PCCP

Solvation stabilizes intercarbonyl n→π* interactions and polyproline II helix

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-02-2022-000857.R1
Article Type:	Paper
Date Submitted by the Author:	19-Apr-2022
Complete List of Authors:	Zondlo, Neal; University of Delaware, Department of Chemistry and Biochemistry

Solvation stabilizes intercarbonyl n→π* interactions and polyproline II helix

Neal J. Zondlo*

Department of Chemistry and Biochemistry

University of Delaware

Newark, DE 19716

United States

* To whom correspondence should be addressed. email: *zondlo@udel.edu*, phone: +1-302-831-0197, fax +1-302-831-6335.

Abstract

n→π* interactions between consecutive carbonyls stabilize the α-helix and polyproline II helix (PPII) conformations in proteins. n→π* interactions have been suggested to provide significant conformational biases to the disordered states of proteins. To understand the roles of solvation on the strength of n→π* interactions, computational investigations were conducted on a model n→π* interaction, the twisted-parallel-offset formaldehyde dimer, as a function of explicit solvation of the donor and acceptor carbonyls, using water and HF. In addition, the effects of urea, thiourea, guanidinium, and monovalent cations on n→π* interaction strength were examined. Solvation of the acceptor carbonyl significantly strengthens the n→π* interaction, while solvation of the donor carbonyl only modestly weakens the n→π* interaction. The n→π* interaction strength was maximized with two solvent molecules on the acceptor carbonyl. Urea stabilized the n→π* interaction via simultaneous engagement of both oxygen lone pairs on the acceptor carbonyl. Solvent effects were further investigated in the model peptides Ac-Pro-NMe_2, Ac-Ala-NMe_2, and Ac-Pro_2-NMe_2. Solvent effects in peptides were similar to those in the formaldehyde dimer, with solvation of the acceptor carbonyl increasing n→π* interaction strength and resulting in more compact conformations, in both the proline *endo* and *exo* ring puckers, as well as a reduction in the energy difference between these ring puckers. Carbonyl solvation leads to an energetic preference for PPII over both the α-helix and β/extended conformations, consistent with experimental data that protic solvents and protein denaturants both promote PPII. Solvation of the acceptor carbonyl weakens the intraresidue C5 hydrogen bond that stabilizes the β conformation.

Introduction

The n→π* interaction between consecutive carbonyls has been recently identified to be an important force in protein structure (Figure 1).[1-8] The n→π* interaction involves the orbital overlap between a lone pair (n) of a donor carbonyl and the π* antibonding orbital of the acceptor carbonyl, which leads to stabilization of specific conformations via through-space electron delocalization. The n→π* interaction stabilizes compact conformations of proteins, including α-helix and polyproline II helix (PPII). n→π* interactions are also important in other diverse contexts, including the structure of organic molecules and stabilizing transition states in catalysis.[9-16]

The n→π* interaction has been implicated as a significant factor in the disordered states of proteins.[4,6,7,17] n→π* interactions stabilize the α-helix conformation locally (as an *i*/*i*+1 interaction), without[7] a need for hydrogen bonding (which involves *i*/*i*+3 or *i*/*i*+4 interactions for 3_{10} helices or α-helices, respectively). n→π* interactions provide an energetic driving force to partially counteract the substantial entropic cost of adopting the first turn of an α-helix. n→π* interactions specifically stabilize the PPII secondary structure, which forms despite lacking hydrogen bonds. Notably, PPII is a major conformation in the unfolded state of proteins.[18-22]

PPII is stabilized by the classical protein denaturant urea.[23-26] In addition, D_2O stabilizes PPII. The effects of D_2O and urea to stabilize PPII both occur through mechanisms that are not well understood.[27] Solvent interactions are globally important in the structure of PPII, with PPII disfavored in non-hydrogen-bonding solvents.[23,28-31] Herein, we sought to develop additional insights into the interplay between solvation, n→π* interaction strength, and structure, via computational analysis of model compounds and peptides.

Methods

Computational chemistry. Calculations were conducted with Gaussian 09.[32] Natural bond orbital (NBO) analysis was conducted using the NBO6 implementation in Gaussian09.[33,34] Models depicting orbital interactions were generated within GaussView 5. For visualization of orbitals, isovalues of 0.02 were used.

Analysis of n→π* interactions in small molecules. For models with formaldehyde, initial geometries were developed with DFT, using the M06-2X method and the 6-311++G(3d,3p) basis set.[35,36] These models were then subjected to further geometry optimization using the MP2

method and the 6-311++G(3d,3p) basis set, followed by additional geometry optimization using the MP2 method and the aug-cc-pVTZ basis set.[37-39] Geometry optimization of the complex with K^+ used the aug-cc-pVTZ basis set on H, C, and O atoms and the 6-311++G(3d,3p) basis set on K, as parameters are not defined for K in the aug-cc-pVTZ basis set. All geometry optimization calculations were conducted with implicit water, using the IEFPCM continuous polarization approach as implemented in Gaussian09.[40] Implicit solvation was critical to geometry optimization: the twisted-parallel-offset dimer was not a local energy minimum in vacuum,[13] presumably due to the large dipole moment (7.3 D) in this structure destabilizing this geometry in vacuum.

In all structures but H_2O•HCHO•••HCHO•H_2O, geometry optimization was conducted without restraints, and geometry optimization resulted in normal termination with low RMS error. However, for the H_2O•HCHO•••HCHO•H_2O model, an unrestrained geometry optimization resulted in the collapse of the two water molecules into a water cluster. Thus, for this model, a restrained optimization was used, using the lowest energy structure observed prior to water collapse for restraints. The related H_2O•HCHO••••HCHO•$2H_2O$ model, which was derived from an initial HF•HCHO•••HCHO•$2H_2O$ model, did not require the use of restrained

optimization. Frequency calculations were conducted on all final structures obtained, and no negative frequencies were identified unless otherwise indicated in the Supporting Information.

Complex interaction energy analysis. The structures derived from MP2/aug-cc-pVTZ optimization calculations were subjected to analysis of the energies of the complexes using two approaches. First, interaction energies were calculated using counterpoise calculations, which are conducted in the gas phase. Counterpoise calculations were used to identify an appropriate basis set of sufficient size to fully account for interaction energies while minimizing basis-set superposition error (BSSE) and allowing acceptable calculation times across all molecules, to allow the comparison of energies in different complexes using the same methods.[41,42] All counterpoise calculations were conducted using the MP2 method. The partially augmented basis set jul-cc-pV5Z, with a full set of diffuse functions (derived from the aug-cc-pV5Z basis set) on heavy atoms but lacking diffuse functions on hydrogens (cc-pV5Z basis set), was identified to be optimal for these analyses, with an acceptable combination of calculation time and BSSE.[43] Similar BSSE-corrected interaction energies were obtained using the aug-cc-pVQZ basis set (within 0.05 kcal mol^{-1}, with more favorable energies using jul-cc-pV5Z), which has diffuse functions on all atoms. However, the aug-cc-pVQZ basis set had significantly larger BSSE (up to

0.3 kcal mol^{-1}), which complicated subsequent analysis in implicit water. Counterpoise energies using the jul-cc-pV5Z basis set had BSSE of 0.05–0.10 kcal mol^{-1} for all complexes except H_2O-HCHO-HCHO-$2H_2O$ (0.13 kcal mol^{-1}) and HCHO-HCHO-H^+ (0.40 kcal mol^{-1}). The small, similar magnitude of the BSSE for all complexes using the jul-cc-pV5Z basis set allowed the analysis of relative complex energies in implicit water without concern of BSSE overly influencing the results.

Complex energies in implicit water were calculated via subtraction of the component MP2 energies using the jul-cc-pV5Z basis set ($\Delta E_{\text{interaction}} = \Delta E_{\text{int}} = E_{\text{complex}} - E_{\text{component1}} - E_{\text{component2}}$), using the energies of the individual components (solvated molecules of formaldehyde) that were optimized independently by the same methods (MP2/aug-cc-pVTZ/H_2O optimization). Thus, component interaction energies explicitly address changes in structure as a result of the interaction (e.g. bond lengths, bond angles, pyramidalization).

In addition to the analysis of all complexes by the MP2 method with the jul-cc-pV5Z basis set, a subset of structures was analyzed using the CCSD(T) method and the jul-cc-pV5Z basis set.[44,45] For all formaldehyde dimer complexes analyzed except those with H^+ or with Li^+, the energies determined by MP2 and by CCSD(T) were within 0.02 kcal mol^{-1} of each other

(Table S1), indicating that the MP2 method provides excellent accuracy compared to the substantially more expensive CCSD(T) method in the analysis of n→π* interaction energies.

Solvation effects on conformation in Ac-Pro-NMe_2 and Ac-Ala-NMe_2 peptides. Solvation was examined in the simple proline model compound Ac-Pro-NMe_2. Initial model structures, with either an *endo* or *exo* ring pucker, each in either a PPII conformation or an α-helix (α_R) conformation, were developed that were derived from prior DFT-based geometry optimization calculations.[7,17] In addition, model structures were similarly developed on Ac-Ala-NMe_2 molecules, in the α_R, PPII, and β/extended (C5-hydrogen-bonded) conformations. All structures had *trans* Ac-Pro or Ac-Ala amide bonds. To each of these structures was added 1-3 hydrogen-bonding molecules as models of solvation or chemical denaturants (urea, thiourea, guanidinium). HF models were used primarily in place of H_2O due to effects of H_2O molecules in promoting alternative structures in order to achieve additional hydrogen bonds to the unsatisfied hydrogen bond donor and acceptors of water. These results emphasize the limitations of implicit solvation models used in quantum chemistry. While the effects of HF as a hydrogen-bond donor solvent are inherently greater than those of water (*vide infra*), the absence of additional complications in geometry optimization (which with water resulted in hydrogen bonds

to the proline molecule [lower overall energy due to the strength of hydrogen bonds] in place of hydrogen bonds to solvent that would occur in a fully solvated molecule) rendered the HF solvation model advantageous at this level of calculation. The Ac-Pro-NMe_2•solvent molecule(s) and Ac-Ala-NMe_2•solvent molecule(s) structures were subjected to initial geometry optimization using the M06-2X DFT method and the 6-311++G(2d,2p) basis set, followed by subsequent optimization using the M06-2X method and the jun-cc-pVTZ basis set, all in implicit water (IEFPCM). All geometry optimization calculations terminated normally to generate structures with low error. Structures were analyzed by frequency calculations, and exhibited no imaginary (negative) frequencies unless otherwise indicated. The resultant structures were analyzed to determine energies using the MP2 method and the aug-cc-pVTZ basis set in implicit water.

Solvation effects on polyproline II helix conformation in Ac-Pro_2-NMe_2 peptides as a function of proline ring pucker. To further examine the effect of solvation on PPII conformation, including effects on propagation of PPII, models of Ac-Pro_2-NMe_2 were generated, with all residues in the PPII conformation and with all 4 combinations of proline *exo* and *endo* ring pucker. All structures had *trans* Ac-Pro and Pro-Pro amide bonds. To the initial models were added zero, one, or two molecules of HF solvation on any of the acetyl, Pro1, or Pro2 carbonyl.

These models were initially generated with the M06-2X method and the 6-311++G(d,p) basis set, followed by subsequent optimization with the same method and the 6-311++G(2d,2p) basis set. Structures were analyzed by frequency calculations, and exhibited no imaginary (negative) frequencies unless otherwise indicated. The energies of the final structures were then calculated using the MP2 method and the 6-311++G(3d,3p) basis set. All calculations were conducted in implicit water (IEFPCM).

