JCTC Journal of Chemical Theory and Computation

Conformational Studies of Polyprolines

Haizhen Zhong* and Heather A. Carlson

Department of Medicinal Chemistry, College of Pharmacy, The University of Michigan, 428 Church Street, Ann Arbor, Michigan 48109-1065

Received July 27, 2005

Abstract: Proline rich peptide sequences are very important recognition elements that have a significant bias toward the *all-trans*-polyproline type II (P_{II}) conformation. Our gas-phase quantum mechanics calculations at the B3LYP/6-31G* level of theory are in good agreement with previous experimental and theoretical studies. They show that *all-trans*-proline conformations are energetically more favorable than *all-cis*-polyprolines (P_I, polyproline type I). Estimates of the solvent effects show that the condensed phase can make the P_I form more populated in the correct environment. Our survey of proline oligomers in the Protein Data Bank confirmed that the predominant conformations from our calculations are seen experimentally. More importantly, we propose two new secondary structures for polyprolines, namely polyproline type III and type IV (P_{III} and P_{IV}). P_{III} is a right-handed, "square helix" from *trans*-proline oligomers. P_{IV} is a β-sheet form of *cis*-prolines. As suggested by its calculated IR spectra, the P_{III} form shares characteristics of both the P_I and P_{II} forms: it has *trans*-amide rotamers similar to P_{II} and forms a right-handed helix like P_I. We propose that the high-energy P_{III} form could exist as a conformational intermediate between P_I and P_{II}. These new forms also show that the handedness of polyproline helices depends not only on the peptide rotamers (cis or trans) but also on the values of the ψ torsions. Changing the ψ torsion from approximately 140° to approximately −30° causes the trans oligomers to flip from a typical left-handed P_{II} to a right-handed helix. Likewise, as the ψ torsion of the *cis*-proline oligomers changes from roughly 165° to −30°, the conformation changes from a characteristic right-handed P_I to a β-sheet.

Introduction

Recognition of proline-rich sequences plays a pivotal role in protein−protein interaction. The most common conformation of these sequences is the polyproline type-II (P_{II}) helix, a left-handed helix consisting of *trans*-prolines with three residues per turn (designated as 3_1 helix), $\phi \sim -75°$, and $\psi \sim 145°$. Another well-studied conformation of polyproline is the polyproline type-I (P_I), a right-handed helix with *cis*-prolines at 3.3 residues per turn (a 10_3 helix), $\phi \sim -75°$, and $\psi \sim 165°$. The ϕ and ψ angles for the P_I and P_{II} helices fall within the allowed β-strands region on the Ramachandran plot, due to the unique property of the imino proline linkage. Peptides adopting the P_{II} conformation have the propensity to function as recognition elements that bind to proline recognition domains, such as Src homology (SH3)[1], WW (named after a conserved Trp-Trp motif),[2] Enabled/VASP homology domains (EVH1),[3] the Gly-Tyr-Phe (GYF) domains,[4,5] and profilin proteins.[6,7] Recent findings have revealed a significant bias toward the P_{II} conformation in unfolded peptides and thus is a dominant component of the denatured states of proteins.[8−10]

Given the importance of proline-rich motifs, a full understanding of the P_{II} and P_I conformations of these species is highly desirable. A large body of experimental data has been reported since the crystallization of polyproline II[11] and polyproline I[12] four decades ago. However, some experiments

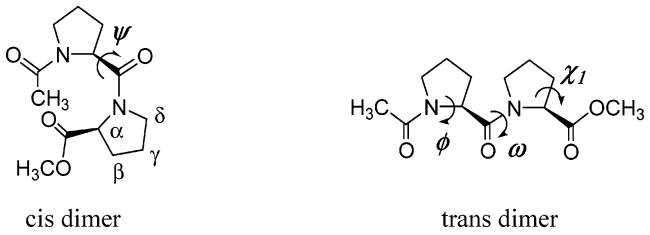
* Corresponding author phone: (336)334-5121; fax: (336)334-5402; e-mail: h_zhong@uncg.edu. Current address: Center for Drug Design, Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC 27402.

10.1021/ct050182t CCC: $33.50
Published on Web 02/02/2006

provide conflicting results. For example, Chao and Bersohn observed that in aqueous solutions the proline oligomer $^{+}H_2$-NPro-$(Pro)_n$-CO_2^- predominantly adopted the P_{II} conformation;[13] the conformations of a series of proline oligomers (*tert*-butyloxycarbonyl-L-Pro_n benzyl esters, Boc-$(Pro)_n$-OBn, $n = 2-6$) in chloroform are found to exist in nearly equal populations of *cis*- and *trans*-proline conformations when $n = 2$, 3, 4 and that they abruptly assume an *all-trans*, P_{II} helical structure when n is greater than 5.[14] However, Zhang and Madalengoitia have found that the ^{1}H NMR spectra of Boc-$(Pro)_n$-OBn (n=2−5) in $CDCl_3$ suggest the *trans*-amide bond is the predominant conformation.[15] Using the $[Pro_n+H]^+$ (n=5−11) model system, Counterman and Clemmer[16] discovered that the *all-cis*, P_I helix is favored and the helix adopted an extended form while the trans adopted a compact form, contrary to previous studies. The preference of P_I helix in Counterman and Clemmer's ionized model is attributed to the N-terminus cation that stabilizes the *cis*-proline. However, this type of ionized proline, capable of forming internal hydrogen bonds, is not found in proteins containing polyproline oligomers.

Many ab initio quantum mechanics calculations on proline motifs have been carried out on proline derivatives, such as *N*-acetyl-*N*′-methylprolineamide (Ac-Pro-NHMe),[17−20] *N*-formyl-L-prolineamide (For-L-Pro-NH_2),[21] proline dimer (For-L-Pro-L-Pro-NH_2),[22] and the neutral form of proline.[23,24] These proline monomer/dimer systems are capable of forming an internal hydrogen bond, increasing the percentage of the cis conformation in both the gas phase and less polar solvents. Tanaka and Scheraga found that trans conformers of the L-proline oligomers, Ac-$(Pro)_n$-OMe where $n = 2-5$, have the lowest energy based on a simplified molecular mechanics energy function.[25] Bour et al. carried out ab initio calculations using the SV and SV(P) split valence basis sets on the proline oligomer Ac-$(L\text{-}Pro)_9$-NHMe for both the P_I and P_{II} helical conformations and found that the relative energy of the optimized P_{II} is 3.3 kcal/mol below that of the P_I.[26] However, the starting structures for optimizing P_I and P_{II} in Bour's study were generated from idealized P_I and P_{II} helical structures and therefore failed to consider the influence of the ring puckering. A statistical survey of nonredundant X-ray protein chains from the 2000 version of PDB by Vitagliano et al. observed a correlation between proline puckering and peptide bond conformation.[27] Of the 178 *cis*-proline residues in these structures, 81% adopt a downward puckered conformation. However, for the *trans*-proline residues, both upward and downward pucker conformations are observed with equal frequency. A survey of the HOMSTRAD database for the P_{II} helices revealed that although P_{II} helices only represent 3% of the residues in the database, about 60% of the proteins chains contain one or more P_{II} helices.[28] The P_{II} helices in the data set are defined as a set of sequences that have the characteristic ϕ and ψ torsions.

Here, we present a survey of the Protein Data Bank (PDB) where we found that proline oligomers are only present in all trans conformations for $(Pro)_n$ ($n \geq 5$). The trans conformation again predominates for $n = 4$, except for 3 cases which have a *cis*-proline at the beginning of the sequence. To understand the preference of *trans*-proline in proteins, we undertook calculations of simple, model polyproline oligomers from monomers to hexamers. The conformational analyses of different proline oligomers, based on ring puckering and peptide rotamers, are also discussed.



cis dimer trans dimer

Figure 1. The structures of *cis*- and *trans*-L-proline dimer. Cα through Cδ are labeled, and the ψ, ϕ, ω, and χ_1 torsions are noted.

