

Conformational Analysis of $n \rightarrow \pi^*$ Interactions in Collagen Triple Helix Models

Felicia A. Etzkorn, Rachel I. Ware, Amanda M. Pester, Diego Troya*

Department of Chemistry, Virginia Tech, Blacksburg VA, 24061

KEYWORDS collagen, triple-helix, $n \rightarrow \pi^*$ interactions, polyproline type II helix, conformational analysis, *ab initio*

ABSTRACT *Ab initio* calculations of three models of collagen at positions Pro–Pro–Gly **1**, Pro–Gly–Pro **2**, and Gly–Pro–Pro **3**, were performed to assess the conformational variation of $n \rightarrow \pi^*$ contributions to the stability of the collagen triple helix. Full conformational analyses by relaxed potential-energy scans of the Ψ dihedral angle of the central residue in models **1**, **2**, and **3** revealed the presence of several $n \rightarrow \pi^*$ interactions. In model **2**, with Gly as the central residue, both the Φ and Ψ dihedral angles of Gly were scanned. Most local minima of each model contained one or two $n \rightarrow \pi^*$ interactions, with pyramidalization at the π^* carbon. We observed pyramidalization at the $n \rightarrow \pi^*$ donor amide nitrogens. Minima with hydrogen bond or non-native $n \rightarrow \pi^*$ interactions compete with the collagen stabilizing $n \rightarrow \pi^*$ interactions. The collagen-like $n \rightarrow re-\pi^*$ conformation was found as the global minimum only in model **3**. The global minimum of **1** had a 5-membered ring hydrogen-bond with an additional weak $n \rightarrow si-\pi^*$ interaction. The

global minimum of **2** was in the extended conformation. We predict that the $n \rightarrow \pi^*$ interactions found in native collagen, while individually small, cumulatively contribute to the stability of the triple helix conformation of collagen.

1. INTRODUCTION

Manifold efforts have been devoted to understanding the folding and stability of the collagen triple helix.¹⁻⁸ The collagen structure has three left-handed single-strand polyproline type II (PPII) helices coiled into a right-handed triple helix.⁹ We found that Gly-Pro and Pro-Pro trans-locked alkene mimics surprisingly destabilized the collagen triple helix.¹⁰⁻¹¹ This led us to investigate the complete local conformational space at each of the three positions of a canonical collagen triple helix: Pro-Pro-Gly. Raines and coworkers have shown with small, single-strand models, both experimentally and computationally, that an electronic $n \rightarrow \pi^*$ interaction stabilizes the PPII conformation of collagen.^{1-5, 7-8} Kamer et al. found that the $n \rightarrow \pi^*$ interaction is not a dipole-dipole interaction.¹² Wennemers and coworkers recently showed that interstrand hydrogen bonding is significantly more important in stabilizing the collagen triple helix than the Yaa C γ -exo pucker.¹³ Our current efforts are designed to examine energetically competing local conformational effects of $n \rightarrow \pi^*$ and intramolecular hydrogen bonding interactions within a single collagen strand. Preorganization of local conformational structure plays a significant role in collagen folding and stability.²⁻⁸

The geometries of $n \rightarrow \pi^*$ interactions are similar to the Bürgi-Dunitz trajectory found in carbonyl reactions.⁵ Bürgi and Dunitz found that, as a nucleophile approaches within 3 Å (δ_{BD}) of a carbonyl, the angle (τ_{BD} : $\angle O_i \cdots C_{i+1}=O$) formed by the approaching nucleophile and the C=O bond is approximately 107° (Figure 1).¹⁴⁻¹⁵ In the $n \rightarrow \pi^*$ interaction that has been described

in collagen, the lone pair of a Pro carbonyl O_i donates electron density to the *re*-face of a Hyp $C_{i+1}=O$ π^* orbital; we call this the forward direction (Figure 1). As a result of this donation of electron density into the $C=O$ π^* orbital, the carbon develops a slight amount of sp^3 character, i.e. pyramidalization, measured by the angle Θ_{BD} (Figure 1).¹⁶⁻¹⁷ We found significant pyramidalization at the Pro nitrogen in some conformations, measured by the angle Θ_N (Figure 1).

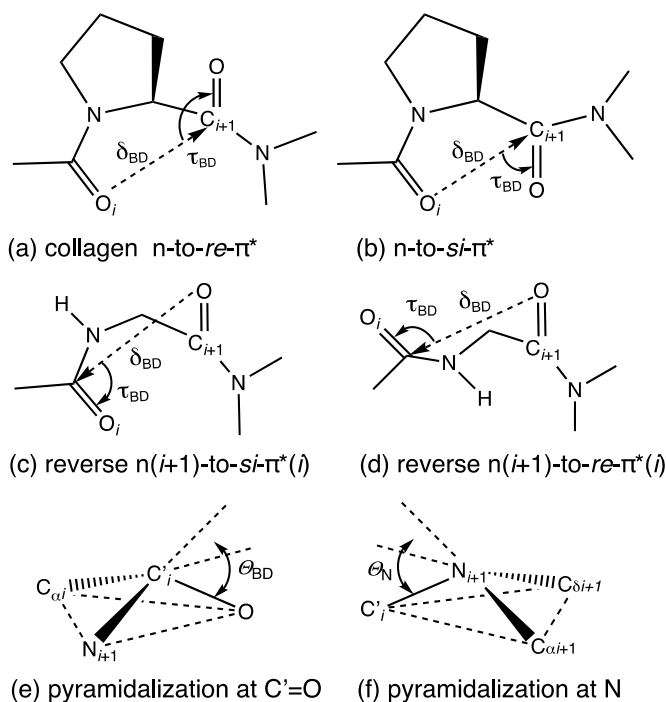


Figure 1. Bürgi-Dunitz trajectory distances (δ_{BD}) and angles (τ_{BD}) of the oxygen lone pair donor (n) with the carbonyl acceptor (π^*) of an adjacent amide bond describe an $n\rightarrow\pi^*$ interaction found in our models: (a) PPII-like $n\rightarrow re\text{-}\pi^*$, (b) $n\rightarrow si\text{-}\pi^*$, (c) reverse $n_{i+1}\rightarrow si\text{-}\pi^*_i$, and (d) reverse $n_{i+1}\rightarrow re\text{-}\pi^*_i$ interactions. Pyramidalization at (e) $C'=O$ is defined as Θ_{BD} ,¹⁶ and at (f) N is defined as Θ_N , with positive values representing endo-Pro lone pairs.

