for the existence of a bond within the AIM framework, was detected between the N_i and N_{i+1} H atoms (Figure 2). The ρ and positive Laplacian values are consistent with weak to moderate H-bonds with little to no covalent character[25] (Table S1).

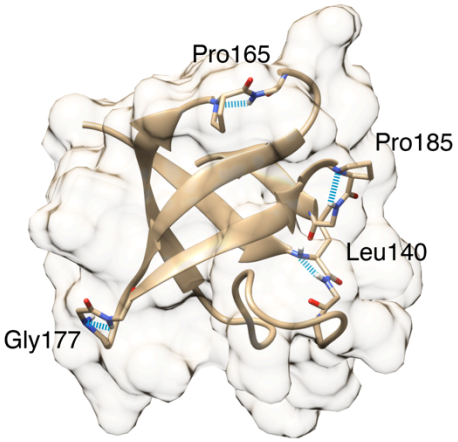

Figure 1. Structure of nSH3 with potential N_{i+1}-H---N_i H-bonds at Pro165, Pro185, Leu140, and Gly177 indicated (PDB ID 1CKA).

To further support the characterization of these N_{i+1}-H---N_i interactions as H-bonds, we employed the criteria of Koch and Popelier,[26] using *N*-methylacetamide and *N*-acetylpyrrolidine calculated at the same level of theory for the required reference states (Table S1). Comparison revealed mutual penetrance, hydrogen atom destabilization, a reduction of hydrogen atom charge, a reduction of the hydrogen dipole moment, and a significant reduction in the volume of the hydrogen atom within the protein N_{i+1}-H---N_i interactions, thereby satisfying all of the criteria for an H-bond.

To ensure that the presence of BCPs was not an artifact of the crystal structure, we optimized without any constraints the structure of 1-acetyl-*N*-methylpyrrolidine-2-carboxamide, and found a minimum with a clear N-H---N interaction. In contrast to an earlier report focused on a similar system performed at a lower level of theory,[27] we observed a BCP with properties that are qualitatively identical to those observed in the protein-derived structures (Supporting Information). This suggests that the BCPs detected with the N_{i+1}-H---N_i interactions in nSH3 are not artifacts of the crystal structure.

Our original experimental evidence that the N_{i+1}-H---N_i interactions constitute H-bonds was based on the absorptions of site-selectively incorporated carbon-deuterium (C-D) bonds, and specifically, $C_\delta D_2$ absorptions of Pro residues as they provide convenient and sensitive observables.[9] However, even a cursory examination of any protein structure immediately reveals that the N_{i+1}-H---N_i interactions are not limited to Pro residues. Indeed, an analysis of nSH3 reveals up to 16 such interactions.[28] Thus, we extended our theoretical studies to include two additional interactions in nSH3 involving N_i residues Gly177 and Leu140 (Figure 1), which are representative of the backbone diversity of the natural amino acids. These structures were treated in the same manner as described above for the Pro residues, with Leu140 truncated to 2-acetamido-*N*-methylpropanamide, thus replacing the isobutyl side chain with a methyl group, and Gly177 modeled as 2-acetamido-*N*-methylacetamide (Figure 2). In these cases, *N*-methylacetamide was used as a reference for both the H and N_i atoms. Again, BCPs were detected in both structures, with very similar electron densities and Laplacians as those observed with the interactions at Pro. In addition, all of the H-bonding criteria established by Koch and Popelier were again satisfied by these interactions (Table S1). We note that the bond path in the case of Gly177 terminates at the N_{i+1} nuclear critical point rather than the H nuclear critical point, but this likely reflects the hyper-sensitivity of BCPs and their associated paths to small perturbations in ρ.[29]