The distribution and conformational behavior of proline oligomers were examined through a series of ab initio RHF/6-31G* and B3LYP/6-31G* calculations. Proline hexamers were constructed using the structures of proline dimers (Figure 1), which were derived from the conformational scans of dimers and monomers at the RHF/6-31G* level. Figure 1 shows the cis and trans conformers of the L-proline dimer. All conformers in the calculations include the peptide rotamer (cis/trans, as 'c' or 't') and pyrrolidine ring puckering (up/down, as 'U' or 'D') conformation. The calculations on the monomer, dimers, and hexamers of L-proline suggest that the trans conformation has the lowest energy. This finding supports our observation in the PDB that *trans*-proline is the dominant conformation in polyprolines. In addition to revealing the ring puckering effect and the solvation effect on the P_I and P_{II} conformers, we also propose two new regular secondary structures of polyproline, herein identified as polyproline type-III (P_{III}) and polyproline type-IV (P_{IV}).

Methods

Survey of Proline Oligomers in the PDB. The July 2003 release of the PDB was searched for polyproline sequences. The classification of oligomers Pro_n is based on the number of proline residues in a contiguous sequence. The dihedral angles ω, ϕ, ψ, and χ_1 were measured for all entries where $n \geq 4$ (Figure 1). As shown in Figure 1, the ω torsion is defined as the dihedral among [$C\alpha_i$-C_i-$N_{(i+1)}$-$C\alpha_{(i+1)}$]; the ϕ torsion as [C_i-$N_{(i+1)}$-$C\alpha_{(i+1)}$-$C_{(i+1)}$]; the ψ torsion as [$N_{(i)}$-$C\alpha_{(i)}$-$C_{(i)}$-$N_{(i+1)}$]; and the χ_1 torsion as [N_i-$C\alpha_i$-$C\beta_i$-$C\gamma_i$]. The torsion ω defines the proline amide bond as cis ($\omega = 0°$) or trans ($\omega = 180°$, abbreviated herein as 'c' and 't' for cis and trans, respectively). The torsional angles ϕ and ψ determine the secondary structure of polyproline oligomers. The torsion χ_1 describes the pucker conformation of the proline ring, where $\chi_1 > 0$ denotes "endo" or "pucker-down" and $\chi_1 < 0$ is "exo" or "pucker-up" (abbreviated as 'D' and 'U', respectively).

Computational Methods. All ab initio calculations were carried out using Gaussian98 and Gaussian03.[29] In all cases, the default convergence criteria were used. Systematic conformational scans were carried out at the RHF/6-31G* level of theory for the monomers and dimers to characterize the minima to be used for constructing the hexamers. All reported minima along the potential energy surface (PES)

Table 1. Distribution of $(Pro)_n$ $n = 4$ Conformations (50 Entries), Based on the Proline Amide Bond (t/c) and the Ring Puckering (D/U)

conformation	no. of entries	percentage (%)	conformation	no. of entries	percentage (%)
tDtUtUtU	9	15.5	tDtDtUtD	3	5.2
tDtDtDtD	8	13.8	tDtUtDtD	3	5.2
tDtUtDtU	7	12.1	tDtDtUtU	2	3.4
tUtDtUtU	6	10.3	cDtDtUtU	2	3.4
tDtUtUtD	6	10.3	tUtDtUtD	1	1.7
tDtDtDtU	5	8.6	tUtUtDtU	1	1.7
tUtUtUtU	4	6.9	cDtDtUtD	1	1.7

were subject to full geometry optimizations and were further confirmed through frequency calculations, which gave no imaginary frequencies.

Full conformational sampling of eight different dimers (i.e., tDtD, tDtU, tUtD, tUtU, cDcD, cUcU, cDcU, and cUcD) was carried out by scanning the ψ torsion for a full 360° at 10° increments. The optimized minima of the dimers were used to construct the initial conformations of the hexamers; full geometry optimizations of these hexamers were carried out without any constraints at RHF/6-31G*. The resulting RHF/6-31G* minima were used as the starting point for the B3LYP/6-31G* optimization. For the RHF/6-31G* minimizations, the optimized hexamers were verified through frequency analysis and were visualized using XChemEdit.[30] Frequency calculations for the B3LYP/6-31G* minima were not possible because of extensive memory requirements, but the structures were very similar to those obtained with RHF/6-31G*. Solvent effects were included for the hexamers by performing single-point energy calculations using self-consistent reaction field (SCRF) theory with the isodensity surface polarized continuum model (IPCM) at the RHF/6-31G* level.[31] The SCRF-IPCM calculations were carried out for solvent dielectric constants of 4.90, 32.63, and 78.39, corresponding to chloroform, methanol, and water, respectively. A combination of 44 phi points and 22 theta points (parameters for the radial grid employed in IPCM) was adopted for the SCRF-IPCM calculations.

Results and Discussion

Distributions of Polyprolines in the Protein Data Bank. Our survey of the PDB showed that more than 6300 of the 22 119 PDB entries (28.5%) contain at least one proline dimer in their sequence; 475 entries (2.1%) contain at least one proline trimer. Some proteins contain nine consecutive proline residues (e.g., farnesyltransferase[32]) and 15 consecutive prolines (e.g., profilin[7]). Downward ring puckers were seen slightly more often, but both up and down ring puckers occurred in large numbers. This implies that the two are nearly equal in free energy with the downward rings being slightly more favorable. It was interesting that no *cis*-prolines were observed in oligomers with five or more consecutive prolines and only 3 of the 58 tetramers were found to contain a *cis*-proline. These *cis*-prolines only occurred at the beginning of the tetramer motifs (Table 1).

Calculations of Proline Monomers. Systematic searching of the conformational space for the L-proline monomer (Ac-Pro-OMe) yielded 8 minima, two for each of the four combinations (i.e., tD, cD, tU, cU) examined. The characteristic torsion angles (ϕ and ψ) and energies are given in Table 2. The PES for scanning ψ (actually, the ester torsion N-Cα-C-OMe) in all four conformers noted above are provided in the Supporting Information (Figures 1S−2S). The energy differences among the eight minima are less than 3.00 kcal/mol [with the exception of cU(−30)]. The trans conformers are energetically more favorable than the cis monomers. For the same rotamers, the endo ('D') conformation is slightly more favorable in energy than the exo ('U'). The global minimum is tD ($\phi = -70°$, $\psi = 157°$). This finding agrees very well with studies from other groups.[18,25,33] Our calculations on the proline monomer agree well with experimental data showing that the endo conformation is the most populated conformation in the model molecule H_2N-Pro-COOH in the gas phase.[34] Table 2 shows that ϕ torsions for pucker-up conformers are about −60°, while the pucker-down conformers have a ϕ torsion around −70°. The conformers with ester torsions in the range of 150°−170° are lower in energy than those with ester torsions in the range of −30° to −20°. The difference in energy and the conformation can be explained by the 1−4 distance between the two carbonyl carbons. The distance is slightly greater in the 'D' conformation, compared to the 'U' conformation, to effectively attenuate the steric hindrance. This observation was also noted by Vitagliano et al.[27] The larger ψ torsions in the cis conformations result from alleviating the steric interactions between the N-terminal methyl group and the C-terminal methoxyl group. In the dimer and hexamer studies, the minima yielded ester torsions between 150°−170°. Additional minima with different ester torsions were not pursued.

Table 2. Energies and Characteristics of Minima for the Proline Monomer, Ac-Pro-OMe

conformer	characteristics[a]	ϕ (°)	ψ (°)	ΔE (kcal/mol)
1	tD(160)	−70.0	157.0	0.00
2	tD(−20)	−67.4	−23.5	1.21
3	tU(150)	−59.5	148.4	0.97
4	tU(−30)	−55.2	−34.3	1.45
5	cD(170)	−75.4	171.0	1.74
6	cD(−20)	−75.9	−16.5	2.48
7	cU(170)	−62.5	167.7	2.83
8	cU(−30)	−60.2	−34.5	3.20

[a] The characteristics note the peptide rotamer, ring pucker, and ester torsion (ψ for monomers).