We now identify $n \rightarrow \pi^*$ interactions in collagen-like models that donate to the opposite *si*-face of the $C_{i+1}=O$ π^* acceptor, as well as in what we call the reverse direction, e.g. O_{i+1} lone pair n donating into either the *re*- or *si*-face of the $C_i=O$ π^* orbital (Figure 1). Forward and reverse interactions can occur within the same pair of carbonyls simultaneously. Rahim et al. refer to two interactions involving the same two carbonyls as reciprocal $n \rightarrow \pi^*$ interactions, which they found in PPII protein structures,¹⁸ and which we now report in collagen models.

Three reduced-dimensionality collagen peptide mimics were modeled in this work (Figure 2). Pro–Pro–Gly (PPG) **1**, Pro–Gly–Pro (PGP) **2**, and Gly–Pro–Pro (GPP) **3** were derived from collagen-like peptide high-resolution X-ray crystal structure, 1CAG,¹⁹ by excising the relevant parts of a sequence from the middle of the triple helix (Figure 2). Models **1–3** were used to compare the energies and geometries of the full range of conformations around each central residue Ψ dihedral angle, and around the Gly Φ angle of model **2**.²⁰ The Pro Φ angles of models **1** and **3** are restricted within the 5-membered ring. All of the models were in the trans conformation of the Pro–Hyp–Gly sequence of collagen.⁹

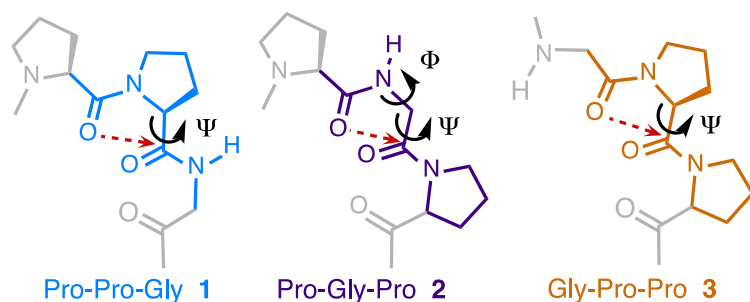


Figure 2. Structures of the PPG **1** (blue), PGP **2** (purple), and GPP **3** (orange) models used to calculate conformational geometries and relative energies. Grey portions of structures are shown to place them in the context of the collagen tripeptide repeat. The dihedral angles, Φ : $C_i-N_{i+1}-$

$C\alpha_{i+1}-C'_{i+1}$, and/or Ψ : $N_{i+1}-C\alpha_{i+1}-C'_{i+1}-N_{i+2}$, of the central residue were scanned in the calculations.

2. COMPUTATIONAL METHODS

Starting structures were derived from residues Hyp20–Gly21–Pro22–Hyp23–Gly24 from the high resolution collagen triple helix structure 1CAG of Bella et al. and modeled as shown in Figure 2.¹⁹ Initial geometry optimizations were performed with second-order Møller-Plesset (MP2) calculations and the 6-31+G(d) basis set with a continuum solvation model of water using Gaussian09 with default convergence thresholds through the WebMO interface.²¹⁻²² MP2 calculations provide an accurate, yet affordable calculation method for these relatively large models.²³ Diffuse functions on heavy atoms were used to improve the likelihood of capturing $n\rightarrow\pi^*$ interactions. All scans performed in this work were relaxed, i.e., all degrees of freedom except for the one being scanned were allowed to fully optimize.

For PPG **1**, scan of the Ψ angle was performed with a 10° step size from -170° to 180° ; the step size was reduced to 5° near the PPII $n\rightarrow\pi^*$ geometry from $\Psi +110^\circ$ to $+160^\circ$ (Figure 3a). For PGP **2**, a scan of the Ψ angle was used with a 10° step size at a fixed angle Φ of -72° , corresponding to the Φ angle in the crystallographic collagen structure (Figure 3a).¹⁹ In addition, for PGP **2**, scans of both the Φ and Ψ angles were performed with a 30° step size (Figure 3b). This double dihedral scan required 169 geometry optimizations, which were accelerated via use of the wB97XD density functional in combination with the 6-31G(d) basis set.²⁵ Using the wB97XD geometries, MP2/6-31G(d) single-point energy calculations were conducted to obtain the contour plot in Figure 3(b). Letters corresponding to minima and maxima of PGP **2** are given on the plot in Figure 3b to designate stationary points in Table 1. For GPP **3**, the scan of the Ψ dihedral angle used a 10° step size from -170° to 180° (Figure 3a).

For all three models, minima and maxima (transition states, TS) identified from the scans were fully optimized unrestrained at the MP2/6-311+G(2d,p) level (Table 1 and Supporting Information). Natural Bond Orbitals (NBOs) and second order perturbation energies stemming from orbital interactions (E2) were calculated at MP2 6-311+G(2d,p) after geometry optimization.²⁴ Energies relative to the lowest energy conformer and Ψ angles of the minima and TS from the fully optimized models **1** – **3** are given in Table 1. Stationary points are labeled as “min” or “TS” and specified by letter. We verified that the split-valence 6-311+G(2d,p) basis set provided reasonable energetic accuracy via sample comparisons with the much larger aug-cc-pVTZ basis set (SI). Animation of imaginary frequencies confirmed relevant Ψ dihedral motion at the transition states. Maximum **2k** did not converge to a TS, so the energy in Table 1 corresponds to an MP2/6-311+G(2d,p) single-point calculation with the geometry at the global maximum from the wB97XD/6-31G(d) scan. Maximum **3a** did not converge to a TS at the higher level, so a single-point energy was calculated at the MP2/6-311+G(2d,p) level from the TS-optimized structure at the MP2/6-31G+(d) level. Natural bond orbital (NBO) wavefunction analyses were used to visualize localized molecular orbitals corresponding to $n \rightarrow \pi^*$ interactions and intramolecular hydrogen bonds (Figures 4-6). Pyramidalization was measured by Θ_{BD} , the angle between the $C_{\alpha i}-C'_i-N_{i+1}$ plane and the $C'_i=O$ vector as described by Choudhary (Figure 1).¹⁶ Here, we define pyramidalization Θ_{BD} as positive when the carbonyl carbon at the apex is pointed toward the donor lone pair. Hydrogen bonding interactions were measured by the H-bond X---H distance (d_{HB}) and X---H---N angle ($\angle_{HB}$) (Table 1). We define a new parameter for pyramidalization at amide nitrogens; Θ_N is the angle between the $C_{\alpha i+1}-N_{i+1}-C_{\delta i+1}$ plane and the $N_{i+1}-C'_i$ vector (Figure 1). Pyramidalization Θ_N was measured to generate positive angles when

the lone pair on the nitrogen at the apex of the pyramid points in the endo Pro direction. Raw energies and complete geometric parameters are reported in the Supporting Information.