Results

In order to examine the roles of solvation on the n→π* interaction via quantum chemistry methods, we initially identified a minimal interaction motif, the twisted-parallel-offset formaldehyde dimer. The twisted-parallel-offset formaldehyde dimer in these studies is aligned for an n→π* interaction, but has no other substantial noncovalent interactions (Figure 1c, Figure 2a), in contrast to the anti-parallel formaldehyde dimers (both head-to-tail and tail-to-head) that represent the lowest energy structures in the gas phase and that are commonly used in computational studies.[13,15,46-50] The closest analogue of the formaldehyde dimer to that employed herein is an anti-parallel dimer that has both an n→π* interaction and a C–H/O interaction.[13,15,46-

[49,51,52] The presence of both interactions renders this structure suboptimal to examine the effects of solvation on solely the n→π* interaction. An n→π* interaction is individually weak, but collectively provides a substantial amount of energy in protein structure due to its role in multiple conformations, with half of all carbonyls in high-resolution protein structures exhibiting an n→π* interaction.[4,5,53] For example, in the α-helix, both carbonyl oxygen lone pairs are involved in noncovalent interactions (Figure 1a): the p-like lone pair (O*p*) engages in a hydrogen bond with the *i*+4 amide hydrogen, while the s-like lone pair (O*s*) engages in an n→π* interaction with the *i*+1 carbonyl (Figure 1b). Hydrogen bonds are inherently stronger than n→π* interactions. Consequently, analysis in systems that have competing hydrogen bonds will result primarily in optimization of the geometry to maximize hydrogen bonding interactions.

In contrast, geometry optimization using the twisted-parallel-offset formaldehyde dimer avoids complications that are associated with more complex molecules. We also considered the twisted-parallel-offset acetone dimer, which is an energy minimum in gas phase calculations.[13] Acetone is electronically more similar to protein carbonyls than is formaldehyde, with the acetone carbonyl a better electron donor for n→π* interactions than formaldehyde (though less electron-donating than an amide carbonyl). Conversely, acetone is a weaker electron-acceptor for

n→π* interactions than is formaldehyde (though still a stronger acceptor than an amide), as electron-donor (Lewis basicity) and electron-acceptor (electrophilicity) capabilities of carbonyls are in general inversely related. However, even the twisted-parallel-offset acetone dimer[13] exhibited evidence of the influence of carbonyl interactions with the methyl C–H groups (C–H/O interactions) during geometry optimization (data not shown).[51,52,54] Due to the inherently weak nature of n→π* interactions, these stronger, competing interactions have the possibility to complicate the analysis of the n→π* interactions in the isolated molecules. The twisted-parallel-offset formaldehyde dimer represents an appropriate minimal model system for the isolation of the effects of the n→π* interaction from other potential competing interactions.

Geometry-optimized structures were obtained through an approach that involved iterative increases in the level of theory and the size of the relevant basis sets. The final geometry-optimized structures (Figure 2, Table 1) were generated using the MP2 level of theory and the aug-cc-pVTZ basis set. All geometry optimization calculations were conducted in implicit water, in order to mimic the appropriate electrostatic environment of proteins and to avoid electrostatic artefacts due to an artificial vacuum environment. Indeed, the twisted-parallel-offset

formaldehyde dimer was not observed as a stable structure in calculations in vacuum, presumably due to the destabilizing effect of its large overall dipole moment (7.3 D).

The structures were analyzed (Figure 1c) for the effect of solvation of the donor carbonyl and/or of the acceptor carbonyl on the intercarbonyl O•••C=O interaction distance (d), where a O•••C distance less than the sum of the van der Waals radii of O and C (d < 3.22 Å) is consistent with an n→π* interaction, with shorter distances indicating more favorable interactions. The structures were also analyzed for the O•••C=O interaction angle ($\angle_{O\cdots C=O}$), with an ideal n→π* interaction resulting in this interaction approximating the Bürgi-Dunitz trajectory (~107˚). In addition, structures were analyzed for pyramidalization (Δ) of either the donor carbonyl or the acceptor carbonyl. Pyramidalization here is defined by the HHOC torsion angle for formaldehyde, with a larger extent of pyramidalization (more non-planar carbonyl geometry) on the acceptor carbonyl indicating greater electron delocalization and a stronger n→π* interaction.

The formaldehyde dimer exhibited an intercarbonyl O•••C=O interaction distance of 2.835 Å and an internuclear O–C–O angle of 114˚, similar to those observed in proteins and consistent with a favorable n→π* interaction.[4,5] The formaldehyde dimer exhibited 0.6˚ pyramidalization on the acceptor carbonyl, but, as expected, exhibited no pyramidalization on the donor carbonyl.

The interaction energy of the formaldehyde dimer was –1.40 kcal/mol in implicit water, using the MP2 method and the large jul-cc-pV5Z basis set, in order to minimize the effects of basis-set superposition error (BSSE) in determining interaction energy. Using the more computationally intensive CCSD(T) method and the same basis set, an interaction of –1.38 kcal mol^{-1} was calculated, essentially identical to that determined by MP2. A BSSE-corrected interaction energy of –1.31 kcal mol^{-1} was determined by counterpoise calculations (MP2/jul-cc-pV5Z) in the gas phase, with only 0.06 kcal mol^{-1} BSSE observed using this combination of method and basis set. These results indicate that the n→π* interaction of the formaldehyde dimer is similar in strength to that observed within peptides and proteins in water. Moreover, the strength of the interaction was similar in the gas phase and in implicit water (Table 1), consistent with a primarily stereoelectronic (molecular orbital-based) nature to the interaction, with only a modest contribution from electrostatics ($C{=}O^{\delta-}\cdots^{\delta+}C{=}O$), despite the substantial partial charges on the C and O atoms of the carbonyls. This similar interaction strength in water and vacuum also suggests only a modest role for either unfavorable or favorable dipole-dipole interactions on complex stability.[14,55,56] Natural bond orbital (NBO) analysis[33,34] confirmed the predominantly

stereoelectronic component to the n→π* interaction, with a second-order perturbation energy of 2.7 kcal mol^{-1} due to electron delocalization via orbital overlap of the s-like and p-like oxygen lone pairs with the π* molecular orbital of the acceptor carbonyl (O_s/π* 0.6 kcal mol^{-1}, O_p/π* 2.1 kcal mol^{-1}) (Figure 2a).[35,38,39,57,58]

In order to investigate the role of solvation on n→π* interactions, models of the formaldehyde dimer were generated with hydrogen-bonding groups on the electron-donor carbonyl and/or the electron-acceptor carbonyl. Models were examined with zero or one hydrogen-bonding groups on the electron-donor carbonyl, and with zero, one, or two hydrogen-bonding groups on the electron-acceptor carbonyl (Figure 2b, Table 1). Both HF and H_2O were examined as hydrogen-bond donor groups in these calculations. HF is a stronger hydrogen-bond donor than H_2O, allowing the examination of the effect of hydrogen bond donor strength on the n→π* interaction. HF also has a technical advantage during geometry optimization calculations: it is not prone to adopt alternative structures that are driven by hydrogen bonding of the additional unsatisfied hydrogen-bond donor in water, which can result due to the substantially greater strength of hydrogen bonds compared to n→π* interactions.

Hydrogen bonding to the acceptor carbonyl led to closer n→π* interaction distances, with the closest interactions occurring with two hydrogen bonds to the acceptor (O···C=O for HCHO•••HCHO 2.835 Å; HCHO••••HCHO•HOH 2.759 Å; HCHO•••HCHO•$(HOH)_2$ 2.683 Å) and/or with the stronger hydrogen bond donor HF (HCHO•••HCHO•HF 2.677 Å; HCHO•••HCHO•$(HF)_2$ 2.551 Å). Increased hydrogen bonding to the acceptor carbonyl also increased its pyramidalization (Figure 2b, Table 1). In contrast, hydrogen bonds to the electron-donor carbonyl led to longer n→π* interaction distances. Closer n→π* interactions were associated with a greater extent of pyramidalization on the acceptor carbonyl. In contrast, no pyramidalization of the donor carbonyl was observed in any case.

Overall, the effects of hydrogen bonding were greater on the acceptor carbonyl than on the donor carbonyl. In addition, the fully explicitly solvated formaldehyde dimer (one hydrogen bond donor on the donor carbonyl, two hydrogen bond donors on the acceptor carbonyl; one oxygen lone pair on the donor carbonyl is used for the n→π* interaction) exhibited closer interaction distances than the unsolvated (not explicitly solvated) formaldehyde dimer, indicating that solvation via hydrogen bonding inherently promotes the n→π* interaction. These computational results are consistent with experimental data on the polyproline II helix (PPII),

which is stabilized by intercarbonyl n→π* interactions and which is promoted in water and disfavored in non-hydrogen-bonding solvents.[18,20,22-25,27,30,59-65] Notably, in proteins, n→π* interactions are competitive with, and can replace, hydrogen bonds as sites of interaction for carbonyl lone pairs.[53,66,67] The role of n→π* interactions explains the "partial occupancy" of water (significantly less than 2 water molecules per carbonyl) that is observed in proteins on solvent-exposed amide carbonyls, whereby n→π* interactions replace hydrogen bonds to water in solvating the carbonyls.[68]

In order to further explore the role of solvation in n→π* interactions, the formaldehyde dimer was analyzed with the acceptor carbonyl hydrogen-bonded to the protein denaturants urea, thiourea, or guanidinium, which can solvate protein carbonyls in the denatured state. These denaturants all can adopt conformations in which they have bidentate hydrogen bonds with the carbonyl, which might provide an entropic advantage compared to hydrogen bonding to water. The results indicated that all three denaturants lead to closer n→π* interactions, with guanidinum resulting in the closest interaction, and with the effect of thiourea greater than urea (Figure 3a, Table 1). Urea stabilizes polyproline II helix (PPII) in peptides and proteins.[23-26]

These results suggest that one mechanism by which urea stabilizes PPII is via promotion of n→π* interactions, which inherently stabilize PPII.

In addition, the intercarbonyl distance in the formaldehyde dimer was examined for the effects of monovalent cations on the acceptor carbonyl (Figure 3b, Table 1). Protonation of the acceptor carbonyl led to nucleophilic attack by the donor carbonyl and a covalent O•••C distance (1.567 Å). Li^+, Na^+, and K^+ also led to closer intercarbonyl distances, with the closest distances observed for the smallest cations, which interact most favorably with carbonyls. These data indicate that the interaction of the acceptor carbonyl with cations leads to closer, more favorable n→π* interactions, with alkali cations functioning as a Lewis acid to promote the n→π* interaction.

All structures were analyzed for the energies of the intercarbonyl interaction as a function of solvation of the donor and acceptor carbonyls, both with analysis in implicit water and via counterpoise calculations in the gas phase (Table 1). The results of these approaches indicated a small BSSE (< 0.1 kcal mol^{-1}) in the vast majority of complexes using the jul-cc-pV5Z basis set, suggesting that the energies calculated in implicit water accurately represented the relative interaction strengths. The use of smaller basis sets was associated with substantially more

significant BSSE that were larger than the differences in interaction energies (data not shown). In addition, larger overall interaction energies were observed using basis sets with diffuse functions (e.g. jul-cc-pV5Z versus cc-pV5Z), as expected for noncovalent interactions that involve electron delocalization.