Calculations for Proline Dimers. The monomer conformations with the ψ torsion in the range of 150°−170° are lower in energy, so they were used to initiate calculations of the dimers (Ac-Pro-Pro-OMe). The torsional scans for dimers gave two minima in most cases (Figure 2 and Table 3), one with ψ torsions in the range of 130°−170° and the other with range of −50° to 10°. In the most favorable states, the range of ψ torsions for the trans dimers is 130°−150°, while that range for the cis dimers is 160°−170°. Among the eight low-energy conformers, the trans states are more energetically favorable than the cis states. The cis dimers are at least 4 kcal/mol higher in energy than their trans counterparts. This finding is consistent with studies from

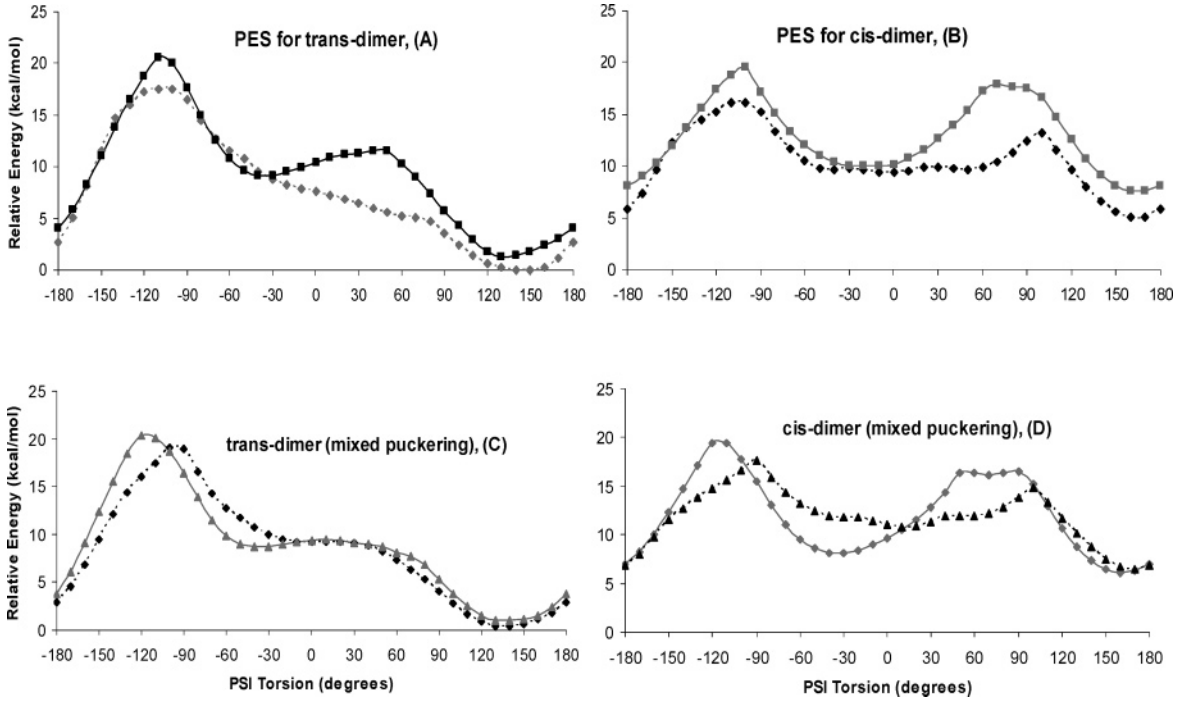


Figure 2. PES for various dimers. All energies are relative to the lowest energy point in the scans, tDtD(148). (A): black square: tUtU; gray diamond: tDtD; (B): gray square: cUcU; black diamond: cDcD; (C): gray triangle: tUtD; black diamond: tDtU; (D): gray diamond: cUcD; black triangle: cDcU.

Table 3. Energies and Characteristics of the Minima for the Proline Dimer (Ac-Pro-Pro-OMe)

confor-mers	charac-teristics[a]	ϕ_1 (°)	ψ (°)	ϕ_2 (°)	ΔE (kcal/mol)	dipole (Debye)
1	tDtD(148)	−70.0	148.2	−74.5	0.00	2.19
2	tDtU(135)	−69.9	135.8	−63.5	0.32	1.29
3	tUtD(135)	−60.8	135.7	−78.8	1.02	1.83
4	tUtU(133)	−60.6	133.4	−64.3	1.30	1.81
5	cDcD(165)	−73.1	165.2	−75.8	4.98	9.11
6	cUcD(161)	−57.2	161.4	−78.3	6.09	9.13
7	cDcU(169)	−73.3	169.0	−65.7	6.43	9.24
8	cUcU(165)	−57.8	165.5	−67.7	7.55	9.35
9	cUcD(−34)	−67.4	−34.2	−83.3	8.10	3.27
10	tUtD(−38)	−66.5	−38.6	−76.5	8.70	6.43
11	tUtU(−35)	−64.9	−35.5	−60.1	9.11	6.71
12	tDtU(4)	−89.8	4.0	−55.8	9.17	6.49
13	cDcD(−3)	−88.6	−3.2	−79.5	9.39	2.47
14	cDcD(−40)	−69.9	−40.5	−82.7	9.64	4.20
15	cDcD(49)	−138.6	49.1	−86.2	9.67	3.82
16	cUcU(−12)	−75.7	−12.8	−59.3	9.96	2.76
17	cDcU(13)	−92.4	13.8	−55.3	10.83	2.51
18	cDcU(56)	−140.4	56.1	−77.3	11.90	4.17
19	cUcD(68)	43.8	68.6	−85.8	16.17	5.32

[a] The characteristics note the peptide rotamer, ring pucker, and ψ torsion. Ester torsions (N_2-Cα_2-C_2-OMe) were between 140° and 175° for all conformers.

other groups.[25] For the same peptide rotamers ('t' or 'c'), pucker-down conformations are slightly lower in energy than the pucker-up conformations. The eight lowest-energy conformers of the dimers are shown in Figure 3, and the higher-energy conformers are shown in Figure 4. The eight low-energy conformers have more favorable steric interactions between the two pyrrolidine rings. The proximity of both termini in the cis dimers (Figure 3) causes them to be higher in energy than the trans dimers. The energetic difference between cis and trans dimers may explain the absence of any *all-cis*-polyprolines ($n \geq 4$) in the current PDB.

For the low-energy trans dimers, the energy difference between ring puckers is small (less than 1.30 kcal/mol), indicating that these conformers would be nearly equally populated at room temperature. Quan and Wu's calculations for two triple helices at HF/6-31G* level suggest that different puckering modes have very similar energies.[35] Because of the closeness in energy between tDtD, tDtU, tUtD, tUtU (Table 3), it is likely that any combination of the above four dimers would give rise to an energetically feasible polyproline oligomer. This hypothesis is consistent with the proline oligomer distribution in the PDB, where the five most common tetramer motifs are comprised of these low-energy dimer moieties (Table 1). For the low-energy dimers with cis rotamers, the downward puckered conformation is energetically more favorable than the upward puckered one. These observations are in good agreement with the survey by Vitagliano et al.[27]

For the proline dimers with ψ torsions in the range of 130°−150°, trans conformers (1−4 in Table 3) have ϕ torsions between −80° and −60°. Therefore, according to the definition of P_I and P_{II} helices, they adopt a P_{II} helix. The cis conformers (5−8 in Table 3) possess a P_I helix, with ϕ torsions in the range of −80° to −55° and ψ torsions of 160°−170°. This finding supports the observations by Zhang and Madelengoitia that dimers can have characteristic values of ϕ and ψ which correspond to P_I and P_{II} helices.[15]

Conformations 9−11 and 14 in Table 3 show that another common minimum conformation includes ψ near −40°, and they are several kcal/mol higher in energy than the lowest global minimum. However, the conformations for tDtD and