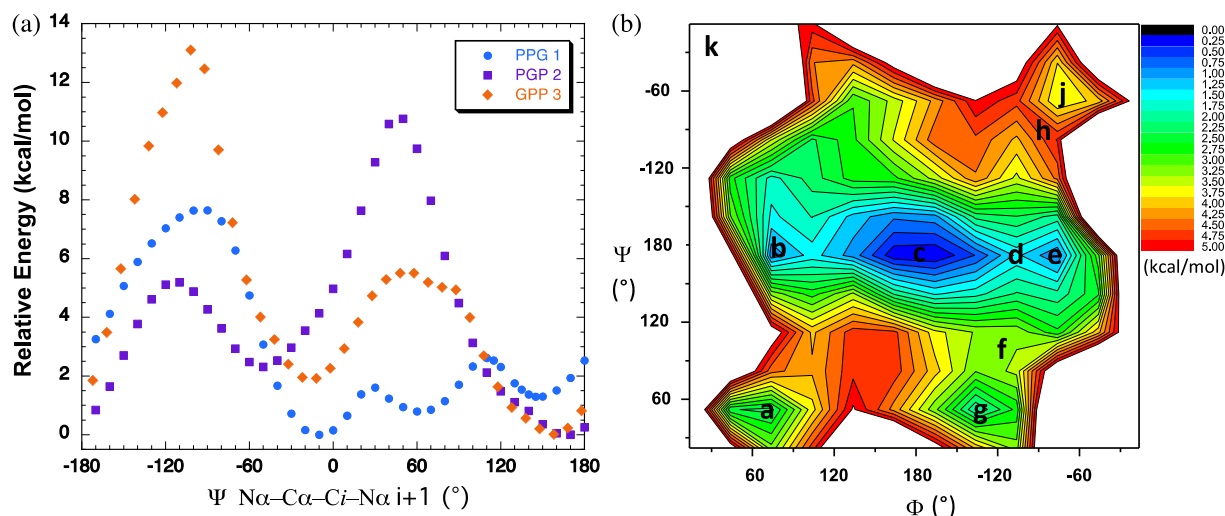


Figure 3. (a) PPG 1, PPG 2, and GPP 3 relative energies at the MP2/6-31+G(d) level are shown as a function of the Ψ dihedral angle. The Φ bond of **2** was fixed at -72° of the crystal structure, Gly21 in 1CAG.¹⁹ (b) PGP 2 relative energies at the MP2//wB97XD/6-31G(d) level are shown as a function of both Φ and Ψ angles. Letters on the plot correspond to PGP 2 labels in Table 1.

3. RESULTS AND DISCUSSION

The extent of interactions in the various minima was determined by measuring the $O_i \cdots C'_{i+1}$ distance (δ_{BD}), the $O_i \cdots C'_{i+1}=O$ angle (τ_{BD}), the degree of pyramidalization at the $C=O$ carbon θ_{BD} , and the NBO second order perturbation energy (NBO E2) of the orbital interactions (Table 1). The sum of the van der Waals radii ($O + C = 3.22 \text{ \AA}$)²⁶ minus the δ_{BD} distance gives an approximation of the amount of overlap between the n and π^* orbitals. For all $n \rightarrow \pi^*$ interactions, θ_{BD} values are small, in line with those observed in high resolution peptide and protein crystal structures.^{16, 27} For H-bonding interactions, the H-bond $X \cdots H$ distance (d_{HB}) and

X---H-N angle ($\angle_{\text{HB}}$) were measured (Table 1). Some of the $n \rightarrow \pi^*$ and H-bonds showed pyramidalization at the Pro nitrogen, Θ_{N} (Table 1).

Table 1. Relative energies and key geometric parameters of minima (blue), PPII-like minima (green) and maxima.

Model	Ψ^a (°)	ΔE^a (kcal/ mol)	π^* face/ H-bond	$\delta_{\text{BD}} /$ d_{HB} (Å)	$\tau_{\text{BD}} /$ $\angle_{\text{HB}}$ (°)	Θ_{BD} (°)	Θ_{N} (°)	NBO E2 ^b (kcal/mol)
PPG 1								
TS 1a	−94	7.0						
min 1b	−12	0.0	$\text{N}_{i+1} \cdots \text{HN}_{i+2}$	2.27	107		11.9	0.80 ^d
"	"	"	<i>si</i>	3.25	126	1.8		< 0.1
TS 1c	+30	1.4						
min 1d	+62	0.3	$\text{O}_i \cdots \text{HN}_{i+1}$	1.92	150		9.4	7.92 ^d
TS 1e	+112	2.3						
min 1f	+150	1.0	<i>re</i>	3.00	98	3.0	−3.3	0.46
"	"	"	<i>si-rev</i> ^c	3.28	85	−0.2		< 0.1
PGP 2								
min 2a	+45	1.4	<i>re</i>	2.92	102	1.2	8.6	1.29 ^d
min 2b	−178	0.8	<i>re-rev</i> ^c	3.16	90	0.7		0.13
"	"	"	<i>si</i>	3.29	85	0.6		< 0.1
min 2c	+176	0.0					0.7	
TS 2d	+171	1.4						
min 2e	+167	0.6	<i>re</i>	3.16	89	0.8	−0.4	0.19
"	"	"	<i>si-rev</i> ^c	3.15	89	0.7		< 0.1
TS 2f	+98	2.9						
min 2g	+54	1.5						
TS 2h	−79	3.8						
min 2j	−50	2.9	<i>si</i>	2.97	99	2.3	11.3	0.93 ^d
max 2k	−8	19.4						
GPP 3								
TS 3a	−105	13.1						
min 3b	−17	1.7	<i>si</i>	3.16	123	0.6	6.3	0.31
TS 3c	+36	5.1						
min 3d	+160	0.0	<i>re</i>	3.08	93	2.9	−3.2	0.41
"	"	"	<i>si-rev</i> ^c	3.22	87	−0.2		< 0.1

^aUnconstrained minima (min) or maxima (TS or max), Ψ angles, and relative energies optimized at the MP2/6-311+G(2d,p) level (except **2k**). ^bNBO second-order perturbation energy for the $n \rightarrow \pi^*$ or the H-bond, $n \rightarrow \sigma^*$ interaction. ^cReverse $n(i+1)$ -to- $\pi^*(i)$ interactions. ^dSum of both oxygen lone pair NBO E2 interacting energies in either the $n \rightarrow \pi^*$ or $n \rightarrow \sigma^*$ interaction.