Overall, these results indicated that solvation of the electron-donor carbonyl in n→π* interactions resulted in a weaker n→π* interaction, as indicated by a longer O•••C=O interaction distance, a less favorable interaction energy, and a smaller extent of pyramidalization of the electron-acceptor carbonyl. In contrast, the addition of one or two hydrogen-bonding groups to the electron-acceptor carbonyl resulted in a stronger n→π* interaction, with shorter O•••C=O interaction distances, more favorable interaction energies, and greater pyramidalization of the acceptor carbonyl. Importantly, the trends that were observed in interaction distances closely correlated with the interaction energies (Figure 4) and with the number and identity of hydrogen bonding groups: the presence of two hydrogen-bond donors on the electron-acceptor carbonyl resulted in a more stable n→π* interaction than when only one hydrogen bond was present, and stronger hydrogen-bond donor groups led to more favorable n→π* interactions. In addition, complexes with explicit hydrogen bonding to both the donor and acceptor carbonyl were more

stable than complexes without explicit hydrogen bonding, indicating that solvation of the carbonyls inherently stabilizes the n→π* interaction.

In order to understand the effects of solvation on n→π* interactions and structure within a more protein-like context, we next examined the effects of solvation as a function of protein conformation on Ac-Pro-NMe_2 molecules. Here, geometry optimization calculations were conducted on Pro with each combination of an *exo* and *endo* ring pucker, and each with Pro in either the PPII or α_R conformation. Models were developed with zero, one, or two molecules of HF on the electron-donor (acetyl) carbonyl, the electron-acceptor (Pro) carbonyl, or both. In addition, models were developed with urea, thiourea, or guanidinium on the electron-acceptor (Pro) carbonyl. The resultant structures were examined for the effects of solvation on conformation and n→π* interaction geometry (Figure 5, Table 2). In addition, all structures were evaluated for their final energies, with the relative energies of the PPII-*exo*, PPII-*endo*, α_R-*exo*, and α_R-*endo* conformations compared. In addition, the effects of solvation with a water cluster were examined. Importantly, solvation by HF (Figure 5b) or by a water cluster (Figure 5d) yielded similar trends in the effects of carbonyl solvation on n→π* interaction distances. This analysis yielded the following conclusions in a more protein-like context: (1) solvation of the

acceptor carbonyl in an n→π* interaction results in a closer interaction; (2) solvation of the acceptor carbonyl with two hydrogen-bond donors (whether two molecules of HF or H_2O, or the denaturants urea, thiourea, or guanidinium [Figure 5c]) yields closer n→π* interactions; (3) in the PPII conformation, solvation of the acceptor carbonyl results in smaller energy differences between proline *endo* and *exo* ring puckers; (4) solvation of the acceptor carbonyl results in a greater energetic preference for PPII over the α-helix (α_R) conformation.

Similar calculations were conducted on the minimal alanine molecule Ac-Ala-NMe_2 as a function of solvation of the electron-donor and/or electron-acceptor carbonyl. Here, three low-energy conformations were examined: PPII, α_R, and the β/extended conformation, which is stabilized by an intraresidue C5 hydrogen bond between the Ala amide hydrogen and the Ala carbonyl oxygen.[69] In implicit solvent, all three conformations represent energy minima, with the β conformation lowest in energy (Figure 6, Table 3).[50,70,71] In contrast, explicit solvation of the electron-acceptor carbonyl, or of both the electron-donor and electron-acceptor carbonyls, results in the lowest energy conformation being the PPII conformation. As expected, solvation effects on intercarbonyl distances were consistent with analyses in the formaldehyde dimer and Ac-Pro-NMe_2, with greater solvation of the acceptor carbonyl leading to closer n→π* interactions. A

substantial effect in stabilizing PPII over both β and α_R conformations was observed with urea, indicating that urea specifically stabilizes PPII relative to other conformations. Most dramatically, in structures solvated with 2 molecules of HF on the acceptor carbonyl and 0 or 1 molecules of HF on the donor carbonyl, the β conformation was not identified as an energy minimum: structures with initial geometries in a C5 β conformation were not stable, with the geometry optimization generating structures with the PPII minimum instead. These results are likely due to the solvation of the acceptor (Ala) carbonyl weakening the inherent strength of the C5 hydrogen bond by reducing the carbonyl electron density associated with the C5 hydrogen bond. Geometry optimization with ϕ fixed to –160°, identical to that observed with implicit solvation, indicated that indeed the C5 hydrogen bond was much longer (weaker) with explicit HFHF solvation on the C5-electron-donor (intraresidue) carbonyl, and that the β conformation was significantly destabilized relative to the PPII or α conformations.

In order to further explore the effect of solvation on PPII in model peptides, we examined solvation of Ac-Pro_2-NMe_2, as a function of number of HF solvent molecules, position of solvent molecules, and proline ring pucker. All peptides were examined in the PPII local energy minima only. In Ac-Pro_2-NMe_2 peptides using implicit solvation, the highest energy conformation had an

exo ring pucker at both prolines, while the lowest energy conformation had an *endo* pucker at both prolines, with these two conformations differing in energy by 1.4 kcal mol^{-1} (Figure 7, Tables 4 and 5). The addition of 2 HF molecules to Pro2 resulted both in closer n→π* interactions at the solvated carbonyl, and a substantial reduction in the energy differences between conformations, with a specific increase in preference for the *exo* ring pucker on Pro2 and an *exo-exo* versus *endo-endo* energy difference of only 0.3 kcal mol^{-1}, with the *exo-exo* combination lower in energy. Similar results were observed in models with 2 HF on Pro2 and 0 or 1 HF on the Ac or Pro1 carbonyls. Most dramatically, in the structure with 1 HF on Ac and 2 HF on each of the Pro1 and Pro2 carbonyls, the energy difference between the lowest and highest energy conformations was only 0.4 kcal mol^{-1}. In the fully solvated system (2 HF on each carbonyl). the *exo-exo* and *endo-endo* conformations were essentially identical in energy. Most broadly, the exact pattern of solvation (1 versus 2 solvent molecules on specific carbonyls) changed the relative conformational and energy preferences at each position, with 2 solvent molecules on the acceptor carbonyl promoting closer (stronger) n→π* interactions and 2 solvent molecules on the donor carbonyl leading to longer (weaker) n→π* interactions.

Discussion

n→π* interactions are important determinants of structure in proteins, as well as in the preferred conformations of small molecules.[1-5,72] Because these individual interactions are weaker than hydrogen bonds and many other noncovalent interactions, they are challenging to study due to the ability of other noncovalent interactions to outcompete the n→π* interaction energetically. Herein, we developed a new model system, the twisted-parallel-offset formaldehyde dimer, in order to investigate n→π* interactions. This structure is an energy minimum in implicit water, but is not an energy minimum observed in calculations in vacuum, presumably due to its large molecular dipole being more unfavorable in vacuum than the favorable nature of the n→π* interaction. To our knowledge, the twisted-parallel-offset formaldehyde dimer has not been described previously. This model system avoids complications due to hydrogen bonding, C–H bond dispersion, and C–H/O interactions that are present in other potential model systems.[13,15,46-49] The twisted-parallel-offset formaldehyde dimer is stabilized by an intermolecular n→π* interaction, with the stabilization resulting due to through-space

electron delocalization, via orbital overlap between the oxygen lone pairs on the electron-donor carbonyl and the π* molecular orbital of the electron-acceptor carbonyl.

The effects of solvation on the n→π* interaction were examined using explicit solvation with a series of hydrogen-bonding groups. Interaction strength was determined as a function of the number and identity of the solvent molecules, via the geometry of the interaction and via interaction energies. The presence of hydrogen-bonding groups on the electron-acceptor carbonyl, or of a hydrogen-bonding group on both the electron-donor and electron-acceptor carbonyls, resulted in a closer and more favorable n→π* interaction. The magnitude of the effect correlated with the identity of the hydrogen-bonding group (stronger interactions with stronger hydrogen-bond donors) and the number of hydrogen bonds to the acceptor carbonyl, with stronger n→π* interactions when two hydrogen bonds were present. Similar effects were observed with alkali-metal Lewis acids on the electron-acceptor carbonyl. In contrast, solvation on only the electron-donor carbonyl resulted in a weaker n→π* interaction. Overall, the solvation effects on the electron-acceptor carbonyl were greater than those on the donor carbonyl (Table 1).

NBO second-order perturbation energies[33,34] (Figure 2) confirmed that the interaction strength was driven by a stereoelectronic (molecular orbital) effect, involving through-space electron delocalization between oxygen lone pairs on the donor carbonyl and the π* molecular orbital on the electron-acceptor carbonyl, as has been previously seen in the analysis of n→π* interactions in peptides and proteins.[1,2,4,5] Consistent with this interaction being driven predominantly by electron delocalization, rather than by simple electrostatics, the interaction energies for the formaldehyde dimer are nearly identical in implicit water and in vacuum (–1.34 versus –1.31 kcal mol^{-1}), despite the favorable electrostatic interactions that are possible between the donor carbonyl O and the acceptor carbonyl C that would be expected to be far stronger in vacuum than in water.

The n→π* interaction was subsequently examined in the model peptide structures Ac-Pro-NMe_2, Ac-Ala-NMe_2, and Ac-Pro_2-NMe_2. As was the case in the formaldehyde dimer, explicit solvation of the acceptor carbonyl yielded closer n→π* interactions. Notably, solvation that promoted n→π* interactions, including solvation by the protein denaturants urea and guanidinium, also resulted in a preference for the PPII conformation over the α-helix and β/extended conformations.

n→π* interactions are observed with diverse relative geometries of the two interacting carbonyls (relative vectors of the carbonyls), which reflect the possible contributions of both lone pairs of the electron-donor oxygen, via the extended loci of electron occupancy around oxygen (Figure 1b). Despite these observed differences in the disposition of the electron-donor carbonyl relative to the electron-acceptor carbonyl, the n→π* interaction inherently exhibits tight geometric preferences in the $\angle_{OCO}$ O•••C=O internuclear angle (~109˚) and in the O•••C intercarbonyl distance (less than the 3.22 Å sum of the van der Waals radii of O and C).[5] The similarities of results in the formaldehyde dimer and in peptides suggest that solvent effects will be general in their impact on the strengths of n→π* interactions.

Collectively, these data provide a model by which solvation by protic solvents and/or by chemical protein denaturants promotes the PPII conformation, as is observed experimentally. Carbonyl solvation inherently promotes the PPII conformation, via making the carbonyl a better electron acceptor of an n→π* interaction. Moreover, carbonyl solvation favors PPII over the competing α-helix conformation, potentially due to stronger n→π* interactions in PPII than in α-helix, and/or alternatively due to stronger hydrogen bonding of the carbonyl to solvent in a PPII conformation than that of the carbonyl in an α-helix conformation. Solvation of the

acceptor carbonyl also significantly destabilizes the β/extended conformation by weakening the intraresidue C5 hydrogen bond that stabilizes the β conformation[69] in the absence of secondary structure (β-sheets). Finally, carbonyl solvation appears to reduce the energy difference between the proline *endo* and *exo* ring puckers. Proline inherently favors an *endo* ring pucker (approximately 2:1 *endo:exo* ratio observed experimentally in water)[73], which was also observed computationally. However, solvation of the proline carbonyl leads to a reduction in the energy difference between the *endo* and *exo* ring puckers. While data in peptides with 4-substituted prolines indicate that *exo*-favoring residues are better for promoting PPII,[1,74-79] a requirement for an *exo* ring pucker for PPII would impose both an enthalpic (higher energy structure) and entropic (selection of one of two ring puckers) penalty to adopt PPII. These results suggest that solvation overcomes both of these energetic costs, both by specifically promoting PPII in an *endo* ring pucker, and by making the *exo* and *endo* puckers more similar in energy for adopting PPII, reducing the entropic cost of adopting PPII. We therefore predict than an analysis of proline ring pucker in proline oligomers under fully denatured conditions will indicate little or no preference for *endo* versus *exo* ring pucker.