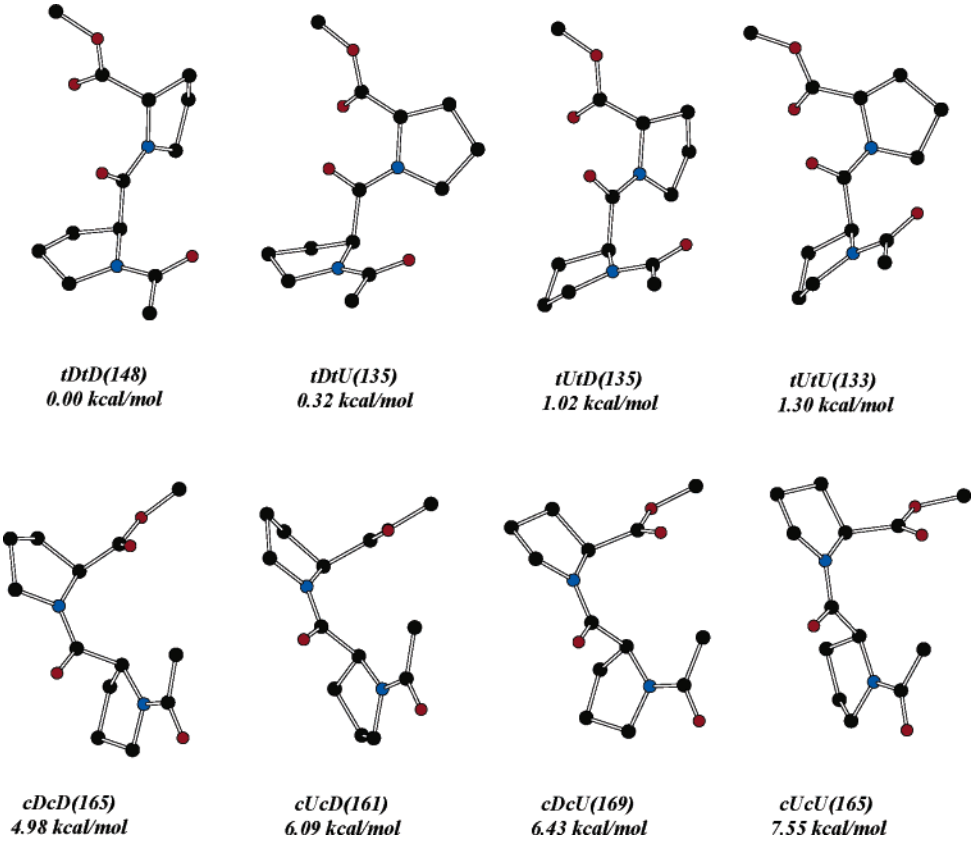


Figure 3. Low-energy minima of dimers illustrating various rotamers and puckering conformations. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

tDtU at $\psi = -40°$ lie along a shoulder on the PES (Figure 2A,C). Full optimization of the tDtU conformer gave a minimum with a ψ torsion of 4.0°. This minimum was verified by the frequency calculations. Optimization of the tDtD conformer failed to yield a minimum outside the range of 130°−170°. A similar situation occurs with cUcU and cUcD, where cUcD gave rise to a minimum with ψ near 70°, while cUcU failed to yield a minimum with ψ near 70° (Figure 2B,D). Interestingly, three energetically close minima for cDcD and cDcU exist (Figure 2D) with ψ torsions in the range of −60° to 80°. Full optimization and frequency calculations reveal that five of these six stationary points yield unique minima. Full optimization of the cDcU conformation ($\psi = -40$) caused an interconversion to the cDcD conformer [designated as cDcD(−40)]. A high-energy conformation like cDcU(−40) most likely has a very small barrier to rearrangement to the lower energy cDcD(−40), and it appears that it is not a stable minimum. Ring inversions such as this one are not uncommon for prolines. Badoni et al.[36] reported that the barrier for ring inversion from the 'U' conformation to the 'D' is 2.1 kcal/mol for *N*-formyl-*trans*-proline amide (For-Pro-NH_2, a proline monomer) based on B3LYP/6-31G* calculations. Kang and Park[37] also reported an estimated energy barrier of 2.2 kcal/mol for ring inversion from the 'D' conformation to the 'U' conformation for *N*-acetyl-L-proline-*N*′ and *N*′-dimethylamide (Ac-Pro-NMe2) at the B3LYP/6-311++G** levels.

The high energy of trans dimers with ψ torsions near −40° and 0° (Figure 4) arises from strong repulsion between the atoms of the N-terminal pyrrolidine ring and the methylene group at the δ-position of the second proline residue. For the cis dimers with similar ψ torsions, the high energy is due to the steric overlap between the first pyrrolidine ring and the C-terminal ester group and/or the steric interactions from both termini.

One new and interesting observation from the torsional scan of the dimers is that the shape of the potential energy surface seems to be determined by the peptide bond conformation and ring puckering of the first proline (Figure 2). The PES of tDtU is more similar to tDtD than tUtU. Similarly, tUtD resembles tUtU, cDcU resembles cDcD, and cUcD resembles cUcU.

Calculations for Proline Hexamers. The energetic penalty of pushing any dimer conformation away from its most favorable ψ torsion of ∼160° is less than 10 kcal/mol (Figure 2) and comparable to the results of Mattice et al.[38] Therefore, we chose to include these higher-energy conformations in our analyses of polyproline hexamers. The proline hexamers were constructed using characteristics of the optimized dimers. We did not pursue conformations of the hexamer based on conformers 15, 18, or 19 in Table 3; propagating helixes with these ϕ and ψ combinations resulted in oligomers that collided back upon themselves in an unrealistic fashion. It should also be noted that helices based on the dimer conformers 13 and 14 in Table 3 minimized to the same conformation (conformation 14 in Table 4). This resulted in a total of 15 conformations for the hexamers.

Each appropriate conformer of the hexamer was fully optimized using the RHF/6-31G* levels of theory, and the resulting minima were further optimized using B3LYP/6-