Three conformational minima were found for model PPG **1** (Figure 3a). The relative conformational energies indicate that **1b** is the global minimum, with minima **1d** and **1f** close in energy (Table 1). The global minimum **1b** has an $n \rightarrow \pi^*$ interaction with a somewhat distorted π^* orbital shape (Figure 4). The geometry (δ_{BD} 3.25 Å, τ_{BD} 126°, θ_{BD} -1.8°) and NBO energy (< 0.1 kcal/mol) indicate a very non-ideal $n \rightarrow \pi^*$ interaction. Images of the overlap between the n and π^* or σ^* (H-bond) NBOs for PPG **1** models are shown in Figure 4.

The additional stability of **1b** can be attributed to an intramolecular, 5-membered ring, H-bond between the amide N_i ---H- N_{i+1} , with geometry, d_{HB} 2.27 Å, and $\angle_{HB}$ 107° (Table 1). The H-bond of **1b** can be visualized as an $n \rightarrow \sigma^*$ interaction (Figure 4). Pyramidalization at the H-bond acceptor N is quite pronounced (θ_N 11.9°). This is probably due the H-bond in addition to deconjugation of the N lone pair from the C=O due to its $n \rightarrow \pi^*$ donation. The NBO E2 energy (0.80 kcal/mol) indicates a relatively weak H-bond compared with the H-bond of **1d** (Figure 4). The angle of this 5-membered ring H-bond is one reason for its weakness, in addition to the H-bond acceptor N lone pair being tied up in an amide bond. Such weak 5-membered ring H-bonds at prolines have been calculated and proposed from experimental results in the transition states of peptidyl-prolyl isomerases, in which it is necessary for the proline N to pyramidalize in the transition state of the isomerization.²⁸⁻²⁹ The conformation of **1b** is very similar to the proposed enzymatic transition states of FKBP and Pin1.²⁸⁻²⁹

Minimum **1d** is characterized by an intramolecular, 7-membered ring, $C_i=O$ ---H- N_{i+1} hydrogen bond (d_{HB} 1.92 Å, $\angle_{HB}$ 150°), a well-known inverse γ -turn H-bond³⁰ that is much

stronger than the H-bond of **1b** (Table 1). The H-bond of **1d** can also be visualized as an $n \rightarrow \sigma^*$ interaction (Figure 4). In the **1d** 7-membered ring H-bonded conformation, the Pro N is also strikingly pyramidalized ($\angle_N 9.4^\circ$), even though there is no obvious interaction with the N lone pair. Possible explanations include alleviation of torsion angle strain in this medium-sized ring with five sp^2 centers, as well as deconjugation of the Pro N lone pair from the C=O that is donating electron density to the Gly HN. The geometry of the **1d** H-bond is such that both lone pairs of the C=O oxygen (only the stronger one is shown in Figure 4) donate to the acceptor H–N σ^* , so the NBO E2 energy of **1d** given in Table 1 is the sum of both NBO E2 energies (7.92 kcal/mol) (Supporting Information). The NBO E2 energies seem to overemphasize H-bonds compared with

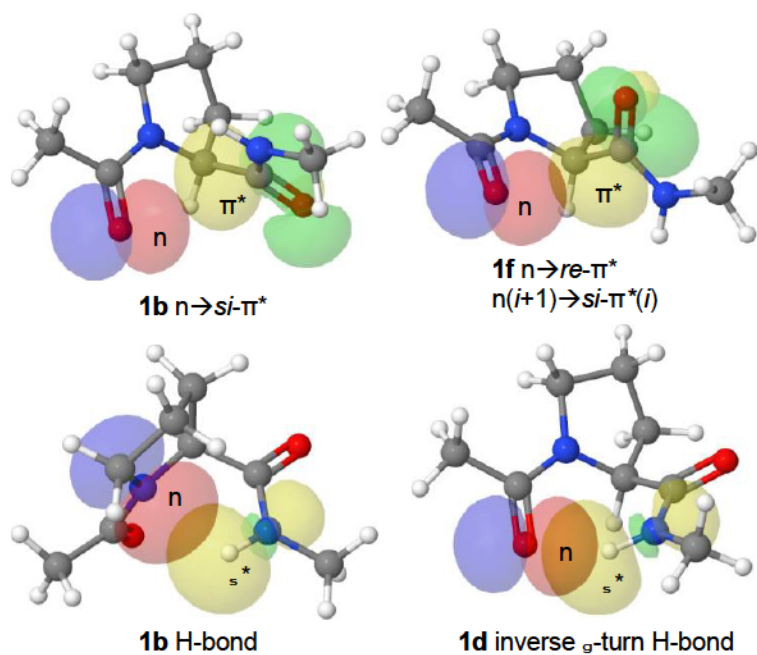


Figure 4. PPG **1** model NBO wavefunctions shown at the three energy minima (Table 1): **1b** with an $n \rightarrow si-\pi^*$ interaction, **1b** showing the N_i n to N_{i+1} –H σ^* H-bond orbitals, PPII-like **1f** with $n \rightarrow re-\pi^*$ and $n(i+1) \rightarrow si-\pi^*(i)$ (not shown) interactions—showing O_i lone pair n , and

$C_{i+1}=O$ anti-bonding π^* orbitals, and **1d** showing the $C_i=O$ n to $N_{i+1}-H$ σ^* H-bond orbitals. Occupied orbitals (n) are shown in blue/red, and unoccupied (π^* and σ^*) are shown in yellow/green. Atom color code: H: white, C: grey, N: blue, O: red.