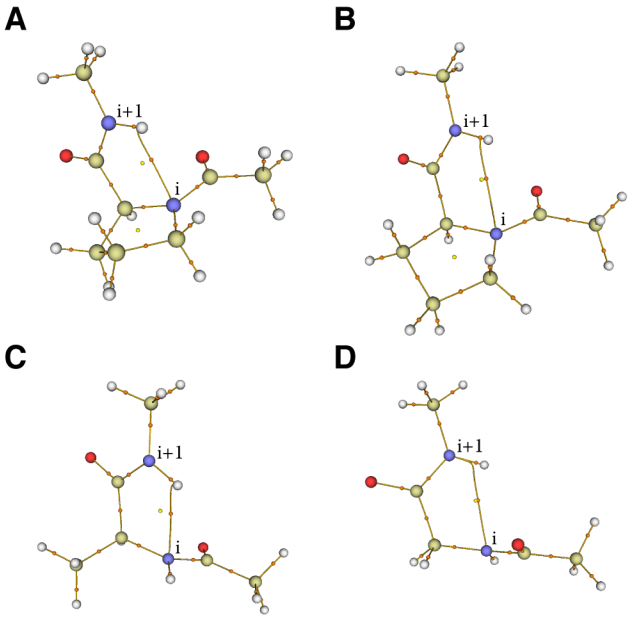

Figure 2. Molecular graphs of the N_{i+1}-H---N_i interactions at (A) Pro165, (B) Pro185, (C) Leu140, and (D) Gly177. Beige, C; red, O; blue, N; white, H; small orange circle, BCP; small yellow circle, ring critical point, a feature of rings of bonded atoms[21] (partially obscured in the perspective shown in D).

To more robustly evaluate the N_{i+1}-H---N_i interactions, we next performed Non-Covalent Interaction (NCI) analysis,[30] which, although founded in the same theory as AIM, has been demonstrated to be a more reliable means of characterizing intramolecular H-bonding.[29] In particular, NCI was demonstrated to be useful in the topological analysis of 1,2-ethandiol, 3-hydroxytetrahydropyran, methyl lactate, and aminoalcohols, all of which form intramolecular bonds resulting in 5-membered rings that are geometrically similar to the N_{i+1}-H---N_i H-bonds examined here.[29,31–33] Rather than focusing on critical points and the paths connecting them to characterize bonding, NCI identifies regions of non-covalent interaction based on the normalized and unitless reduced density gradient ($s=1/[2(3\pi^2)^{1/3}]|\nabla\rho|/\rho^{4/3}$) and ρ. These interactions can then be characterized as attractive or repulsive

according to the second derivative of ρ, and the strength of the interaction is taken to be proportional to ρ. Using Multiwfn to generate the isosurfaces for all four residues examined, we found that in each case there is a clear attractive region between the N_{i+1}-H and N_i atoms, as well a region of repulsive interactions within the resulting five-membered ring, indicating the presence of the same steric interactions observed with other similar intramolecular H-bonded structures[29,31] (Figure 3). Thus, each of the N_{i+1}-H---N_i interactions, including at Gly177, is predicted by NCI to be an H-bond.

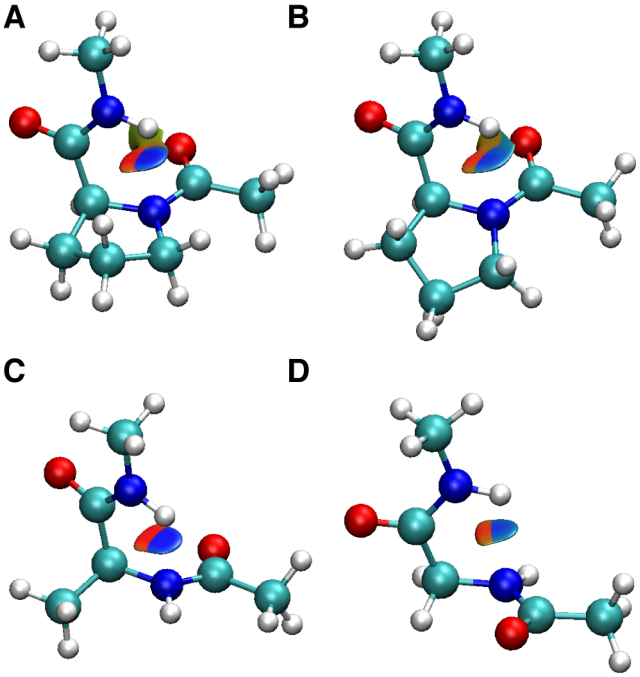

Figure 3. NCI isosurfaces (s = 0.5) of the N_{i+1}-H---N_i interactions at (A) Pro165, (B) Pro185, (C) Leu140 and (D) Gly177. Color corresponds to a blue (attractive)-green-red (repulsive) scale from -0.02 < sign(λ_2)ρ < 0.02. Isosurfaces were generated using 512,000 points in a 1.5 × 1.5 × 1.5 Å cube centered at the midpoint of the N_{i+1}-H---N_i interaction.