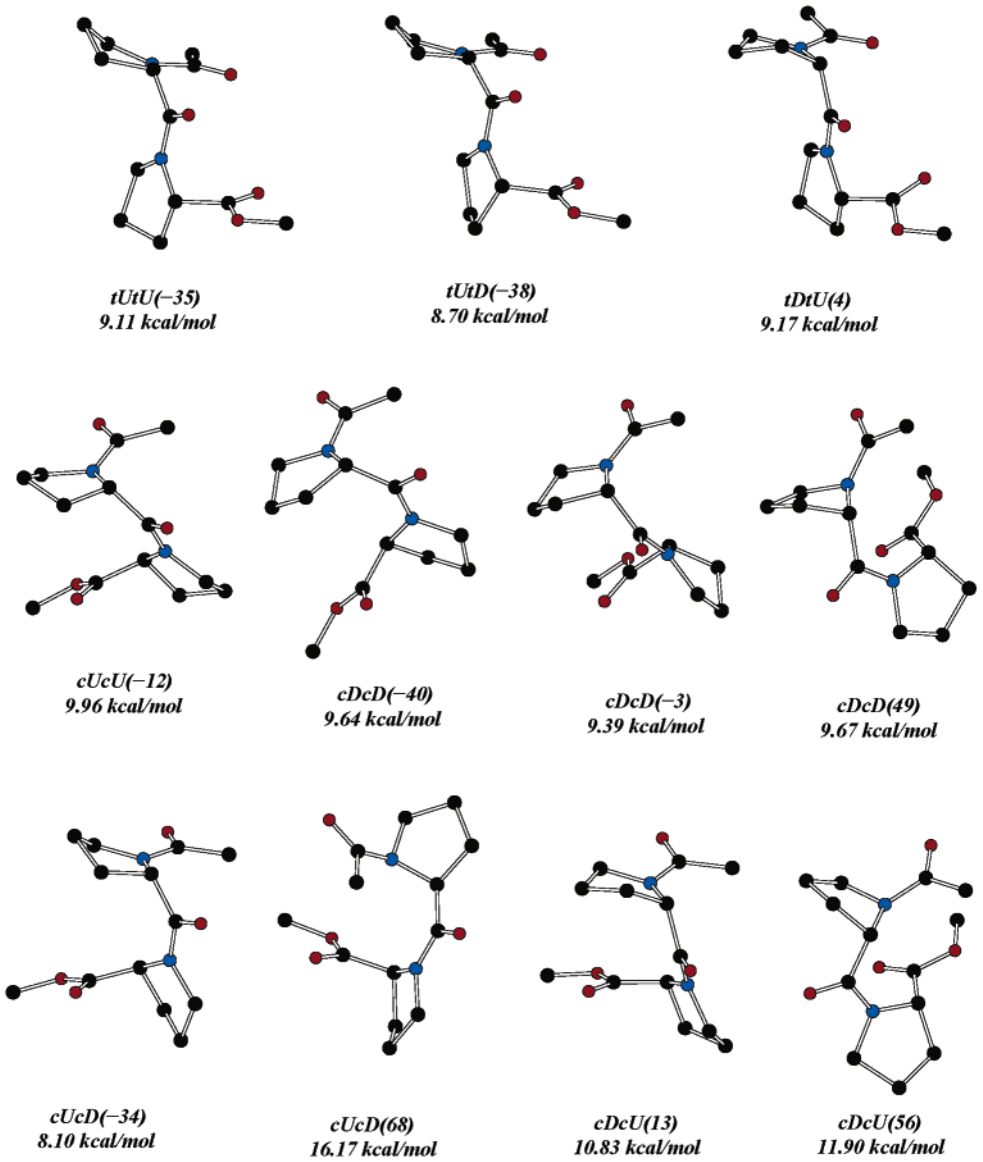


Figure 4. High-energy minima of dimers illustrating various rotamers and puckering conformations. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

31G* levels. The energy differences derived from the B3LYP method were smaller than those from the RHF method. Proline hexamers composed of all *trans*-proline units (conformers 1−4 in Table 4) were found to have lower energies in the gas phase than those containing all *cis*-prolines (conformers 5−8 in Table 4). These observations are consistent with the lack of *all-cis*-proline oligomers in the PDB. The trans conformation, tDtD-hex-(148), forms an ideal left-handed 3_1 P_{II} helix (Figure 5). Conformers 2−4 in Table 4 (tDtU, tUtD, and tUtU hexamers) show similar left-handed P_{II} helices, with ϕ torsions in the range of −75° to −65° and ψ torsions in the range of 125°−152°. These variations from the ideal structure are energetically accessible and are seen in the PDB structures. The cis, endo hexamer, cDcD-hex-(165) shows an ideal, right-handed 10_3 P_I helix (Figure 6). The low-energy structures of the cDcU, cUcD, and cUcU hexamers (conformers 6−8 in Table 4) show similar right-handed P_I helices. Conformational studies of homologous β-proline oligomers have been reported, and the handedness of the β-proline oligomers is the reverse of these α-polyprolines; *trans*-α-polyprolines adopt left-handed P_{II} helices, but the *all-trans*-β-proline oligomers yield right-handed ones.[39]

The relative free energy, ΔG, for each minimum at the RHF/6-31G* level was estimated based on the calculated frequencies of the normal modes (the frequencies were scaled back by a factor of 0.89).[40] The trends in the free energy mirror those of ΔE. The ΔG values show that trans conformers are the most favorable and should be the most populated.

Here, we compare tDtD-hex-(148), our representative structure of a P_{II} helix to P_{II} structures determined by NMR (pdb code: 1JVR[41]) and X-ray crystallography (pdb file: 1F34[42]). The average ϕ and ψ torsions in the minimized trans, endo hexamer are −72.0° and 152.9°, respectively. This is in good agreement with the ϕ and ψ values from the NMR structure and the X-ray structures (Table 5). The comparison of calculated and experimental geometrical parameters for the trans, exo hexamer tUtU-hex-(133) are listed in Table

Table 4. Energies (in kcal/mol), Free Energies (in kcal/mol), and Conformational Characteristics of the Proline Hexamer Minima

conformers	characteristics[a]	ϕ (°)[b]	ψ (°)[b]	ΔE (RHF)	ΔG (RHF)[c]	ΔE (B3LYP)	handedness
1	tDtD-hex-(148)	−72.0	152.9	0	0	2.46	left
2	tDtU-hex-(135)	−74.8	125.0	2.33	1.79	0.00	left
3	tUtD-hex-(135)	−71.7	125.7	3.06	2.02	1.01	left
4	tUtU-hex-(133)	−65.7	126.1	4.31	4.98	1.00	left
5	cDcD-hex-(165)	−75.8	163.4	10.31	12.38	7.33	right
6	cUcD-hex-(161)	−68.6	163.1	14.06	15.85	10.06	right
7	cDcU-hex-(169)	−71.3	164.0	14.18	15.95	10.28	right
8	cUcU-hex-(165)	−65.8	164.5	17.88	18.96	13.00	right
9	tUtU-hex-(−35)	−66.1	−36.5	27.00	30.25	21.00	right
10	tUtD-hex-(−38)	−71.8	−26.1	29.13	31.19	23.62	right
11	tDtU-hex-(4)	−71.4	−21.6	29.47	31.64	19.68	right
12	cUcU-hex-(−12)	−70.3	−37.7	48.78	52.93	36.08	left
13	cUcD-hex-(−34)	−77.7	−30.6	49.75	53.87	37.70	left
14	cDcD-hex-(−3)	−91.0	−21.0	51.52	55.69	39.67	left
15	cDcU-hex-(13)	−80.1	−23.4	51.75	56.01	39.40	left

[a] The characteristics note the peptide rotamer, ring pucker, and ψ torsion of the minimized dimers used to build hexamers. [b] The average ϕ and ψ of the hexamers minimized with B3LYP/6-31G*. [c] The relative free energy, ΔG, was derived from the frequency calculations at the RHF/6-31G* level. The frequencies were scaled by 0.890, and the calculation was determined for $T = 298.15$ K.

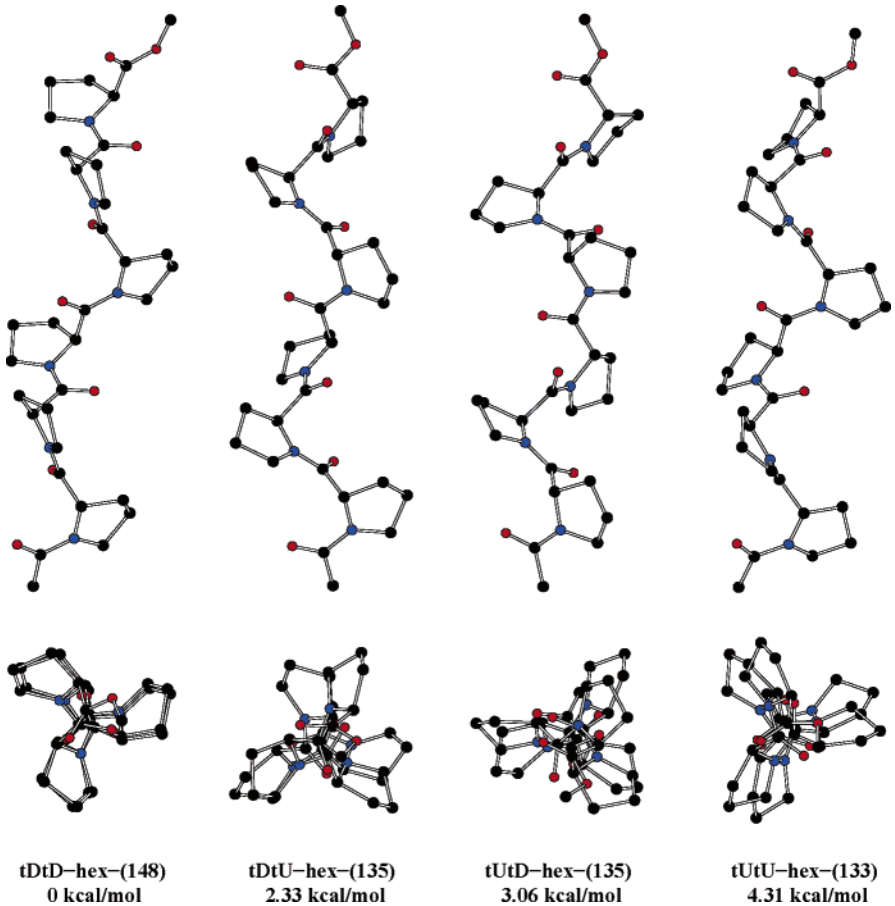


Figure 5. Side views and axial views for the low-energy, trans hexamers determined at the RHF/6-31G* level of theory. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

1S. Again, the calculated average ϕ torsions are very close to those determined from the X-ray structure (pdb file: 1CF0[43]). The choice of tDtD-hex-148 and tUtU-hex-(133) was primarily based on the available NMR and/or X-ray structures with the same puckering pattern, i.e., trans, endo (tDtD-hex) and trans, exo (tUtU-hex).