$n \rightarrow \pi^*$ interactions, since **1d** is overall *less* stable than **1b** (Table 1). In the collagen triple helix, these intramolecular H-bonds would compete with the requisite interstrand H-bond, Pro $C=O$ to Gly $N-H$.³¹⁻³²

Intriguingly, PPG **1f** (Figure 4), which is near the native collagen PPII-like conformation, exhibits reciprocal forward $n \rightarrow re-\pi^*$ and reverse $n(i+1) \rightarrow si-\pi^*(i)$ interactions (Table 1). The forward interaction is stronger than the reverse. The PPG **1f** $n \rightarrow re-\pi^*$ interaction has a short δ_{BD} of 3.00 Å, with good O---C overlap (0.22 Å smaller than the sum of the C and O van der Waals radii), a τ_{BD} angle of 98°, the strongest pyramidalization for PPG **1** with Θ_{HB} of 3.0°, and NBO E2 of 0.46 kcal/mol (Table 1). Yet PPG **1f**, with $n \rightarrow re-\pi^*$ and $n \rightarrow si-\pi^*$ interactions, is 1.0 kcal/mol higher in energy than the global minimum **1b** that includes both weak H-bond and $n \rightarrow si-\pi^*$ interactions. In addition to pyramidalization at the acceptor $C=O$, PPG **1f** shows *exo* pyramidalization at the Pro nitrogen ($\Theta_N -3.3^\circ$). This appears to be due to repulsion between the Pro N_{i+1} lone pair and the Pro $C_{i+1}=O$ oxygen lone pair; the N---O distance is 2.80 Å. Deconjugation of the Pro N_{i+1} from the $C=O_i$ due to the $n \rightarrow re-\pi^*$ interaction is also likely to contribute. Exo pyramidalization could also allow a more favorable Pro ring conformation given the relatively strong 5-membered ring $n \rightarrow re-\pi^*$ interaction.

The $n \rightarrow si-\pi^*$ global minimum **1b**, the H-bonded minimum PPG **1d**, and the PPII-like minimum **1f**, have similar energies, with low barriers, TS **1c** and TS **1e**, between them (Table 1). There are no obvious repulsive interactions in these maxima, just a lack of stabilizing $n \rightarrow \pi^*$ or

H-bond interactions. Therefore, for model PPG **1**, conformations **1b** and **1d** compete with $n \rightarrow \pi^*$ stabilization of the native collagen conformation **1f** (Figure 4). H-bonded **1d** is also an intermediate between $n \rightarrow \pi^*$ conformations **1b** and **1f**, providing a low energy path between the two (Figure 3a). These minima are all within 1.0 kcal/mol in energy, resulting in a broad, shallow potential energy surface between Ψ about -60° and 180° (Figure 3a and Table 1).

Model PGP **2** has a more complex conformational energy profile due to flexibility at the central Gly residue.²⁰ In addition to the Φ -fixed scan (Figure 3a), both Φ and Ψ dihedral angles were varied every 30° , starting with the collagen crystal structure, 1CAG (Figure 3b). This model contains no stereocenters, so we expected a very symmetric plot. However, because we started with the model extracted from a single enantiomer, there were residual effects of the local environment, including the C γ -endo Pro ring conformation.^{1, 33-34} We chose to retain these idiosyncrasies in our model because we think they represent real local conformational differences found in collagen.

For the PGP **2** model, the global minimum **2c** is in the extended conformation (not shown), and does not exhibit any $n \rightarrow \pi^*$ stabilization.²⁰ The PPII-like minimum **2e** is only 0.6 kcal/mol above the global minimum **2c** with a low barrier TS **2d** between them (Table 1). Similar to PPII-like PPG **1f**, minimum **2e**, adopts a conformation with reciprocal forward $n \rightarrow \pi^*$ and reverse $n(i+1) \rightarrow \pi^*(i)$ interactions; these have very nearly the same δ_{BD} , τ_{BD} , and θ_{BD} geometric values (Table 1, Figure 5). The E2 energies of the two interactions are both very small. PPII-like PGP **2e** shows a slight exo pyramidalization at the Pro nitrogen ($\theta_N -0.4^\circ$), similar to PPG **1f**.

Newberry et al. showed that $n \rightarrow \pi^*$ interactions compete effectively with, and even strengthen, hydrogen-bonds in certain conformations.³⁵ In the collagen triple helix, the Pro C=O is involved in an interstrand H-bond with Gly N-H that withdraws electron density and strengthens the

$n \rightarrow \pi^*$ interaction when the π^* of the Pro C=O would be the acceptor as in PGP **2e**.^{19, 31, 35} These interactions are not within the scope of this study.

Minimum **2b** is approximately the mirror image of **2e**; only the Pro ring conformations are not mirrored (Figure 5). They also have nearly mirror image geometric interactions; **2b** has a reciprocal forward $n \rightarrow si-\pi^*$ and a reverse $n(i+1) \rightarrow re-\pi^*(i)$ (Figure 5). In the **2b** case, the $n \rightarrow si-\pi^*$ is slightly weaker. PGP **2b** is only 0.8 kcal/mol above the global minimum **2c** (Table 1). The geometry of models with two amides involved in reciprocal $n \rightarrow \pi^*$ interactions does not accommodate much pyramidalization at either C=O; we observe θ_{BD} less than 1° for each of these interactions in **2b** and **2e** (Table 1). The geometric parameters of the reciprocal $n \rightarrow \pi^*$ interactions are not significantly different.¹⁸ For example in **2b**, although the $n \rightarrow si-\pi^*$ δ_{BD} is slightly longer (3.29 Å) than the $n \rightarrow re-\pi^*$ δ_{BD} (3.16 Å), the τ_{BD} , θ_{BD} , and E2 values are very similar (Table 1). It appears that two weaker reciprocal $n \rightarrow \pi^*$ interactions give a slightly more stable overall conformation than one strong $n \rightarrow \pi^*$. For example, **2b** with two interactions is more stable than **2a** with a strong $n \rightarrow re-\pi^*$ interaction (ΔE 0.8 kcal/mol), and **2e** is more stable than **2j** (ΔE 2.3 kcal/mol).