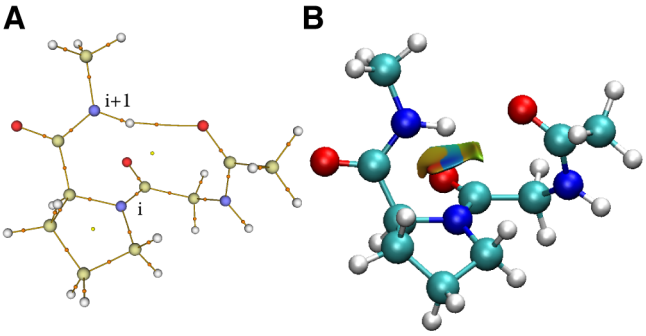

Figure 4. (A) Molecular graph and (B) NCI isosurface of the N_{i+1}-H---N_i interaction at Pro185 with the additional oxygen H-bond acceptor observed in the crystal structure.

As noted in our original report,[9] many of the N_{i+1}-H---N_i interactions appear to be part of a bifurcated H-bond that includes a second H-bond acceptor, such as another amide carbonyl or side chain heteroatom. More recent works by others have also suggested that such bifurcation is common.[11] Indeed, the N_{i+1}-H---N_i interaction at Pro185 in nSH3 appears to be part of a bifurcated H-bond that includes the backbone oxygen of Pro183. To examine the effects of bifurcation on this N_{i+1}-H---N_i H-interaction, we performed the same AIM and NCI analysis with the Ac-Gly-Pro-NMe model dipeptide, which includes the second H-bond acceptor (Figure 4). With this structure, no BCP associated with an N_{i+1}-H---N_i H-bonding interaction is observed. However, the NCI isosurface for the bifurcated N_{i+1}-H---N_i interaction is virtually identical to its unbifurcated counterpart described above, with an attractive surface between the N_i acceptor and N_{i+1}-H donor that reflects a stabilizing interaction. (Figure 4). Moreover, we found that a plot of *s* versus sign(λ_2)ρ (the sign of the second eigenvalue of the Hessian matrix weighted by ρ) showed no difference in the ordinate value of the observed low *s*, low sign(λ_2)ρ spike (a measure of the presence and strength of an attractive interaction[29,30]) for the bifurcated and unbifurcated cases (Figure 5, Table S1). Thus, while bifurcation formally prevents the gradient from vanishing to zero, it does not eliminate the local minimum in *s* or even significantly change its position. Following the observation of Lane and coworkers that minima in *s* exist on a continuum between those that are detected as BCPs by AIM and those that are not[29] and that the ordinate of the minima is predictive of strength regardless of whether or not the gradient vanishes,[33] we treated the minimum in *s* as a BCP and found ρ and Laplacian values that were virtually identical to those found in the unbifurcated case. Moreover, with an *N*-methylacetamide dimer with the same N-H---O=C distance and internal angles as the reference structure, the N_{i+1}-H---N_i interaction again meets all of the Koch and Popelier criteria for being characterized as an H-bond (Table S1).

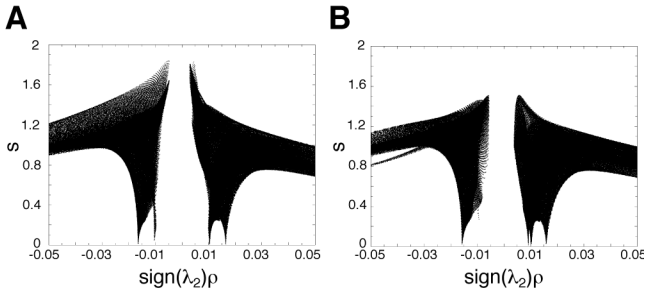

Figure 5. NCI sign(λ_2)ρ vs s plots of the N_{i+1}-H---N_i interactions at Pro185 (A) without and (B) with bifurcation.

Having validated the classification of the N_{i+1}-H---N_i interactions as H-bonds, we finally estimated their stability via the relationship between ρ and stabilization energy detailed by Sathyamurthy and coworkers.[34] This method predicts that each N_{i+1}-H---N_i H-bond contributes 3 to 5 kcal/mol to the stability of the protein (Supporting Information).