Numerous experiments have shown that in polar solvents such as water, aliphatic acids, or benzyl alcohol, the cis P_I will rearrange to create the trans P_{II} form.[44,45] In contrast, the trans P_{II} will isomerize to P_I in organic solvents such as propanol or butanol. To probe this, solvation effects were estimated for the different hexamer conformations in three

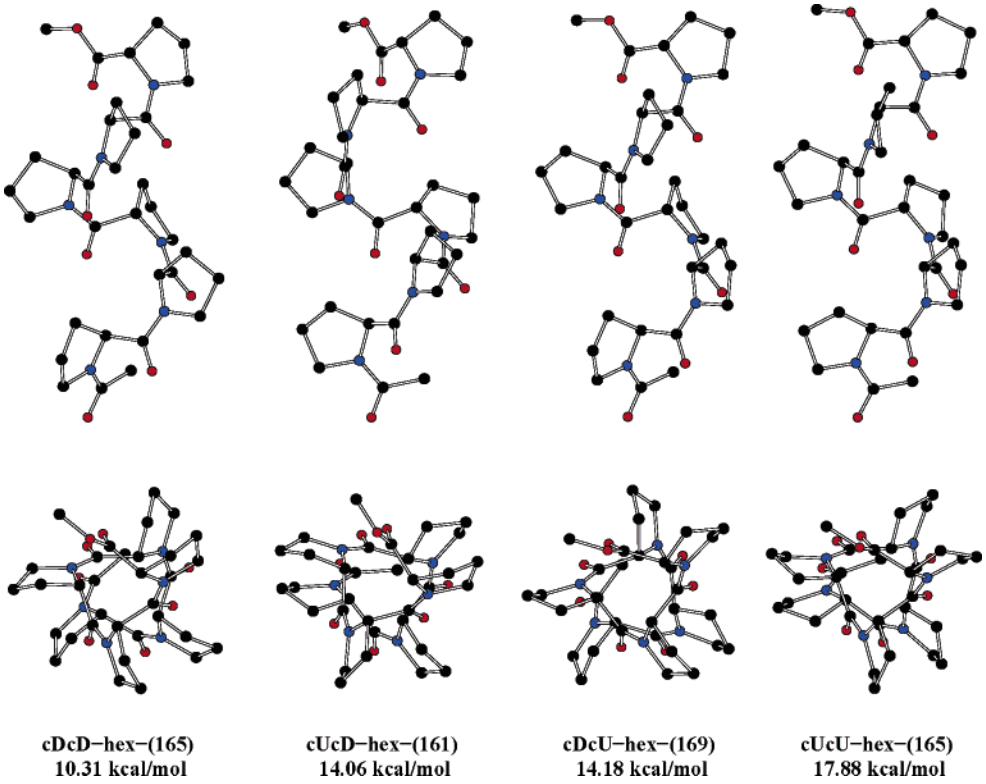


Figure 6. Side views and axial views for the low-energy, cis hexamers determined at the RHF/6-31G* level of theory. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

Table 5. Comparison of Calculated Main Chain Torsion Angles (°) from a 3_1-Helical tDtD-hex-(148) with NMR and X-ray Data

	B3LYP/6-31G*			NMR (1JVR)			X-ray (1F34)		
res	ω	ϕ	ψ	ω	ϕ	ψ	ω	ϕ	ψ
1	176.7	−71.9	147.8	180.0	−75.0	159.3	179.6	−54.3	137.8
2	171.1	−73.4	153.9	180.0	−75.0	164.2	179.9	−64.4	159.4
3	171.1	−72.5	154.1	179.9	−75.0	155.0	179.7	−81.8	148.8
4	170.3	−70.5	153.7	180.0	−75.0	159.3	179.9	−61.1	147.4
5	172.1	−71.6	149.8	180.0	−75.0	171.4	179.8	−50.9	147.1
6	173.5	−72.1	158.1	180.0	−75.1	170.0			
7				179.9	−75.1	56.5			
av	172.5	−72.0	152.9	180.0	−75.0	161.8	179.8	−62.5	148.1

different environments (Table 6). The effects were calculated using the SCRF-IPCM method with increasing dielectrics for three solvents: chloroform, methanol, and water. A uniform decrease in ΔE between trans and cis low-energy conformers is observed as the solvent polarity increases. These trends can be explained by the dielectric better complementing the larger dipole moments for the cis low-energy hexamers. However, the IPCM calculations indicate that for lower-energy hexamers, the cis forms are better solvated in water and methanol than the corresponding trans forms (Table 6). This result is somewhat surprising because *trans*-proline oligomers are overwhelmingly preferred in water according to numerous experiments. However, some research groups point out that (1) the P_{II} content is solvent dependent and that the population of the P_{II} decreases in such order, water > methanol > ethanol > 2-propanol,[46] and (2) the stability of the P_{II} structure of proline oligomers is chain-length dependent. Proline oligopeptides composed of 13 Pro residues are quite stable in water, while proline hexamers and tetramers show decreasing stability due to molecular thermal fluctuation.[47]

Of course, protic solvents such as water and alcohols do not behave like simple dielectrics. There are factors other than the dipole moments that can contribute to the stability of the polyproline helices in the condensed phase. The ICPM calculations cannot account for the hydrogen bonding between the solvents and the polyproline helices. These effects strongly influence the stability of the helices. The most important point to the ICPM calculations is the fact that environmental effects overcome the large energy difference between P_I and P_{II} helices in the gas phase, making both forms stable in the condensed phase. The condensed phase also lowers the energy of conformers 9−11 which is relevant to our discussions of high-energy helices below.

The calculated IR spectrum (Table 2S in Supporting Information) from the frequency calculations show that all

Table 6. Dipole Moment (μ) and SCRF Energies (kcal/mol) of the Hexamer Minima at the RHF/6-31G* Level in the Gas Phase, $CHCl_3$, MeOH, and H_2O

conformer	characteristics[a]	μ (Debye)	ΔE (gas)	ΔE ($CHCl_3$)	ΔE (MeOH)	ΔE (H_2O)
1	tDtD-hex-(148)	4.79	0.00	0.00	0.00	0.00
2	tDtU-hex-(135)	3.04	2.33	2.70	3.28	3.25
3	tUtD-hex-(135)	3.62	3.06	2.23	2.93	2.94
4	tUtU-hex-(133)	5.62	4.31	4.97	5.58	5.49
5	cDcD-hex-(165)	25.59	10.31	2.03	−0.80	−1.32
6	cUcD-hex-(161)	26.20	14.06	5.67	3.69	3.31
7	cDcU-hex-(169)	26.31	14.18	5.73	3.17	2.86
8	cUcU-hex-(165)	26.93	17.88	8.52	5.08	4.53
9	tUtU-hex-(−35)	21.54	27.00	19.38	17.06	16.58
10	tUtD-hex-(−38)	19.32	29.13	24.49	22.98	22.68
11	tDtU-hex-(4)	19.93	29.47	22.71	20.24	19.83
12	cUcU-hex-(−12)	8.49	48.78	46.35	45.71	45.52
13	cUcD-hex-(−34)	7.86	49.75	44.93	44.04	43.48
14	cDcD-hex-(−3)	5.79	51.52	46.44	44.64	44.29
15	cDcU-hex-(13)	5.75	51.75	46.48	42.54	42.14

[a] The characteristics note the peptide rotamer, ring pucker, and ψ_1 torsion of the minimized dimers, the parents to build hexamers. The numbers in brackets denote the conformer numbers in the table.

calculated hexamers contain absorption bands at 1756 cm^{-1} which is characteristic of the ester carbonyl C═O stretch and between 1680 and 1700 cm^{-1} characteristic of the amide carbonyl stretch.[48] A strong band at 1421 cm^{-1} was also observed in all hexamers and has been reported previously for both P_I and P_{II} types of polyprolines.[49] Each of the low-energy, cis hexamers display a characteristic P_I absorption band at 960 cm^{-1} and have no P_{II} specific bands at between 670 and 400 cm^{-1}.[50] This finding suggests that the most favorable cis hexamers have the properties of P_I. Low-energy, trans hexamers have no band at 960 cm^{-1} (except tDtD-hex-148) but have bands at 400 and/or 670−675 cm^{-1}, indicating the characteristics of P_{II} helices in the most favorable trans hexamers. The higher-energy conformations from both the cis and trans hexamers have characteristic bands at 400 and 960 cm^{-1}. The calculated spectra suggest that these conformations may have some characteristics of both P_I and P_{II}.