The effects of solvation, as observed in structures of Ac-Pro_2-NMe_2, were inherently local: any specific carbonyl solvation affected that residue significantly, and the prior residue only very modestly. These results are consistent with a lack of cooperativity in PPII and the inherently local nature of interactions on PPII stability.[21,63,64,80,81]

The denatured state of proteins exhibits evidence of substantial PPII conformation.[18-20,82] PPII is also promoted by protic solvents over aprotic solvents, and by more polar protic solvents over less polar protic solvents.[22] D_2O also promotes PPII over H_2O.[27] Moreover, the chemical protein denaturant urea promotes PPII directly, through mechanisms that are not well understood.[19,24-26] Urea denatures protein secondary structures through direct hydrogen bonding of urea to protein amide groups, competing against the hydrogen bonding patterns necessary for α-helix, β-sheet, and β-turn protein secondary structures. However, this disruption of hydrogen bonding to destabilize secondary structures does not explain why the PPII conformation is *preferred* in the denatured state, as both the α and β conformations are stabilized by noncovalent interactions in the absence of interresidue hydrogen bonds. Herein, we show that urea *directly* promotes the PPII conformation over both the α and β conformations, via selective stabilization

of the n→π* interaction in PPII *and* via destabilization of the C5-hydrogen bonded conformation of individual residues in the β conformation.

Rose and coworkers have suggested, based on molecular mechanics-based calculations, that solvation inherently promotes PPII over the α and β conformations, via greater solvent accessibility and solvent-backbone hydrogen bonding that is possible in PPII.[83,84] The results herein confirm those predictions using more rigorous quantum-mechanical calculations, and expand on the implications based on the analysis of the inherent noncovalent interactions that stabilize secondary structures in individual amino acids. Solvation specifically increases the stability of PPII, via stronger/more favorable n→π* interactions, and decreases the stability of the β/extended conformation, via weakening the C5 hydrogen bond. The effects of urea in stabilizing PPII thus appear to be significantly enthalpic: dual hydrogen bonding to carbonyls directly promotes PPII via closer n→π* interactions. The dual hydrogen bonding of ureas is also central to the function of small-molecule urea, thiourea, and guanidinium asymmetric catalysts.[85,86] In addition, urea exhibits an entropic advantage over water in promoting n→π* interactions, via bivalent hydrogen bonding, whereby the establishment of a second hydrogen bond does not have a translational entropy cost.[87] Notably, protein carbonyls in the urea-

denatured states exhibit approximately 50% urea occupancy (~ 1 urea per two amino acids), which includes urea interactions with both the backbone and with sidechains.[88] The urea-bound carbonyls thus are expected to be particularly strong loci as electron acceptors to stabilize n→π^* interactions in the denatured state of proteins. Thus, the results herein suggest that urea can promote n→π* interactions via both enthalpic and entropic effects that stabilize PPII.

These results confirm a critical role for solvation in the stability of PPII. PPII is a sterically favorable conformation.[83,89] However, PPII is directly stabilized both by n→π* interactions between consecutive carbonyls *and* by carbonyl solvation, which further promotes PPII. The importance of solvation in PPII can also be observed in a recent small-molecule crystal structure of a proline oligomer in the PPII conformation.[8] While no water molecules are bound to the proline carbonyls, each carbonyl exhibits an intermolecular solvation interaction, including a hydrogen bond to the C-terminal carboxylic acid and C–H/O interactions of carbonyls with the polarized C–H bonds of solvent (acetonitrile), an aromatic ring, or a proline ring. We also propose that the observed[27] stabilization of PPII in D_2O versus H_2O is due to the stronger hydrogen bonds of the main chain carbonyls to D_2O than to H_2O,[90,91] which yields a more stable polyproline helix via stronger carbonyl solvation. Collectively, the sum of all experimental and

computational data indicates that PPII is stabilized both by the n→π* interaction and by carbonyl solvation, the latter of which further promotes the n→π* interaction and PPII.

Acknowledgements

We thank NSF (CHE-2004110, CHE-1412978, and BIO-1616490) for funding. We thank Joaquin Barroso for helpful discussions.

Supporting Information Available

Comparison of energies of formadehyde dimer complexes determined by the MP2 and CCSD(T) methods, energies of all structures in hartrees, and coordinates of all geometry-optimized structures. This material is available free of charge via the Internet at the journal web site.

Figures

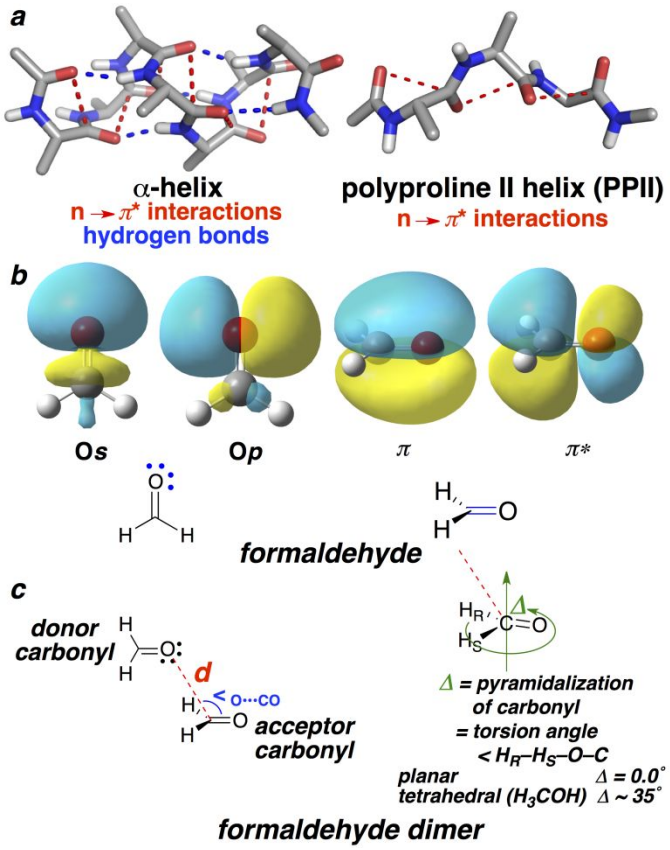

Figure 1. (a) n→π* interactions (red) between consecutive carbonyls (*i*/*i*+1 interactions) stabilize α-helices and polyproline II helices (PPII), shown using polyalanine (Ac-Ala_7-NHMe and Ac-Ala_3-NHMe) models. (b) Localized representations (NBO) of key molecular orbitals in formaldehyde. Left: the *s*-like (O*s*) and *p*-like (O*p*) oxygen lone pairs. Right: the π and π^* orbitals of the carbonyl. Blue and yellow colors indicate opposite signs of the wave function. (c) The twisted-parallel-offset formaldehyde dimer and definitions of geometric variables used in the analysis herein.

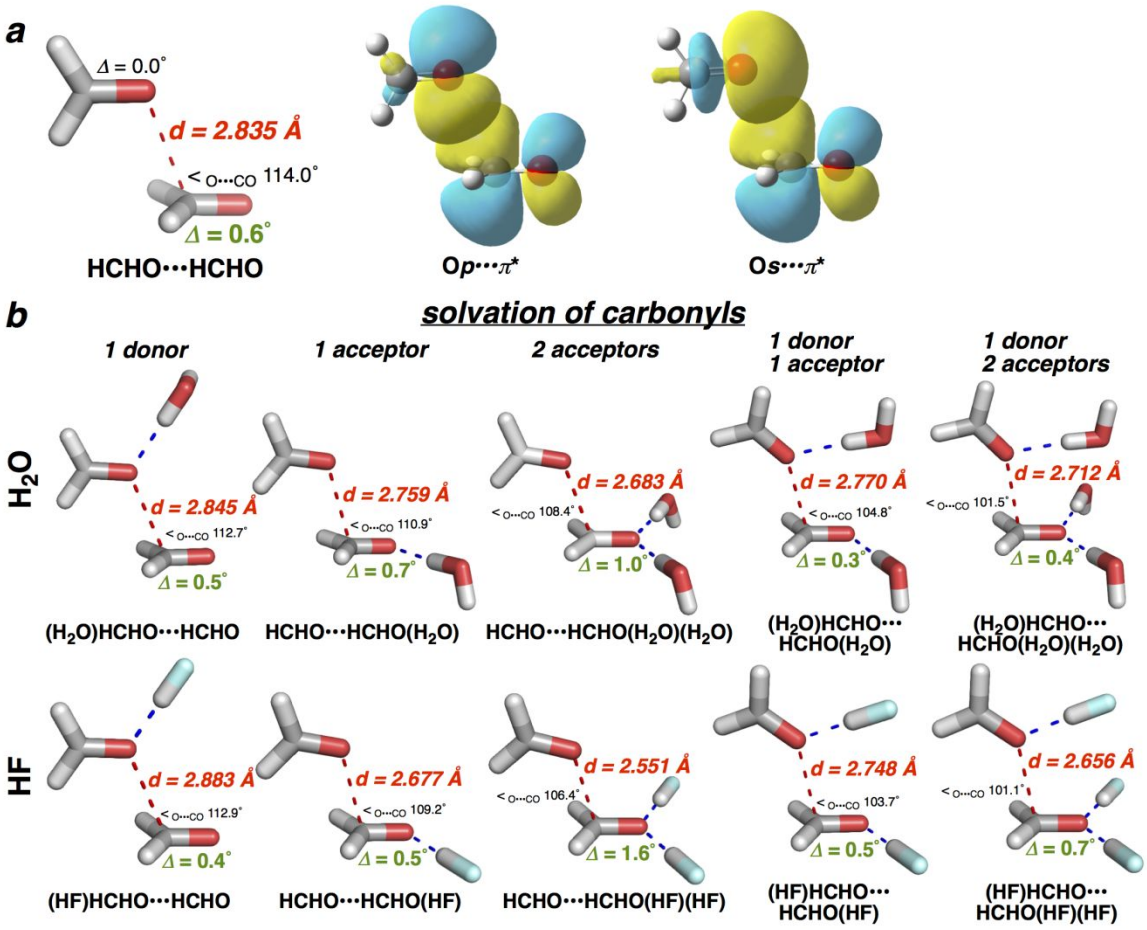

Figure 2. Geometry-optimized structures of formaldehyde dimer complexes in implicit water as a function of explicit carbonyl solvation. All structures were generated using the MP2 method and the aug-cc-pVTZ basis set in implicit water. Red dashed lines indicate n→π* interactions, while blue dashed lines indicate hydrogen bonds. (a) Formaldehyde dimer structure with no explicit solvation (implicit H_2O solvation only). Middle and right: NBO representation of orbital overlap between the donor *p*-like (O*p*, middle) and *s*-like (O*s*, right) oxygen lone pairs and the acceptor π* molecular orbital that contribute to electron delocalization and stabilization of the complex. The extent of orbital overlap (overlapping yellow lobes of lone pair donor and π* acceptor orbitals) correlates with the extent of stabilization, with greater orbital overlap resulting in greater stabilization. The second-order perturbation energies (determined using the M06-2X method and the aug-cc-pV6Z basis set) were 2.08 (O*p*/π*) and 0.63 (O*s*/π*) kcal mol^{-1}, respectively. Similar second-order perturbation energies (±0.15 kcal mol^{-1} for most) were calculated with a range of DFT functionals (20 methods examined) and large quadruple-zeta basis sets with diffuse functions (jun-cc-pVQZ, aug-cc-pVQZ, Def2QZVPPD). While NBO represents the role of key frontier molecular orbitals in stabilization (here, with the largest effect

due to the O*p*/π* interaction), which can be used to describe and explain the observed geometry of the interaction, the overall stabilization due to delocalization involves all of the molecular orbitals of both molecules. (b) Effects of solvation by H_2O (top) and HF (bottom) on the structure of the formaldehyde dimer, as a function of site of solvation (donor and/or acceptor carbonyl), the number of hydrogen-bonded solvent molecules, and the strength of hydrogen bond donor. Geometric parameters are defined in Figure 1c.