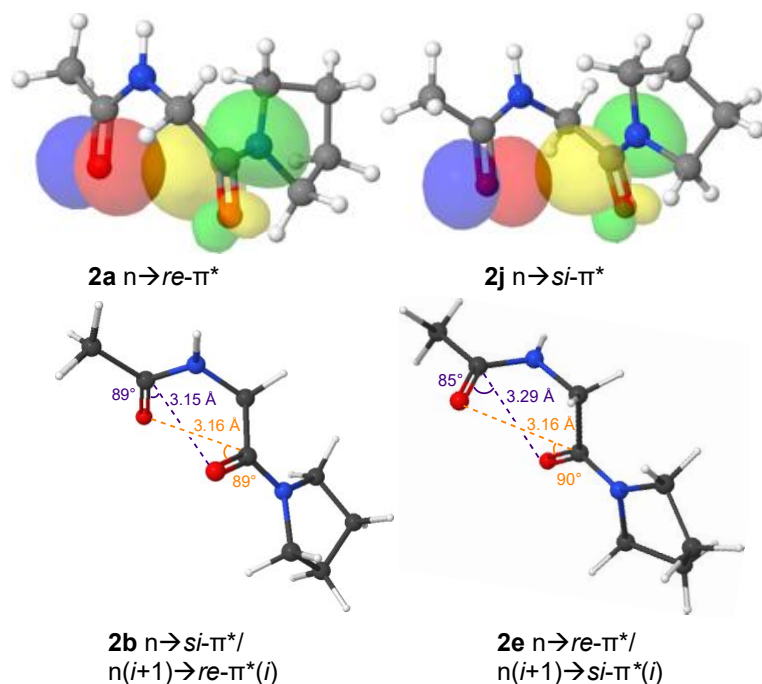


Figure 5. PPG **2** model relevant NBO wavefunctions at local minima: **2a** with $n \rightarrow re-\pi^*$; **2j** with $n \rightarrow si-\pi^*$; **2b** with $n \rightarrow si-\pi^*$ and $n(i+1) \rightarrow re-\pi^*(i)$; and PPII-like **2e** with $n \rightarrow re-\pi^*$ and $n(i+1) \rightarrow si-\pi^*(i)$ interactions (Table 1). Same orbital and atom color code as Fig. 4.

Minima **2a** and **2j** are approximately mirror image conformations with relatively strong single $n \rightarrow \pi^*$ interactions—both C=O lone pairs donate in these conformations (Table 1, Figure 5). The interacting carbonyls are nearly parallel, as in the α -helix conformation.¹⁶ The Pro- N_{i+1} is pyramidalized (**2a** $\angle_N$ 8.6°, **2j** $\angle_N$ 2.5°), probably due to deconjugation from the C=O, and sterics in the 5-membered ring of the $n \rightarrow \pi^*$ interactions (Table 1). For comparison, the extended conformation of global minimum **2c** has almost no pyramidalization at the Pro- N_{i+1} ($\angle_N$ 0.7°).

PPII-like minimum **2e** is separated from nearby minima **2c**, **2g**, and **2j** by low energy barriers, TS **2d**, TS **2f**, and TS **2h** (Table 1). Thus, the potential energy surface of model **2** is broad with many shallow minima, as expected due to the conformational flexibility of Gly,²⁰ with multiple $n \rightarrow \pi^*$ interactions energetically accessible (Figure 3b). The presence of multiple conformations

with low barriers between them for model **2** indicates even less stabilization of the PPII-like conformation by $n \rightarrow \pi^*$ interactions. Of course, the geometry of Gly within the context of the collagen triple helix is constrained by other factors. These results are in accord with Gly being the least pre-organized residue during collagen triple-helix folding.

Model GPP **3**, with two Pro rings, was much more computationally intensive than the first two models. We chose to use a pyrrolidine ring to model the second Pro ($i+1$) because the geometries of a similar dimethylamide model were different enough to be of concern. The global minimum **3d** was found to have both the PPII-like $n \rightarrow re-\pi^*$ interaction (δ_{BD} 3.08 Å, τ_{BD} 93°, θ_{BD} 2.9°, E2 0.41 kcal/mol), as initially discovered by Raines,³⁻⁷ as well as a weaker reverse $n(i+1) \rightarrow si-\pi^*(i)$ interaction (δ_{BD} 3.22 Å, τ_{BD} 87°, θ_{BD} 0.2°, E2 < 0.1 kcal/mol) (Figure 6). PPII-like GPP **3d** also shows *exo* pyramidalization at the Pro nitrogen (θ_N of -3.2°), almost identical to PPG **1f** (Table 1). This is probably due to repulsion between the Pro N_{i+1} lone pair and the $C_{i+1}=O$ oxygen lone pair; the N---O distance is 2.75 Å. Deconjugation of the Pro N_{i+1} from the $C=O_i$ due to the $n \rightarrow re-\pi^*$ interaction also allows pyramidalization. It could also be a favorable Pro ring conformation given the relatively strong 5-membered ring $n \rightarrow re-\pi^*$ interaction.

Similar to **1b**, local minimum **3b** contains only one relatively weak $n \rightarrow si-\pi^*$ interaction (δ_{BD} 3.16 Å, τ_{BD} 123°), with a small degree of pyramidalization (θ_{BD} 0.6°), and a low NBO E2 energy (0.31 kcal/mol). Pyramidalization at Pro- N_{i+1} is relatively large for **3b** (θ_N 6.3°), like PPG **1b**, probably due to deconjugation from the $C=O$, and strain due to steric interactions between the two Pro rings ($\gamma CH_2(i)-\delta CH_2(i+1)$ at 2.16 Å and $\delta CH_2(i)-\delta CH_2(i+1)$ at 2.29 Å (Figure 6). These steric interactions also prevent **3b** from adopting a more ideal $n \rightarrow si-\pi^*$ conformation (Table 1). Thus, **3b** is higher in energy relative to the global minimum PPII-like **3d** (Table 1). In the analogous **1b**, there is a favorable H-bond instead of unfavorable steric

interactions, so **1b** is more stable than **1f** (ΔE 1.0 kcal/mol), while **3b** is less stable than **3d** (ΔE 2.7 kcal/mol).

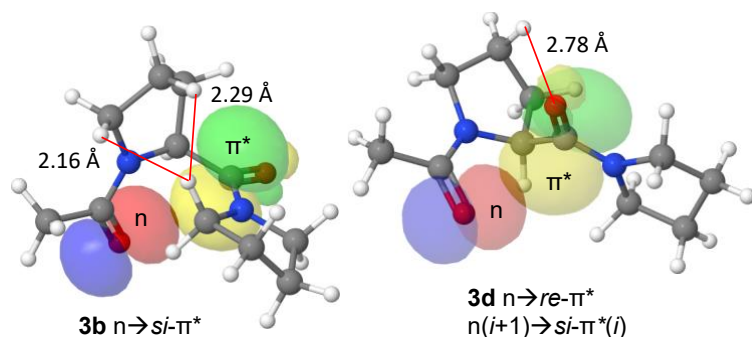


Figure 6. GPP **3** model relevant NBO wavefunctions showing minima: **3b** with an $n \rightarrow si-\pi^*$ interaction, and PPII-like **3d** with $n \rightarrow re-\pi^*$ and reverse $n(i+1) \rightarrow si-\pi^*(i)$ interactions (Table 1). Steric interactions destabilize **3b** relative to **3d** (interatomic distances shown as red lines). Same orbital and atom color codes as Fig. 4.