The high-energy, trans hexamers are right-handed helices with four residues per turn (4_1), adopting a "square helix" form with a proline ring at each corner (Figure 7). We refer to this novel secondary structure for *trans*-proline oligomers as a polyproline type-III conformation (P_{III}). The square, P_{III} helix is more compact than both the P_{II} (3_1 helix) and the classic α-helix (3.6 residues per turn). P_{III} has ϕ torsions near −70° and ψ torsions of approximately −35°. This combinations of ϕ and ψ angles lie in the allowed α-helix region ($\phi \sim -57°$ and $\psi \sim -47°$), in contrast to the ϕ and ψ angles of P_{II} which are located in the β-sheet region. The conformational state for proline with ψ near −50° has been shown to be stable both experimentally[51] and computationally.[38]

It is noteworthy that trans oligomers with ψ torsions between 133° and 155° adopt a left-handed P_{II} helix, but

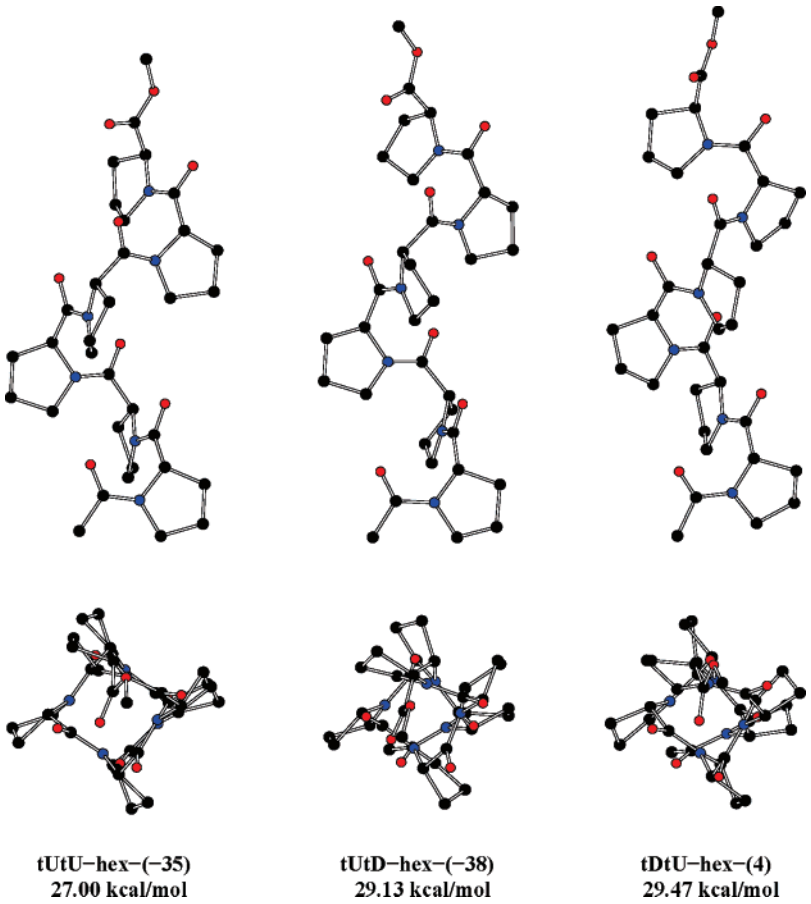


Figure 7. Side views and axial views for the higher-energy, trans hexamers determined at the RHF/6-31G* level of theory. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

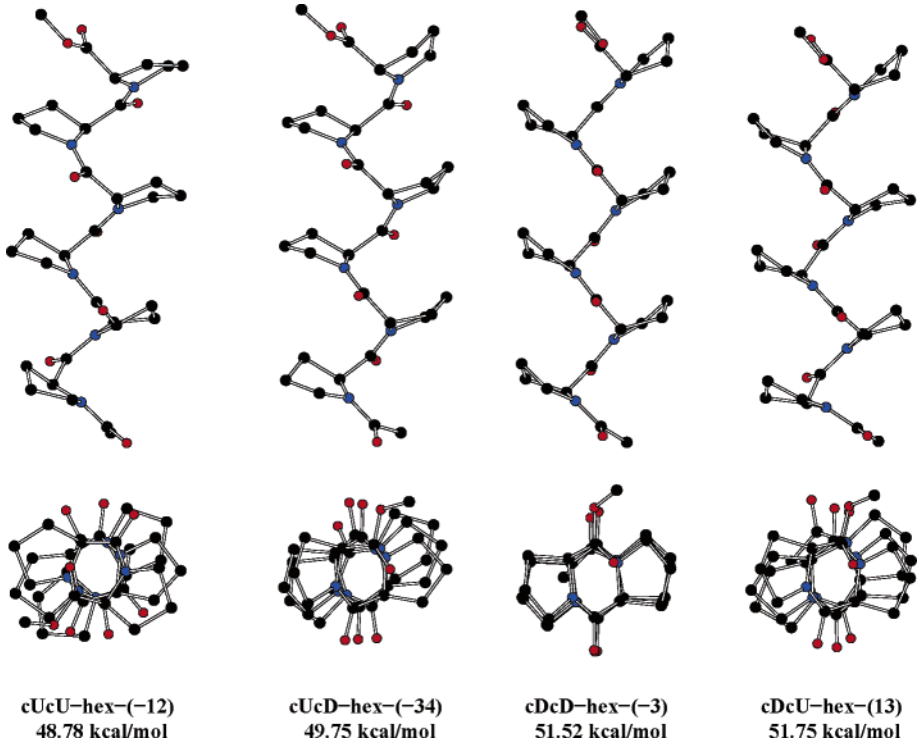


Figure 8. Side views and axial views for the higher-energy, cis hexamers determined at the RHF/6-31G* level of theory. Hydrogen atoms are not shown for clarity. Color codes for atoms: black: C; red: O; blue: N.

trans oligomers with ψ torsions near −35° give rise to the right-handed P_{III} structure. As suggested by the calculated IR spectra, the P_{III} form does indeed share characteristics of both the P_I and P_{II} forms: it has *trans*-amide rotamers similar to P_{II} and forms a right-handed helix like P_I. The handedness of the polyprolines depends not only on the peptide rotamers (cis or trans) but also on the values of the ψ torsions. We propose that the high-energy P_{III} form could exist as conformational intermediates between P_I and P_{II}. As a polyproline strand converts from the P_I to P_{II} form, it has to flip the amide rotamers and change handedness of the helix. Whether the amides flip first or the helix changes handedness first, the polyproline will have to at least partially adopt a trans, right-handed form (or a less likely cis left-handed form, see below). The ICPM calculations show that the condensed phase lowers the energies of these less favorable, trans states, making them energetically accessible and more likely than the high-energy, cis forms.