CONCLUSIONS

H-bonds between backbone N-H and CO moieties are ubiquitous in proteins, and their contribution to structure is undisputed. While originally noted over 45 years ago,[3] the less ubiquitous but still common N_{i+1}-H---N_i interactions have received almost no attention. However, the combined AIM and NCI analysis presented here, along with our previous experimental study,[9] indicates that the N_{i+1}-H---N_i interactions also constitute stabilizing H-bonds. They appear to participate in both conventional H-bonds and bifurcated H-bonds. At least in the case of bifurcation examined, the reliance of AIM on a strictly vanishing density gradient appears to makes NCI a more reliable approach for their detection, as has also

been recently observed with other H-bonding systems.[29,31,32] Overall, the ρ and positive Laplacian values indicate that the N_{i+1}-H---N_i H-bonds are primarily electrostatic, which is consistent with the calculated stabilization energies being significantly greater than the second order perturbation energies we calculated previously using NBO analysis,[9] and which were later reproduced by others.[11] Given that these interactions appear to be common in every protein, that their formation depends on backbone structure, and that a given protein conformation is only marginally stable relative to other conformations or even relative to the unfolded state, the N_{i+1}-H---N_i H-bonds are expected to make a substantial contribution to protein stability, structure, and, correspondingly, function. The further experimental and theoretical characterization of these previously overlooked H-bonds is currently in progress.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Supporting commentary, table, figures, and references (PDF)

AUTHOR INFORMATION

Corresponding Author

* floyd@scripps.edu

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (Grant No. NSF/MCB-1716864 to F.E.R.). We would like to acknowledge Dr. Tian Lu for instructive conversations on Multiwfn and topological analysis.

REFERENCES

(1) Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *23*, 283–438.

(2) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. Intrinsic Phi, Psi Propensities of Amino Acids, Derived from the Coil Regions of Known Structures. *Nat. Struct. Biol.* **1995**, *2*, 596–603.

(3) Pohl, F. M. Empirical Protein Energy Maps. *Nat. New Biol.* **1971**, *234*, 277–279.

(4) Gieren, A.; Dederer, B.; Schanda, F. Some Aspects Concerning Conformation of Polypeptide Chains in Proteins. *Z. Naturforsch. C* **1980**, *35*, 741–746.

(5) Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schaefer, L.; Momany, F. A. Ab Initio Studies of Molecular Geometries. 27. Optimized Molecular Structures and Conformational Analysis of N.Alpha.-Acetyl-N-Methylalaninamide and Comparison with Peptide Crystal Data and Empirical Calculations. *J. Am. Chem. Soc.* **1983**, *105*, 3438–3445.

(6) Karplus, P. A. Experimentally Observed Conformation-Dependent Geometry and Hidden Strain in Proteins. *Prot. Sci.* **1996**, *5*, 1406–1420.

(7) Cox, C.; Lectka, T. Intramolecular Catalysis of Amide Isomerization: Kinetic Consequences of the 5-NH- -N□ Hydrogen Bond in Prolyl Peptides. *J. Am. Chem. Soc.* **1998**, *120*, 10660–10668.

(8) Cox, C.; Young, V. G.; Lectka, T. Intramolecular Catalysis of Amide Isomerization. *J. Am. Chem. Soc.* **1997**, *119*, 2307–2308.

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

(10) Dhar, J.; Kishore, R.; Chakrabarti, P. A Novel Secondary Structure Based on Fused Five-Membered Rings Motif. *Sci Rep* **2016**, *6*, 31483.

(11) Deepak, R. N.; Sankararamakrishnan, R. Unconventional N-H...N Hydrogen Bonds Involving Proline Backbone Nitrogen in Protein Structures. *Biophys. J.* **2016**, *110*, 1967–1979.

(12) Lipsitz, R. S.; Sharma, Y.; Brooks, B. R.; Tjandra, N. Hydrogen Bonding in High-Resolution Protein Structures: A New Method to Assess NMR Protein Geometry. *J. Am. Chem. Soc.* **2002**, *124*, 10621-10626.

(13) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, UK, 2013.