The barrier, TS **3c** (5.1 kcal/mol) between minima **3b** and **3d** is higher than the local barriers for the other two models (Table 1). Stabilization of the PPII-like conformation **3d** thus arises from several factors: (1) stabilization by the $n \rightarrow re-\pi^*$ interaction of **3d**, (2) destabilization by repulsion between rings in the competing $n \rightarrow si-\pi^*$ conformation **3b**, (3) the lack of a competing γ -turn H-bonded conformation, and (4) the higher local barrier **3c** between the minima.

The global maxima, **1a** and **3a**, around Ψ angles of -100° are due to obvious steric interactions between the two carbonyl O atoms in each model (Figure SI-1). This repulsion is compounded for GPP **3** by a $\beta CH_2(i)-\delta CH_2(i+1)$ steric interaction (Figure SI-1). For maximum **2k**, transition state searches near the global maximum all failed because of the flexibility of the Gly Φ and Ψ angles. The highest energy structure, PGP **2k** at $\Phi +14^\circ$ and $\Psi -8^\circ$, of the scan shows steric contact between $C=O(i)$ and $\delta CH_2(i+1)$ (Figure SI-1). The high rotational barriers for Φ and Ψ

angles constrain the local conformational space towards both sides of these global maxima, especially for PPG **1** and GPP **3**, while for PGP **2**, the conformational space is confined to broader well.

Near the PPII-like conformations, minima **1f**, **2e**, and **3d** have reasonably good $n \rightarrow re-\pi^*$ parameters (Table 1). Unexpectedly, the collagen-like $n \rightarrow re-\pi^*$ interactions in conformations **1f** and **2e** are not at the global minima. Only for GPP **3** does the PPII-like conformation **3d** occur at the global minimum. The latter is consistent with the previous calculations of $n \rightarrow \pi^*$ interactions in collagen models.²⁻⁵ Clearly, the $n \rightarrow \pi^*$ interactions are not the only local forces affecting the conformations of these collagen models.

The NBO E2 energies for single interactions do not tell the whole story of $n \rightarrow \pi^*$ interactions in these models. In addition, hydrogen bond NBO E2 interaction energies do not explain why **1d** (H-bond E2 7.92 kcal/mol) is *less* stable than **1b** (H-bond E2 0.8 kcal/mol) (Table 1). For model PGP **2**, the NBO $n \rightarrow \pi^*$ E2 energies also do not correlate with the total conformational energy differences (Table 1). In order to bring the two carbonyls closer for $n \rightarrow \pi^*$ interactions or intramolecular H-bonds, other steric interactions become repulsive. Only for model GPP **3**, are the NBO $n \rightarrow \pi^*$ energies in the same direction as the total conformational energies (Table 1). Again, we find the $n \rightarrow \pi^*$ interactions are not the only stabilizing or destabilizing forces in these collagen models.

4. CONCLUSIONS

The PPII conformation is a minimum in our models at each position of Pro–Pro–Gly, a repeating unit of the collagen triple helix, indicating that the PPII-like $n \rightarrow re-\pi^*$ conformation is readily accessible and contributes to collagen triple-helix stability.¹⁻⁸ Intriguingly, while the PPII-like $n \rightarrow re-\pi^*$ interaction leads to low-lying local minima in collagen models PPG **1f** and

PGP **2e** of this work, they do not correspond to the global minimum. PPG **1b** and **1d** are hydrogen-bonded conformations that are 1.0 and 0.7 kcal/mol, respectively, more stable than the PPII conformation **1f**. The extended conformation at the global minimum PGP **2c** is 0.6 kcal/mol more stable than the PPII conformation **2e**. Only in model GPP **3d** are the PPII conformation and the global minimum one and the same. Interestingly, in the PPII-like $n\rightarrow re-\pi^*$ conformations **1f**, **2e**, and **3d**, the Pro N lone pair is pyramidalized exo to the Pro ring, which appears to be a previously unrecognized indicator of a good $n\rightarrow re-\pi^*$ interaction in prolyl peptides and proteins.

Models similar to GPP **3d** have been studied before, but to our knowledge the native conformation has not been compared with neighboring conformers.¹⁻⁸ The conformational space of models like PPG **1** and PGP **2** has not been examined before. Placing these collagen-like conformations within the broader energetic landscape of possibilities leads to the conclusion that the PPII conformation is not the only stable conformation available. In the native collagen triple helix, additional long-range forces contribute to overcome the competition with the other interactions described in this paper.^{19, 31, 36} As with hydrogen bonds in folded proteins, this work suggests that each single $n\rightarrow re-\pi^*$ interaction does not contribute dramatically to the stability of the collagen triple helix, but the interactions combine to give a globally stable folded structure.³⁷

These results show that local conformations with $n\rightarrow\pi^*$ stabilizing interactions compete with other low-energy conformations on a relatively broad, shallow energy landscape. We are continuing these studies to design collagen-like alkene mimics that stabilize the triple helix conformation by restricting the backbone to *trans*-Pro PPII-like conformations, without losing $n\rightarrow\pi^*$ interactions.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge.

Supplementary Figure S1, Cartesian coordinates of model geometries at all energy minima and maxima described in the text. (PDF)

Energies, distances, and angles for all models. (xlsx)

AUTHOR INFORMATION

Corresponding Author

*fetzkorn@vt.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

NSF CHE-0749061, and the Department of Chemistry, Virginia Tech

ACKNOWLEDGMENT

No competing financial interests have been declared. We thank NSF CHE-0749061, and the Department of Chemistry, Virginia Tech for funding. We thank Nick Mayhall and Paul Carlier of Virginia Tech for helpful discussions. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper.

ABBREVIATIONS

Bürgi-Dunitz distance, δ_{BD} ; Bürgi-Dunitz angle τ_{BD} : $\angle O_i \cdots C_{i+1}=O$; Bürgi-Dunitz

pyramidalization angle, θ_{BD} ; amide nitrogen pyramidalization angle, θ_N ; PPII, polyproline type

II; PPG, Pro-Pro-Gly; PGP, Pro-Gly-Pro; GPP, Gly-Pro-Pro; Φ , angle $C(i)-N(i+1)-C\alpha(i+1)-$

$C'(i+1)$; Ψ , angle $N(i+1)-C\alpha(i+1)-C'(i+1)-N(i+2)$; MP2, Møller-Plesset-2; NBOs, Natural

Bond Orbitals; TS, transition state.