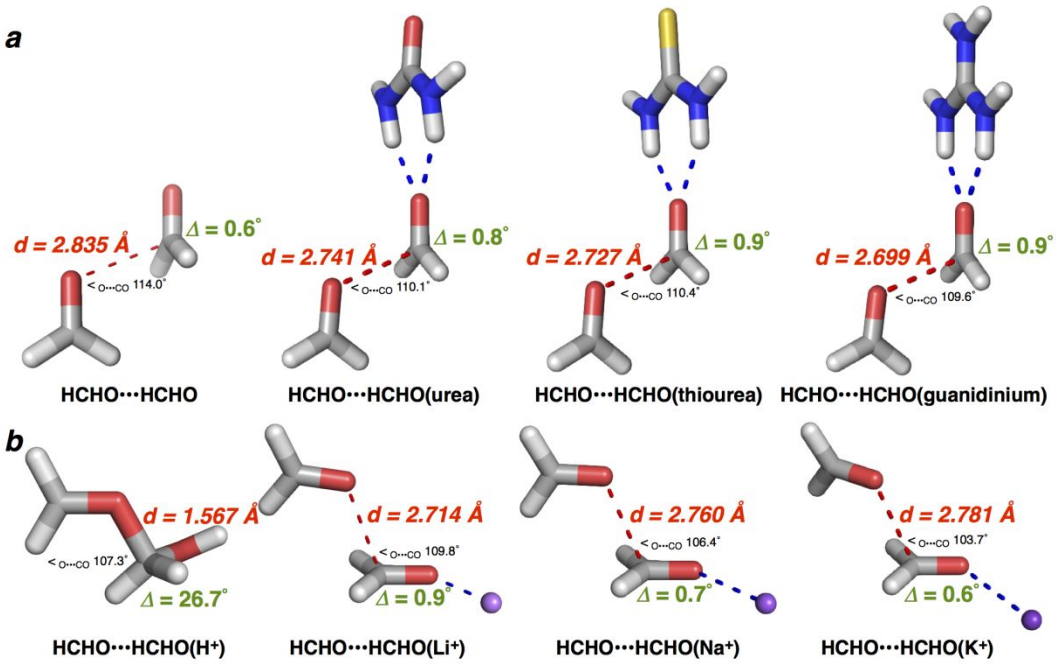

Figure 3. Geometry-optimized structures of the formaldehyde dimer with (a) hydrogen-bonded protein denaturants or (b) Bronsted acids or alkali-metal Lewis acids. The complex with H^+ was initially generated as a Bronsted acid complex of the formaldehyde dimer, which during geometry optimization generated the covalent bond. A similar structure resulted from the Bronsted acid hydrogen-bonded complex with H_3O^+. Red dashed lines indicate $n \rightarrow \pi^*$ interactions, while blue dashed lines indicate hydrogen bonds or Lewis acid complexes. Geometric parameters are defined in Figure 1.

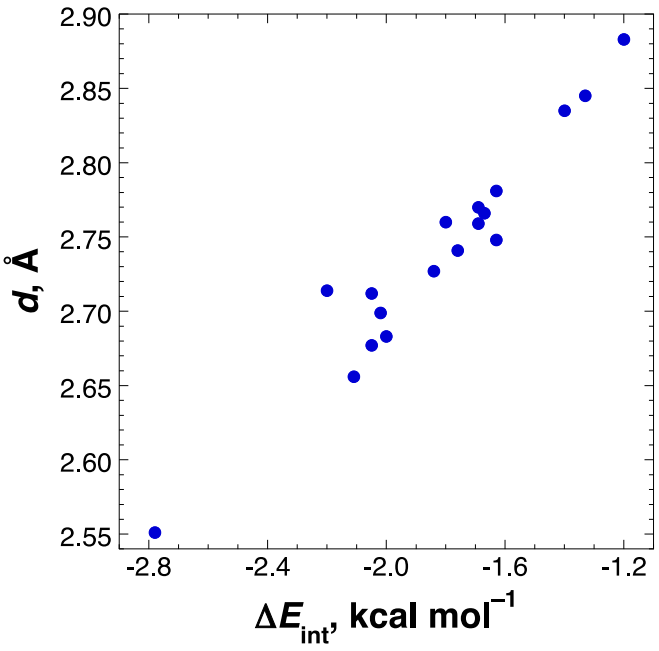

Figure 4. Correlation of formaldehyde dimer interaction energy (ΔE_{int}, kcal mol^{-1}) in implicit water with the O•••C n→π* interaction intercarbonyl distance (O•••C=O, Å) across all formaldehyde dimer complexes in Table 1 (excluding the fully covalent complex resulting from interaction of the acceptor carbonyl with H$^+$). These distances are all well below the 3.22 Å sum of the van der Waals radii of O and C.

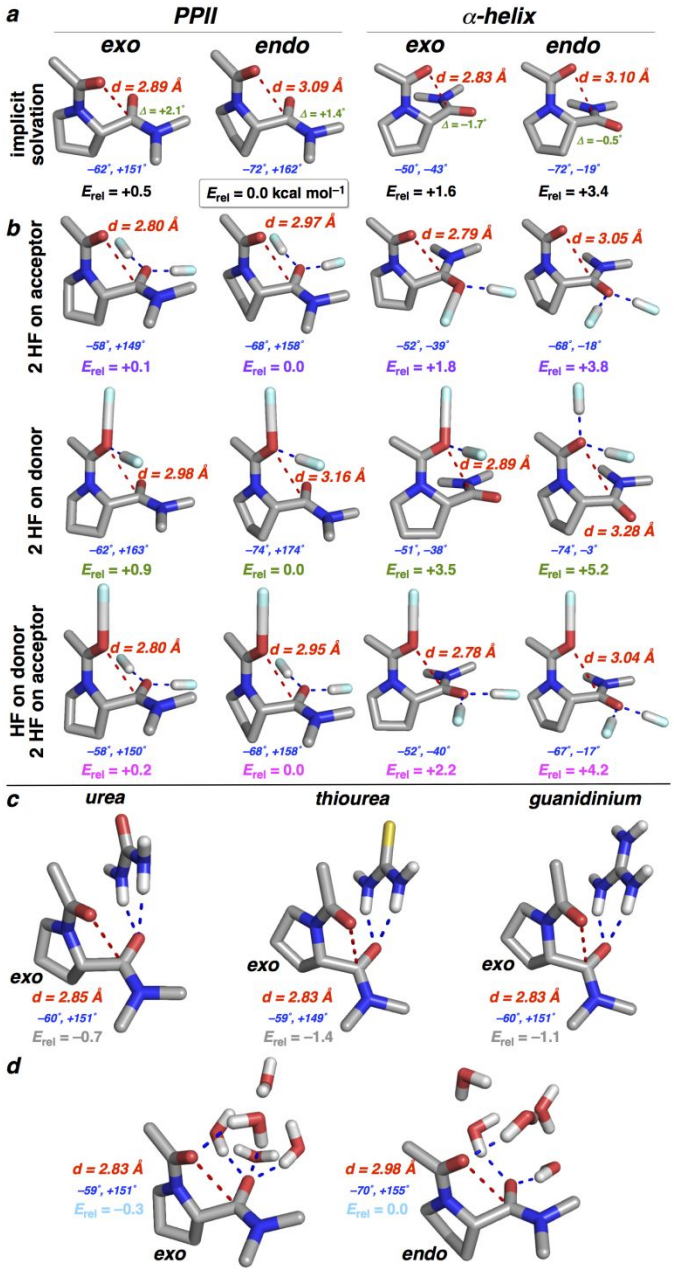

Figure 5. Structures of Ac-Pro-NMe_2 with *exo* and *endo* ring puckers in PPII and α-helix conformations and the indicated representative solvation. n→π* interaction distances (*d*) and (ϕ,ψ) main chain torsion angles are indicated. Energies are relative to the PPII conformation with an *endo* ring pucker for a given pattern of solvation (E_{rel} defined as 0.0 kcal mol^{-1} for each solvation pattern). Nonpolar hydrogens are not shown for clarity. Full sets of values and structures are in Table 2 and the Supporting Information. Pyramidalization (Δ) of the acceptor carbonyl is defined by the torsion angle (Cα–N–O–C). A positive value of Δ is consistent with the expected puckering for an n→π* interaction in a PPII conformation, while a negative value

of Δ is consistent with the expected puckering for an n→π* interaction in an α-helix conformation. All molecules exhibited the expected sign of Δ for their secondary structure; however, in molecules with explicit solvation, the magnitude was variable, due to the value of Δ being dependent on the positioning of all atoms. Any displacement of any atom (for example, a steric clash of a solvent molecule with the dimethyl amide, as was observed) will result in a change in Δ that does not directly relate to puckering of the carbonyl due to the n→π* interaction. (a) Structures obtained and energies determined using implicit water solvation. (b) Structures and energies using the indicated explicit solvation and using implicit water solvation. n→π* interactions are indicated by red dashed lines, while hydrogen bonds to carbonyls are indicated by blue dashed lines. In some structures, the hydrogen bond of a carbonyl with HF is close enough that the hydrogen bond was represented as a covalent bond by Pymol. (c) Structures in the PPII conformation with an *exo* ring pucker in the presence of the protein denaturants urea, thiourea, and guanidinium, with energies relative to the structures in the PPII conformation and the *endo* ring pucker. (d) Structures in the PPII conformation with the *exo* and *endo* ring puckers with a water cluster on the acceptor carbonyl; these initial water clusters optimized to include a hydrogen bond to the donor carbonyl. For (c) and (d), these relative energies are given with significantly reduced confidence, due to the presence of new noncovalent interactions with the denaturants, or changes in energy due to differences in the structure of H_2O-H_2O hydrogen bonds, in addition to carbonyl hydrogen bonding, that were obtained during geometry optimization. In addition, in general, the energies of complex molecules that interact with solvents (such as those herein) are subject due to greater variability due to solvation energy that is not fully addressed in these calculations. Thus, the energies of peptides in this study should be considered as relative conformational energies, rather than absolute conformational free energies, with the changes in patterns in energies as a result of changes in solvation more important than the absolute numbers. However, in all cases, the observed effect of solvation by denaturants or by a water cluster was a closer n→π* interaction relative to implicit solvation.