(14) Blasiak, B.; Ritchie, A. W.; Webb, L. J.; Cho, M. Vibrational Solvatochromism of Nitrile Infrared Probes: Beyond the Vibrational Stark Dipole Approach. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18094–18111.

(15) Kim, H.; Cho, M. Infrared Probes for Studying the Structure and Dynamics of Biomolecules. *Chem. Rev.* **2013**, *113*, 5817–5847.

(16) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, NY, 1997.

(17) Newberry, R. W.; Raines, R. T. A Prevalent Intraresidue Hydrogen Bond Stabilizes Proteins. *Nat. Chem. Biol.* **2016**, *12*, 1084–1088.

(18) Walsh, P. S.; Blodgett, K. N.; McBurney, C.; Gellman, S. H.; Zwier, T. S. Inherent Conformational Preferences of Ac-Gln-Gln-Nhbn: Sidechain Hydrogen Bonding Supports a □-Turn in the Gas Phase. *Angew Chem Int Ed Engl* **2016**, *55*, 14618–14622.

(19) Sohn, W. Y.; Brenner, V.; Gloaguen, E.; Mons, M. Local NH-□ Interactions Involving Aromatic Residues of Proteins: Influence of Backbone Conformation and □□* Excitation on the □ H-Bond Strength, as Revealed from Studies of Isolated Model Peptides. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29969–29978.

(20) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. A Backbone-Based Theory of Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623–16633.

(21) Bader, R. F. W. *A Quantum Theory*; Oxford University Press: London, 1990.

(22) Wu, X.; Knudsen, B.; Feller, S. M.; Zheng, J.; Sali, A.; Cowburn, D.; Hanafusa, H.; Kuriyan, J. Structural Basis for the Specific Interaction of Lysine-Containing Proline-Rich Peptides with the N-Terminal SH3 Domain of c-Crk. *Structure* **1995**, *3*, 215–226.

(23) Frisch, M. J., *et al. Gaussian 09*, Wallingford, CT, 2016.

(24) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comp. Chem.* **2012**, *33*, 580–592.

(25) Grabowski, S. J. What Is the Covalency of Hydrogen Bonding? *Chem. Rev.* **2011**, *111*, 2597–2625.

(26) Koch, U.; Popelier, P. L. A. Characterization of C-H-O Hydrogen Bonds on the Basis of the Charge Density. *J. Phys. Chem.* **1995**, *99*, 9747–9754.

(27) Ramek, M.; Kelterer, A.-M.; Teppen, B. J.; Schäfer, L. Theoretical Structure Investigations of N-Acetyl-L-Proline Amide. *J. Mol. Struct.* **1995**, *352*, 59–70.

(28) Adhikary, R.; Zimmermann, J.; Romesberg, F. E. Transparent Window Vibrational Probes for the Characterization of Proteins with High Structural and Temporal Resolution. *Chem. Rev.* **2017**, *117*, 1927–1969.

(29) Lane, J. R.; Contreras-García, J.; Piquemal, J.-P.; Miller, B. J.; Kjaergaard, H. G. Are Bond Critical Points Really Critical for Hydrogen Bonding? *J. Chem. Theory Comput.* **2013**, *9*, 3263–3266.

(30) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.

(31) Solha, D. C.; Barbosa, T. M.; Viesser, R. V.; Rittner, R.; Tormena, C. F. Experimental and Theoretical Studies of Intramolecular Hydrogen Bonding in 3-Hydroxytetrahydropyran: Beyond Aim Analysis. *J. Phys. Chem. A* **2014**, *118*, 2794–2800.

(32) Schroder, S. D.; Wallberg, J. H.; Kroll, J. A.; Maroun, Z.; Vaida, V.; Kjaergaard, H. G. Intramolecular Hydrogen Bonding in Methyl Lactate. *J Phys Chem A* **2015**, *119*, 9692–9702.

(33) Lane, J. R.; Schroder, S. D.; Saunders, G. C.; Kjaergaard, H. G. Intramolecular Hydrogen Bonding in Substituted Aminoalcohols. *J. Phys. Chem. A* **2016**, *120*, 6371-6378.

(34) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. Hydrogen Bonding without Borders: An Atoms-in-Molecules Perspective. *J. Phys. Chem. A* **2006**, *110*, 3349–3351.

TOC Graphic