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* (4), 777-778.
2. 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* (11), 2497-2505.
3. Hinderaker, M. P.; Raines, R. T., An Electronic Effect on Protein Structure. *Protein Sci.* **2003**, *12* (6), 1188-1194.
4. Hodges, J. A.; Raines, R. T., Energetics of an $n \rightarrow \pi$ Interaction that Impacts Protein Structure. *Org. Lett.* **2006**, *8* (21), 4695-4697.
5. Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T., Nature of Amide Carbonyl--Carbonyl Interactions in Proteins. *J. Am. Chem. Soc.* **2009**, *131* (21), 7244-7246.
6. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.
7. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N., $n \rightarrow \pi^*$ Interactions in Proteins. *Nat. Chem. Biol.* **2010**, *6* (8), 615-620.
8. Newberry, R. W.; Raines, R. T., The $n \rightarrow \pi^*$ Interaction. *Acc. Chem. Res.* **2017**, *50* (8), 1838-1846.
9. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1955**, *176* (4482), 593-595.
10. Dai, N.; Etzkorn, F. A., Cis-Trans Proline Isomerization Effects on Collagen Triple-Helix Stability are Limited. *J. Am. Chem. Soc.* **2009**, *131* (38), 13728-13732.
11. Dai, N.; Wang, X. J.; Etzkorn, F. A., The Effect of a Trans-Locked Gly-Pro Alkene Isostere on Collagen Triple Helix Stability. *J. Am. Chem. Soc.* **2008**, *130* (16), 5396-5397.
12. Kamer, K. J.; Choudhary, A.; Raines, R. T., Intimate Interactions with Carbonyl Groups: Dipole-Dipole or $n \rightarrow \pi^*$? *J. Org. Chem.* **2013**, *78* (5), 2099-2103.
13. Erdmann, R. S.; Wennemers, H., Importance of Ring Puckering versus Interstrand Hydrogen Bonds for the Conformational Stability of Collagen. *Angew. Chem. Int. Ed.* **2011**, *50* (30), 6835-6838.
14. Bürgi, H. B.; Dunitz, J. D.; Shefter, E., Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.* **1973**, *95* (15), 5065-5067.

15. Bürgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G., Stereochemistry of Reaction Paths at Carbonyl Centres. *Tetrahedron* **1974**, *30* (12), 1563-1572.
16. Choudhary, A.; Raines, R. T., Signature of n $\rightarrow$ π^* interactions in α -helices. *Protein Sci* **2011**, *20* (6), 1077-81.
17. Newberry, R. W.; Bartlett, G. J.; VanVeller, B.; Woolfson, D. N.; Raines, R. T., Signatures of n $\rightarrow$ π^* interactions in proteins. *Protein Sci* **2014**, *23* (3), 284-8.
18. Rahim, A.; Saha, P.; Jha, K. K.; Sukumar, N.; Sarma, B. K., Reciprocal Carbonyl–Carbonyl Interactions in Small Molecules and Proteins. *Nat. Comm.* **2017**, *8*, 78.
19. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M., Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **1994**, *266* (5182), 75-81.
20. Ramachandran, G. N.; Sasisekharan, V., Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *23*, 283-437.
21. 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.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; 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 A.02*. Gaussian, Inc.: Wallingford CT, 2009.
22. *WebMO*, 14.0.007e; 2014.
23. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*. J. Wiley: New York, 2002.
24. Weinhold, F.; Landis, C. R., Natural Bond Orbitals and Extensions of Localized Bonding Concepts. *Chem. Educ.: Res. Pract. Eur.* **2001**, *2* (2), 91-104.
25. Chai, J.-D.; Head-Gordon, M., Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615-6620.
26. Bondi, A., Van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441-451.
27. Esposito, L.; Vitagliano, L.; Zagari, A.; Mazzarella, L., Pyramidalization of backbone carbonyl carbon atoms in proteins. *Protein Science* **2000**, *9* (10), 2038-2042.
28. Fischer, S.; Michnick, S.; Karplus, M., A Mechanism for Rotamase Catalysis by the FK506 Binding Protein (FKBP). *Biochemistry* **1993**, *32*, 13830-13837.
29. Mercedes-Camacho, A. Y.; Mullins, A. B.; Mason, M. D.; Xu, G. G.; Mahoney, B. J.; Wang, X.; Peng, J. W.; Etzkorn, F. A., Kinetic isotope effects support the twisted amide mechanism of Pin1 peptidyl-prolyl isomerase. *Biochemistry* **2013**, *52* (44), 7707-13.
30. Perskie, L. L.; Street, T. O.; Rose, G. D., Structures, basins, and energies: a deconstruction of the Protein Coil Library. *Protein Sci* **2008**, *17* (7), 1151-61.
31. Bella, J.; Brodsky, B.; Berman, H. M., Hydration Structure of a Collagen Peptide. *Structure* **1995**, *3* (9), 893-906.

32. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., Crystal structure of a collagen-like polypeptide with repeating sequence Pro-Hyp-Gly at 1.4 Å resolution: implications for collagen hydration. *Biopolymers* **2001**, *56* (1), 8-13.
33. Hodges, J. A.; Raines, R. T., Stereoelectronic Effects on Collagen Stability: the Dichotomy of 4-Fluoroproline Diastereomers. *J. Am. Chem. Soc.* **2003**, *125* (31), 9262-9263.
34. Horng, J. C.; Raines, R. T., Stereoelectronic Effects on Polyproline Conformation. *Protein Sci.* **2006**, *15* (1), 74-83.
35. Newberry, R. W.; Orke, S. J.; Raines, R. T., n $\rightarrow$ π^* Interactions are Competitive with Hydrogen Bonds. *Org. Lett.* **2016**, *18* (15), 3614-3617.
36. Kramer, R. Z.; Bella, J.; Brodsky, B.; Berman, H. M., The Crystal and Molecular Structure of a Collagen-Like Peptide with a Biologically Relevant Sequence. *J. Mol. Biol.* **2001**, *311* (1), 131-147.
37. Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N., Interplay of Hydrogen bonds and n $\rightarrow$ π^* Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135* (49), 18682-18688.

SYNOPSIS