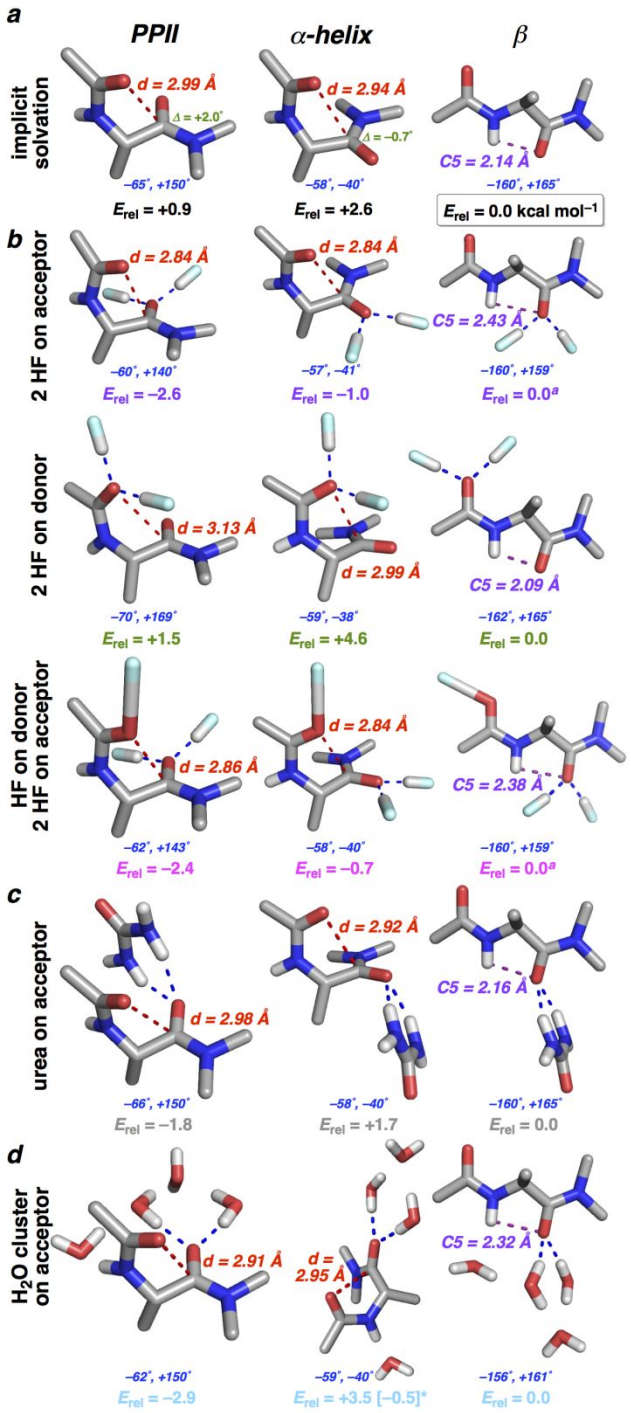

Figure 6. Structures of Ac-Ala-NMe_2 in the PPII, α-helix, and β conformations and the indicated representative solvation. n→π* interaction distances (d, red), C5 hydrogen bond distances (purple), and (ϕ,ψ) main chain torsion angles (blue) are indicated. Energies are relative to the β conformation for a given pattern of solvation (E_{rel} defined as 0.0 kcal mol^{-1} for each pattern of solvation). Nonpolar hydrogens are not shown for clarity. Full sets of values and structures are in

Table 3 and the Supporting Information. n→π* interactions are indicated by red dashed lines. (a) Structures obtained and energies determined using implicit water solvation. (b) Structures and energies using the indicated explicit solvation and using implicit water solvation. Hydrogen bonds to carbonyls are indicated by blue dashed lines. In some structures, the hydrogen bond of a carbonyl with HF is close enough that the hydrogen bond was represented as a covalent bond by Pymol. [a] Geometry optimization of these structures in the β conformation generated the equivalent structures with a PPII conformation; the structures such indicated resulted from geometry optimization with the ϕ torsion angle fixed at –160°, which was the observed value at the local minimum of the structure with implicit solvation. (c) Structures with urea on the acceptor carbonyl. (d) Structures with a water cluster on the acceptor carbonyl and with the amide hydrogen solvated by water. For clarity, only carbonyl-water hydrogen bonds are indicated by blue dashed lines. The structure in the α-helix conformation fundamentally differs from those in the PPII and β conformations due to a loss of one water-water hydrogen bond, and thus appears to be significantly higher in energy. At this level of theory in implicit water, a water-water hydrogen bond is stabilizing by 4.0 kcal mol^{-1} (Table S2); the approximate corrected E_{rel} of –0.5 kcal mol^{-1} (indicated by *) accounts for this lost hydrogen bond interaction energy.[50,70,71]

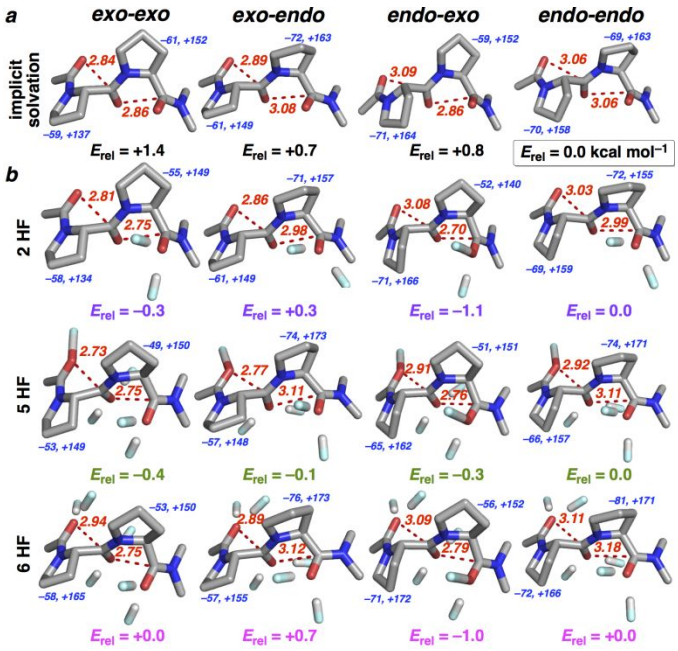

Figure 7. Structures of Ac-Pro$_2$-NMe$_2$ as a function of ring pucker and explicit solvation. Intercarbonyl n→π* interaction distances (red) and main chain (ϕ,ψ) torsion angles (blue) are indicated. Energies are relative to the conformation with an *endo* ring pucker on both prolines (E_{rel} defined as 0.0 kcal mol^{-1}) for a given pattern of solvation. Nonpolar hydrogens are not shown for clarity. Full sets of values and structures are in Table 4 and the Supporting Information. (a) Structures obtained and energies determined using implicit water solvation. (b) Structures and energies using the indicated explicit solvation and using implicit water solvation. n→π* interactions are indicated by red dashed lines. In some structures, the hydrogen bond of a carbonyl with HF is close enough that the hydrogen bond was represented as a covalent bond by Pymol.

Tables

Table 1. Summary of computational data (optimization MP2/aug-cc-pVTZ/H_2O, energies MP2/jul-cc-pV5Z/H_2O) on n→π* interactions in formaldehyde dimer complexes.[a]

	groups on C=O				donor	acceptor	ΔE_{int}, kcal mol^{-1}				
							H_2O			vacuum	
interaction	donor	acceptor	d	$\angle_{O\cdots C=O}$	Δ, °	Δ, °	raw	corr.[b]	E_{rel}[c]	corr.[d]	BSSE[e]
HCHO•HCHO	–	–	2.835	114.0	0.0	0.6	-1.40	*-1.34*	0.00	-1.31	-0.06
HCHO•HCHO	H_2O	–	2.845	112.7	0.0	0.5	-1.33	*-1.27*	0.07	-0.71	-0.06
HCHO•HCHO	–	H_2O	2.759	110.9	0.0	0.7	-1.69	*-1.62*	-0.28	-1.87	-0.07
HCHO•HCHO	–	H_2O H_2O	2.683	108.4	0.0	1.0	-2.00	*-1.91*	-0.57	-2.54	-0.09
HCHO•HCHO	H_2O	H_2O	2.770	104.8	0.0	0.3	-1.69	*-1.60*	-0.26	-0.28	-0.09
HCHO•HCHO	H_2O	H_2O H_2O	2.712	101.5	0.0	0.4	-2.05	*-1.92*	-0.58	-0.55	-0.13
HCHO•HCHO	HF	–	2.883	112.9	0.0	0.4	-1.20	*-1.14*	0.20	-0.15	-0.06
HCHO•HCHO	–	HF	2.677	109.2	0.0	1.1	-2.05	*-1.97*	-0.63	-2.60	-0.08
HCHO•HCHO	–	HF HF	2.551	106.4	0.0	1.6	-2.78	*-2.68*	-1.34	-4.10	-0.10
HCHO•HCHO	HF	HF	2.748	103.7	0.0	0.5	-1.63	*-1.55*	-0.21	-0.16	-0.08
HCHO•HCHO	HF	HF HF	2.656	101.1	0.0	0.7	-2.11	*-2.01*	-0.67	-0.88	-0.10
HCHO•HCHO	HF	H_2O H_2O	2.766	98.0	0.0	0.3	-1.67	*-1.57*	-0.23	0.21	-0.10
HCHO•HCHO	–	urea	2.741	110.1	0.0	0.8	-1.76	*-1.69*	-0.35	-2.46	-0.07
HCHO•HCHO	–	thiourea	2.727	110.4	0.0	0.9	-1.84	*-1.77*	-0.43	-2.87	-0.07
HCHO•HCHO	–	guanidinium	2.699	109.6	0.0	0.9	-2.02	*-1.95*	-0.61	-6.71	-0.07

HCHO•HCH			1.56				-18.0	-	-16.2		
O	–	H^+	7	107.3	-0.1	26.7	1	*17.61*	7	-51.24	-0.40
HCHO•HCH			2.71								
O	–	Li^+	4	109.8	0.0	0.9	-2.20	*-2.13*	-0.79	-9.03	-0.07
HCHO•HCH			2.76								
O	–	Na^+	0	111.3	0.0	0.7	-1.80	*-1.73*	-0.39	-7.38	-0.07
HCHO•HCH			2.78								
O	–	K^+	1	111.3	0.0	0.6	-1.63	*-1.57*	-0.23	-6.83	-0.06

[a] *d* = distance (Å) between the oxygen of the electron-donor carbonyl and the carbon of the electron-acceptor carbonyl. $\angle_{O\bullet\bullet\bullet C=O}$ = angle (˚) between oxygen of the electron-donor carbonyl, the carbon of the electron-acceptor carbonyl, and the oxygen of the electron-acceptor carbonyl. Δ = pyramidalization of the carbonyl, as defined by the torsion angle (˚) between the *pro-R* H, the *pro-S* H, the carbonyl O, and the carbonyl C. These geometry measurements are defined schematically in Figure 1c. The torsion angle-based method for quantification of carbonyl pyramidalization is compared with alternative methods to quantify carbonyl pyramidalization in Table S3.

[b] Correction for potential errors due to basis set superposition error (BSSE). Corrected H_2O interaction energy = raw interaction energy (H_2O) – BSSE (vacuum). While BSSE determination is inherently a vacuum-based calculation, this error in water was approximated using the magnitude of BSSE in gas phase.

[c] $E_{rel} = E_{\text{solvated HCHO•HCHO, corrected}} - E_{\text{HCHO•HCHO, corrected}}$. This number indicates the relative formaldehyde complex interaction energy in the presence of the indicated hydrogen bonding groups compared to the interaction energy of the formaldehyde dimer in the absence of explicit hydrogen bonding groups.

[d] BSSE-corrected complex interaction energy in the gas phase, as determined via counterpoise calculations.

[e] Magnitude of the gas-phase BSSE, which was used for energy corrections to complexes in water.

Table 2. Summary of geometry data and energies of Ac-Pro-NMe_2 as a function of ring pucker, region of the Ramachandran plot, and explicit solvation. All geometry optimization calculations were conducted with the M06-2X method and jun-cc-pVTZ basis set in implicit water. Energies were determined by the MP2 method with the aug-cc-pVTZ basis set in implicit water.

		solvation					E_{rel}[a]
							kcal mol^{-1}
pucker	*2°*	donor	acceptor	*d*, Å	ϕ	ψ	
exo	*PPII*	–	–	2.890	-61.6	150.7	0.48
		–	HF *cis*	2.832	-58.2	150.7	0.34
		–	HF *trans*	2.848	-60.1	149.2	0.37
		–	HF HF	2.803	-58.2	148.7	0.15
		HF	–	2.867	-61.0	152.6	0.55
		HF	HF HF	2.799	-58.4	149.7	0.24
		–	urea	2.846	-60.1	151.2	-0.66
		–	thiourea	2.827	-59.3	149.4	-1.35
		–	guanidinium	2.825	-59.5	150.7	-1.13
		–	H_2O *cis*	2.875	-61.7	150.8	0.57
		–	H_2O HF	2.830	-60.5	149.2	0.47
		HFHF	HF HF	2.930	-60.8	158.1	0.68
		HFHF	–	2.979	-62.1	163.2	0.92
endo	*PPII*	–	–	3.094	-72.2	162.4	*0.00*
		–	HF *cis*	2.992	-67.8	159.1	*0.00*
		–	HF *trans*	3.046	-70.5	162.3	*0.00*
		–	HF HF	2.973	-68.3	157.6	*0.00*
		HF	–	3.062	-72.0	164.6	*0.00*
		HF	HF HF	2.949	-67.9	157.6	*0.00*
		–	urea	3.022	-70.3	159.4	*0.00*
		–	thiourea	3.023	-71.3	156.5	*0.00*
		–	guanidinium	3.040	-73.3	158.2	*0.00*
		–	H_2O *cis*	3.031	-70.2	159.9	*0.00*
		–	H_2O HF	3.025	-71.1	160.4	*0.00*
		HFHF	HF HF	3.132	-73.4	172.2	*0.00*
		HFHF	–	3.159	-73.8	174.4	*0.00*
exo	α	–	–	2.828	-49.7	-42.9	1.60
		–	HF *cis*	2.820	-52.0	-39.9	1.90
		–	HF *trans*	2.809	-50.5	-39.8	1.23
		–	HF HF	2.793	-52.2	-39.3	1.78
		HF	–	2.834	-51.4	-40.8	1.88
		HF	HF HF	2.777	-51.7	-40.1	2.20

		–	urea	2.844	-52.4	-39.7	3.43
		–	thiourea	2.841	-52.6	-39.4	3.38
		–	guanidinium	2.800	-50.2	-43.0	2.58
		–	H_2O *cis*	2.849	-52.5	-39.7	2.69
		–	H_2O HF	2.819	-52.5	-38.4	2.58
		HFHF	HF HF	2.860	-51.4	-36.6	3.77
		HFHF	–	2.888	-51.1	-38.2	3.54
endo	α	–	–	3.095	-66.7	-19.5	3.39
		–	HF *cis*	3.070	-67.4	-19.2	3.78
		–	HF *trans*	3.052	-66.1	-19.8	3.21
		–	HF HF	3.054	-67.8	-17.7	3.77
		HF	–	3.079	-67.6	-19.5	3.90
		HF	HF HF	3.038	-67.5	-17.1	4.19
		–	urea	3.065	-66.3	-21.2	5.26
		–	thiourea	3.062	-66.2	-20.0	5.24
		–	guanidinium	3.056	-66.2	-20.2	4.49
		–	H_2O *cis*	3.118	-68.9	-17.1	4.55
		–	H_2O HF	3.066	-67.6	-18.2	4.44
		HFHF	HF HF	3.254	-75.1	-6.2	5.02
		HFHF	–	3.283	-74.3	-3.3	5.24

[a] Energies are relative to the energy of the molecule with the same explicit solvation with an *endo* ring pucker and a PPII conformation.

Table 3. Summary of geometry data and energies of Ac-Ala-NMe_2 as a function of region of the Ramachandran plot and explicit solvation. All geometry optimization calculations were conducted with the M06-2X method and jun-cc-pVTZ basis set in implicit water. Energies were determined by the MP2 method with the aug-cc-pVTZ basis set in implicit water.

	C=O solvation					E_{rel}[a], kcal mol^{-1}		
2°	donor	acceptor	d, Å	ϕ	ψ	to α	to PPII	to β
PPII	–	–	2.985	-64.8	150.2	-1.69	*0.00*	0.91
	–	HF HF	2.841	-59.6	139.9	-1.69	*0.00*	-2.64
	HF	HF HF	2.859	-62.0	142.9	-1.76	*0.00*	-2.44
	HFHF	HF HF	3.083	-67.0	163.4	-3.25	*0.00*	-1.71
	HFHF	–	3.134	-70.0	168.5	-3.05	*0.00*	1.53
	HFHF	HF	3.118	-69.4	167.2	-2.92	*0.00*	0.57
	–	urea	2.977	-66.1	150.0	-3.52	*0.00*	-1.85
	–	4 H_2O	2.910	-62.1	149.2	-2.4[b]	*0.00*	-2.89
α	–	–	2.940	-58.0	-39.7	*0.00*	1.69	2.60
	–	HF HF	2.840	-57.0	-40.5	*0.00*	1.69	-0.95
	HF	HF HF	2.838	-57.8	-40.3	*0.00*	1.76	-0.67
	HFHF	HF HF	2.936	-58.2	-36.2	*0.00*	3.25	1.53
	HFHF	–	2.990	-59.3	-38.1	*0.00*	3.05	4.58
	HFHF	HF	2.942	-58.1	-38.4	*0.00*	2.92	3.49
	–	urea	2.918	-58.1	-39.7	*0.00*	3.52	1.68
	–	4 H_2O	2.949	-59.4	-39.9	*0.00*	2.4[b]	-0.5[b]
β	–	–	–	-160.3	164.6	-2.60	-0.91	*0.00*
	–	HF HF[c]	–	-160.0	159.0	0.95	2.64	*0.00*
	HF	HF HF[c]	–	-160.0	159.1	0.67	2.44	*0.00*
	HFHF	HF HF	–	-151.0	160.6	-1.53	1.71	*0.00*
	HFHF	–	–	-162.4	164.9	-4.58	-1.53	*0.00*
	HFHF	HF	–	-160.4	162.9	-3.49	-0.57	*0.00*
	–	urea	–	-159.9	165.0	-1.68	1.85	*0.00*
	–	4 H_2O	–	-155.5	161.0	0.5[b]	2.89	*0.00*

[a] Energies are relative to the energy of the molecule with the same explicit solvation, with an α, PPII, or β conformation, respectively, as indicated in the column header.

[b] These numbers have been corrected from the raw numbers in order to control for the energetic cost of one less H_2O-H_2O hydrogen bond in the 4-H_2O cluster of the complex with the α-helix

conformation compared to the complexes with the PPII and β conformations, since an equivalent hydrogen bond would be present in a full explicit H_2O solvation model.

[c] Attempts to achieve geometry optimization of this solvation pattern in a β conformation led to a change in conformation to PPII. In order to understand the energetics, these structures were examined via geometry optimization with ϕ fixed at –160˚, as is present in the structure with implicit solvation. Torsion angle scans from $\phi = -140^{\circ}$ to $\phi = -180^{\circ}$ confirmed that there was no energy minimum in this region for these structures. These structures, while not local energy minima, nonetheless exhibited zero imaginary (negative) frequencies.

Table 4. Geometric parameters of Ac-Pro_2-NMe_2 as a function of pattern of explicit solvation and combinations of proline ring pucker.[a]

	explicit C=O solvation			*d* O•••C=O Ac•••P1, P1•••P2 (Å)				(*ϕ*,*ψ*) P1, (*ϕ*,*ψ*) P2 (°)		
name	Ac	P1	P2	*exo-exo*	*exo-endo*	*endo-exo*	*endo-endo*	*exo-exo*	*exo-endo*	*endo-exo*
0HF	–	–	–	2.84, 2.86	2.89, 3.08	3.09, 2.86	3.06, 3.06	-59 137, -61 152	-61 149, -72 163	-71 164, -59 152
2HF	–	–	HFH F	2.81, 2.75	2.86, 2.98	3.08, 2.70	3.03, 2.99	-58 134, -55 149	-61 149, -71 157	-71 166, -52 140
3HF	–	HF	HFH F	2.84, 2.67	2.78, 2.93	3.01, 2.66	2.97, 2.95	-58 155, -50 141	-56 147, -70 153	-69 164, -50 139
4HF	HF	HF	HFH F	2.81, 2.67	2.77, 2.93	3.01, 2.67	2.92, 2.96	-57 155, -50 142	-56 148, -69 156	-70 166, -51 140
5HF	HF HFH	F HFH	F HFH	2.73, 2.75 2.99,	2.77, 3.11 2.93,	2.91, 2.76 3.14,	2.92, 3.11	-53 149, -49 150 -62 168, -54	-57 148, -74 173 -59 158, -77	-65 162, -51 151 -74 172, -57
5aHF	F HFH	HF HFH	F	2.69 2.94,	3.04 2.91,	2.72 3.11,	3.14, 3.04	143 -58 165, -55	168 -59 155, -78	146 -72 173, -59
5bHF	F HFH	F HFH	HF HFH	2.78 2.94,	3.17 2.89,	2.84 3.09,	3.11, 3.22	149 -58 165, -53	170 -57 155, -76	151 -71 172, -56
6HF	F	F	F	2.75	3.12	2.79	3.11, 3.18	150	173	152

[a] Geometry optimization and analysis was conducted only on the PPII conformation with a *trans* amide bond. Geometry optimization was conducted using the M06-2X functional and the 6-311++G(2d,2p) basis set in implicit water. *d* = intercarbonyl O•••C distance in Å for the Ac•••Pro1 and Pro1•••Pro2 n→π* interactions, respectively.

Table 5. Relative energies of Ac-Pro_2-NMe_2 as a function of the combination of proline ring pucker and explicit solvation.[a]

				E_{rel}[a], kcal mol^{-1}			
	explicit C=O solvation			P1-P2 ring pucker			
name	Ac	P1	P2	*exo-exo*	*exo-endo*	*endo-exo*	*endo-endo*
0HF	–	–	–	+1.4	+0.7	+0.8	*0.0*[a]
2HF	–	–	HFHF	–0.3	+0.3	–1.1	*0.0*[a]
3HF	–	HF	HFHF	–0.6	+0.1	–1.3	*0.0*[a]
4HF	HF	HF	HFHF	–0.6	+0.2	–1.3	*0.0*[a]
5HF	HF	HFHF	HFHF	–0.4	–0.1	–0.3	*0.0*[a]
5aHF	HFHF	HF	HFHF	–0.0	+1.3	–1.7	*0.0*[a]
5bHF	HFHF	HFHF	HF	+0.8	+1.0	–0.1	*0.0*[a]
6HF	HFHF	HFHF	HFHF	+0.0	+0.7	–1.0	*0.0*[a]

[a] Energies were determined using the MP2 method and the 6-311++G(3d,3p) basis set in implicit water. Energies (E_{rel}) are referenced to the peptide with *endo* ring pucker on both prolines with the indicated solvation.

References

(1) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001**, *123*, 777-778.

(2) Hinderaker, M. P.; Raines, R. T. An electronic effect on protein structure. *Protein Sci.* **2003**, *12*, 1188-1194.

(3) Hodges, J. A.; Raines, R. T. Energetics of an n->pi* interaction that impacts protein structure. *Org. Lett.* **2006**, *8*, 4695-4697.

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

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

(6) Zondlo, N. J. Fold Globally, Bond Locally. *Nat. Chem. Biol.* **2010**, *6*, 567-568.

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

(8) Wilhelm, P.; Lewandowski, B.; Trapp, N.; Wennemers, H. A Crystal Structure of an Oligoproline PPII-Helix, at Last. *J. Am. Chem. Soc.* **2014**, *136*, 15829-15832.

(9) Choudhary, A.; Kamer, K. J.; Raines, R. T. An n ->pi* Interaction in Aspirin: Implications for Structure and Reactivity. *J. Org. Chem.* **2011**, *76*, 7933-7937.

(10) Newberry, R. W.; Raines, R. T. n ->pi(star) interactions in poly(lactic acid) suggest a role in protein folding. *Chem. Commun.* **2013**, *49*, 7699-7701.

(11) Newberry, R. W.; Raines, R. T. A Key n ->pi* Interaction in N-Acyl Homoserine Lactones. *ACS Chem. Biol.* **2014**, *9*, 880-883.

(12) Kilgore, H. R.; Raines, R. T. n ->pi* Interactions Modulate the Properties of Cysteine Residues and Disulfide Bonds in Proteins. *J. Am. Chem. Soc.* **2018**, *140*, 17606-17611.

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

(14) Muchowska, K. B.; Pascoe, D. J.; Borsley, S.; Smolyar, I. V.; Mati, I. K.; Adam, C.; Nichol, G. S.; Ling, K. N. B.; Cockroft, S. L. Reconciling Electrostatic and n ->pi* Orbital Contributions in Carbonyl Interactions. *Angew. Chem., Int. Ed.* **2020**, *59*, 14602-14608.

(15) Sahariah, B.; Sarma, B. K. Spectroscopic evidence of n -> pi* interactions involving carbonyl groups. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26669-26681.

(16) Vik, E. C.; Li, P.; Pellechia, P. J.; Shimizu, K. D. Transition-State Stabilization by n ->pi* Interactions Measured Using Molecular Rotors. *J. Am. Chem. Soc.* **2019**, *141*, 16579-16583.

(17) Costantini, N. V.; Ganguly, H. K.; Martin, M. I.; Wenzell, N. A.; Yap, G. P. A.; Zondlo, N. J. The distinct conformational landscapes of 4S-substituted prolines that promote an endo ring pucker. *Chem. Eur. J.* **2019**, *25*, 11356-11364.

(18) Shi, Z. S.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9190-9195.

(19) Shi, Z. S.; Woody, R. W.; Kallenbach, N. R. Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **2002**, *62*, 163-240.

(20) Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R. The Pentapeptide GGAGG Has PII Conformation. *J. Am. Chem. Soc.* **2003**, *125*, 8092-8093.

(21) Chen, K.; Liu, Z. G.; Kallenbach, N. R. The polyproline II conformation in short alanine peptides is noncooperative. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15352-15357.

(22) Liu, Z. G.; Chen, K.; Ng, A.; Shi, Z. S.; Woody, R. W.; Kallenbach, N. R. Solvent dependence of PII conformation in model alanine peptides. *J. Am. Chem. Soc.* **2004**, *126*, 15141-15150.

(23) Tiffany, M. L.; Krimm, S. Circular Dichroism of Poly-L-Proline in an Unordered Conformation. *Biopolymers* **1968**, *6*, 1767-1770.

(24) Tiffany, M. L.; Krimm, S. Extended Conformations of Polypeptides and Proteins in Urea and Guanidine Hydrochloride. *Biopolymers* **1973**, *12*, 575-587.

(25) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry* **2005**, *44*, 6269-6275.

(26) Elam, W. A.; Schrank, T. P.; Campagnolo, A. J.; Hilser, V. J. Temperature and Urea Have Opposing Impacts on Polyproline II Conformational Bias. *Biochemistry* **2013**, *52*, 949-958.

(27) Chellgren, B. W.; Creamer, T. P. Effects of H_2O and D_2O on polyproline II helical structure. *J. Am. Chem. Soc.* **2004**, *126*, 14734-14735.

(28) Harrington, W. F.; Sela, M. Studies on the Structure of Poly-L-Proline in Solution. *Biochim. Biophys. Acta* **1958**, *27*, 24-41.

(29) Steinberg, I. Z.; Berger, A.; Katchalski, E. Reverse mutarotation of poly-L-proline. *Biochim. Biophys. Acta* **1958**, *28*, 647-648.

(30) Bovey, F. A.; Hood, F. P. Circular Dichroism Spectrum of Poly-L-Proline. *Biopolymers* **1967**, *5*, 325-326.

(31) Schimmel, P. R.; Flory, P. J. Confromational energies and configurational statistics of copolypeptides containing L-Proline. *J. Mol. Biol.* **1968**, *34*, 105-120.

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

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

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

(35) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and

transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

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

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

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

(39) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron-Affinities of the 1st-Row Atoms Revisited - Systematic Basis-Sets and Wave-Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.

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

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

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

(43) Papajak, E.; Zheng, J. J.; Xu, X. F.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027-3034.

(44) Purvis, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model - the Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910-1918.

(45) Pople, J. A.; Headgordon, M.; Raghavachari, K. Quadratic Configuration-Interaction - a General Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968-5975.

(46) Ford, T. A.; Glasser, L. Ab initio calculations of the structural, energetic and vibrational properties of some hydrogen bonded and van der Waals dimers .3. The formaldehyde dimer. *J. Mol. Struct. THEOCHEM* **1997**, *398*, 381-394.

(47) Hermida-Ramon, J. M.; Rios, M. A. A new intermolecular polarizable potential for a formaldehyde dimer. Application to liquid simulations. *J. Phys. Chem. A* **1998**, *102*, 10818-10827.

(48) Dolgonos, G. A. Which isomeric form of formaldehyde dimer is the most stable - a high-level coupled-cluster study. *Chem. Phys. Lett.* **2013**, *585*, 37-41.

(49) Van Dornshuld, E.; Holy, C. M.; Tschumper, G. S. Homogeneous and Heterogeneous Noncovalent Dimers of Formaldehyde and Thioformaldehyde: Structures, Energetics, and Vibrational Frequencies. *J. Phys. Chem. A* **2014**, *118*, 3376-3385.

(50) Rezac, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038-5071.

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

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

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

(54) Scheiner, S.; Kar, T. Effect of solvent upon CH center dot center dot center dot O hydrogen bonds with implications for protein folding. *J. Phys. Chem. B* **2005**, *109*, 3681-3689.

(55) Kamer, K. J.; Choudhary, A.; Raines, R. T. Intimate Interactions with Carbonyl Groups: Dipole-Dipole or n ->pi*? *J. Org. Chem.* **2013**, *78*, 2099-2103.

(56) Siebler, C.; Maryasin, B.; Kuemin, M.; Erdmann, R. S.; Rigling, C.; Grunenfelder, C.; Ochsenfeld, C.; Wennemers, H. Importance of dipole moments and ambient polarity for the conformation of Xaa-Pro moieties - a combined experimental and theoretical study. *Chem. Sci.* **2015**, *6*, 6725-6730.

(57) Wilson, A. K.; vanMourik, T.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations .6. Sextuple zeta correlation consistent basis sets for boron through neon. *J. Mol. Struct. THEOCHEM* **1996**, *388*, 339-349.

(58) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.

(59) Kurtz, J.; Berger, A.; Katchalski, E. Mutarotation of Poly-L-Proline. *Nature* **1956**, *178*, 1066-1067.

(60) Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E. The Configurational Changes of Poly-L-Proline in Solution. *J. Am. Chem. Soc.* **1960**, *82*, 5263-5279.

(61) Forsythe, K. H.; Hopfinger, A. J. Influence of Solvent on Secondary Structures of Poly(L-Alanine) and Poly(L-Proline). *Macromolecules* **1973**, *6*, 423-437.

(62) Kuemin, M.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. Effects of Terminal Functional Groups on the Stability of the Polyproline II Structure: A Combined Experimental and Theoretical Study. *J. Am. Chem. Soc.* **2009**, *131*, 15474-15482.

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

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

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

(66) Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N. Interplay of Hydrogen Bonds and n->pi* Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682-18688.

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

(68) Walsh, S. T. R.; Cheng, R. P.; Wright, W. W.; Alonso, D. O. V.; Daggett, V.; Vanderkooi, J. M.; DeGrado, W. F. The hydration of amides in helices; a comprehensive picture from molecular dynamics, IR, and NMR. *Protein Sci.* **2003**, *12*, 520-531.

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

(70) Rezac, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.

(71) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034-12046.

(72) Newberry, R. W.; Bartlett, G. J.; VanVeller, B.; Woolfson, D. N.; Raines, R. T. Signatures of n ->pi* interactions in proteins. *Protein Sci.* **2014**, *23*, 284-288.

(73) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002**, *124*, 2497-2505.

(74) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. Code for collagen's stability deciphered. *Nature* **1998**, *392*, 666-667.

(75) Hodges, J. A.; Raines, R. T. Stereoelectronic effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers. *J. Am. Chem. Soc.* **2003**, *125*, 9262-9263.

(76) Horng, J. C.; Raines, R. T. Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006**, *15*, 74-83.

(77) Sonntag, L. S.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. The "azido gauche effect"-implications for the conformation of azidoprolines. *J. Am. Chem. Soc.* **2006**, *128*, 14697-14703.

(78) Kumin, M.; Sonntag, L. S.; Wennemers, H. Azidoproline containing helices: Stabilization of the polyproline II structure by a functionalizable group. *J. Am. Chem. Soc.* **2007**, *129*, 466-467.

(79) Tressler, C. M.; Zondlo, N. J. (2S,4R)- and (2S,4S)-Perfluoro-tert-butyl 4-Hydroxyproline: Two Conformationally Distinct Proline Amino Acids for Sensitive Application in 19F NMR. *J. Org. Chem.* **2014**, *79*, 5880-5886.

(80) Creamer, T. P. Left-handed polyproline II helix formation is (very) locally driven. *Proteins* **1998**, *33*, 218-226.

(81) Chen, K.; Liu, Z. G.; Zhou, C. H.; Shi, Z. S.; Kallenbach, N. R. Neighbor effect on PPII conformation in alanine peptides. *J. Am. Chem. Soc.* **2005**, *127*, 10146-10147.

(82) Ferreon, J. C.; Hilser, V. J. The effect of the polyproline II (PPII) conformation on the denatured state entropy. *Protein Science* **2003**, *12*, 447-457.

(83) Fitzkee, N. C.; Rose, G. D. Sterics and solvation winnow accessible conformational space for unfolded proteins. *J. Mol. Biol.* **2005**, *353*, 873-887.

(84) Fleming, P. J.; Fitzkee, N. C.; Mezei, M.; Srinivasan, R.; Rose, G. D. A novel method reveals that solvent water favors polyproline II over beta-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). *Protein Sci.* **2005**, *14*, 111-118.

(85) Doyle, A. G.; Jacobsen, E. N. Small-molecule H-bond donors in asymmetric catalysis. *Chem. Rev.* **2007**, *107*, 5713-5743.

(86) Uyeda, C.; Jacobsen, E. N. Transition-State Charge Stabilization through Multiple Non-covalent Interactions in the Guanidinium-Catalyzed Enantioselective Claisen Rearrangement. *J. Am. Chem. Soc.* **2011**, *133*, 5062-5075.

(87) Mammen, M.; Choi, S.-K.; Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem., Int. Ed.* **1998**, *37*, 2754-2794.

(88) Makhatadze, G. I.; Privalov, P. L. Protein Interactions with Urea and Guanidinium Chloride - a Calorimetric Study. *J. Mol. Biol.* **1992**, *226*, 491-505.

(89) Pappu, R. V.; Rose, G. D. A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **2002**, *11*, 2437-2455.

(90) Schowen, R. L. Mechanistic Deductions from Solvent Isotope Effects. *Prog. Phys. Org. Chem.* **1972**, *9*, 275-332.

(91) Schowen, K. B.; Schowen, R. L. Solvent Isotope Effects on Enzyme Systems. *Methods in Enzymology* **1982**, *87C*, 551-606.