The UV-Raman spectroscopy of a 21-residue, Ala-based peptide shows a conversion from an α-helix to a P_{II} conformation.[52] Under compressive strain the polyalanine α-helix (3.66 residues per turn) was reported to transform to a π-helix (4.5 residues per turn) which is more compact than α-helix.[53] Our calculated square, P_{III} helix is very similar to the reported π-helix of polyalanine. Whether a right-handed, trans P_{III} conformation plays a role during the meltdown of the right-handed α-helix to the left-handed P_{II} conformation remains unknown. By providing the characteristics of the P_{III} conformation through calculations, it may be possible to design experiments to observe that form.

The high-energy, cis hexamers adopt conformations similar to β-strands with two residues per turn (Figure 8). They tend to adopt a slight left-handed twist if they deviate away from the ideal conformation seen for cDcD-hex-(−3). We refer to this novel secondary structure as polyproline type IV (P_{IV}). The P_{IV} sheets have ϕ torsions in the range of −90° to −70° and ψ torsions in the range of −40° to −20°, lying in the allowed α-helix region. This is different from the low-energy, cis form of P_I that has ϕ and ψ torsions in the β-sheet region of the Ramachandran plot. P_{IV} structures have relatively low dipole moments which show little stabilization in the condensed phase (Table 6). It is unlikely that this form could be observed experimentally, but it may be useful in understanding the conformational behavior of peptides or in the design of biomaterials with unique properties.

Conclusions

Our calculations provide a basis for understanding the conformational behavior of polyproline and provide an explanation for the proline oligomer distribution in the current PDB. Our calculations show that in the gas phase, *trans*-proline P_{II} helices are energetically more favorable than *cis*-polyprolines. In the condensed phases, the P_I and P_{II} forms become much closer in energy. The energy difference in ring puckering is small but slightly biased toward down-puckering. Both states would be highly populated.

To our knowledge, this is the first report of novel secondary structures for polyproline, the P_{III} and P_{IV} forms. P_{III} forms a square, right-handed helix, and P_{IV} is a β-sheet form. This is also the first report of the interconversion between left- and right-handed forms due solely to changes in the ψ torsion. Frequency calculations on the P_{III} and P_{IV} forms show that they possess the IR bands characteristic of both P_I and P_{II}. It is quite possible that the P_{III} form is an

intermediate state in the mutarotation of polyproline from P_{II} to P_I helices. Although P_{III} and P_{IV} would be less populated because of their high energy, their existence might aid in our understanding of the conformational behavior of polyproline in protein folding and provide some insight for better understanding the interconversion between P_{II} and P_I helices.

Acknowledgment. The authors are indebted to Prof. William L. Jorgensen and Dr. D. C. Lim for their generous donation of the XChemEdit program used to make many figures and to visually analyze the normal modes of optimized hexamers. The authors also thank Dr. Eugene Stewart for his critical review of this paper. This work has been supported by the National Institutes of Health (GM 65372).

Supporting Information Available: Comparison of calculated geometry parameters from a 3_1-helical tUtU-hex-(133) with X-ray data, the calculated IR frequencies for hexamers, the absolute energies (in hartrees) for proline monomers, dimers, and hexamers, the PES maps for proline monomers, and the Cartesian coordinates for all reported minima. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Zarrinpar, A.; Bhattacharyya, R. P.; Lim, W. A. *Sci. STKE* **2003**, *179*, re8., 1−10.

(2) Wiesner, S.; Stier, G.; Sattler, M.; Macias, M. J. *J. Mol. Biol.* **2002**, *324*, 807−822.

(3) Gertler, F. B.; Niebuhr, K.; Reinhard, M.; Wehland, J.; Soriano, P. *Cell* **1996**, *87*, 227−239.

(4) Freund, C.; Kühne, R.; Yang, H.; Park, S.; Reinherz, E. L.; Wagner, G. *Embo. J.* **2002**, *21*, 5985−5995.

(5) Gu, W.; Kofler, M.; Antes, I.; Freund, C.; Helms, V. *Biochemistry* **2005**, *44*, 6404−6415.

(6) Mahoney, N. M.; Janmey, P. A.; Almo, S. C. *Nat. Struct. Biol.* **1997**, *4*, 953−960.

(7) Mahoney, N. M.; Rozwarski, D. A.; Fedorov, E.; Fedorov, A. A.; Almo, S. C. *Nat. Struct. Biol.* **1999**, *6*, 666−671.

(8) Ramakrishnan, V.; Ranbhor, R.; Durani, S. *J. Am. Chem. Soc.* **2004**, *126*, 16332−16333.

(9) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry* **2005**, *44*, 6269−6275.

(10) Hamburger, J. B.; Ferreon, J. C.; Whitten, S. T.; Hilser, V. J. *Biochemistry* **2004**, *43*, 9790−9799.

(11) Cowan, P. M.; McGavin, S. *Nature* **1955**, *176*, 501−503.

(12) Straub, W.; Shmueli, U. *Nature* **1963**, *198*, 1165−1166.

(13) Chao, Y.-Y. H.; Bersohn, R. *Biopolymers* **1978**, *17*, 2761−2767.

(14) Deber, C. M.; Bovey, F. A.; Carver, J. P.; Blout, E. R. *J. Am. Chem. Soc.* **1970**, *92*, 6191−6198.

(15) Zhang, R.; Madalengoitia, J. S. *Tetrahedron. Lett.* **1996**, *37*, 6235−6238.

(16) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2004**, *108*, 4885−4898.

(17) Jhon, J. S.; Kang, Y. K. *J. Phys. Chem. A* **1999**, *103*, 5436−5439.

(18) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341−350.

(19) Fischer, S.; Dunbrack, R. L., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931−11937.

(20) Kang, Y. K. *J. Mol. Struct. (THEOCHEM)* **2004**, *675*, 37−45.

(21) Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233−249.

(22) Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135−140.

(23) Czinki, E.; Császár, A. G. *Chem. Eur. J.* **2003**, *9*, 1008−1019.

(24) Ramek, M.; Kelterer, A.-M.; Nikolić, S. *Int. J. Quantum Chem.* **1997**, *65*, 1033−1045.

(25) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 698−705.

(26) Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2002**, *65*, 45−59.

(27) Vitagliano, L.; Berisio, R. A.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2001**, *10*, 2627−2632.

(28) Cubellis, M. V.; Caillez, F.; Blundell, T. L.; Lovell, S. C. *Proteins: Struct., Funct., Bioinformatics* **2005**, *58*, 880−892.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, revision A.7*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(30) Lim, D. C.; Jorgensen, W. L. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons Ltd.: Athens, GA, 1998; Vol. 5, p 3295.

(31) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(32) Long, S. B.; Hancock, P. J.; Kral, A. M.; Hellinga, H. W.; Beese, L. S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12948−12953.

(33) Kang, Y. K. *J. Phys. Chem. B* **2002**, *106*, 2074−2082.

(34) Lesarri, A.; Mata, S.; Cocinero, E. J.; Blanco, S.; Lüpez, J. C.; Alonso, J. L. *Angew. Chem., Int. Ed.* **2002**, *41* (24), 4673−4676.

(35) Quan, J. M.; Wu, Y. D. J. *Theor. Comput. Chem.* **2004**, *3*(2), 225−243.

(36) Badoni, H. A.; Rodriguez, A. M.; Zamora, M. A.; Zamarbide, G. N.; Enriz, R. D.; Farkas, Ö.; Császár, P.; Torday, L. L.; Sosa, C. P.; Jàkli, I.; Perzel, A.; Papp, J. G.; Hollosi, M.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **1999**, *465*, 79−91.

(37) Kang, Y. K.; Park, H. S. *J. Mol. Struct. (THEOCHEM)* **2005**, *718*, 17−21.

(38) Mattice, W. L.; Nishikawa, K.; Ooi, T. *Macromolecules*. **1973**, *6*, 443−446.

(39) Sandvoss, L. M.; Carlson, H. A. *J. Am. Chem. Soc.* **2003**, *125*, 15855−15862.

(40) http://www.nist.gov/compchem/irikura/prog/thermo_cgi.html.

(41) Christensen, A. M.; Massiah, M. A.; Turner, B. G.; Sundquist, W. I.; Summers, M. F. *J. Mol. Biol.* **1996**, *264*, 1117−1131.

(42) Ng, K. K.; Petersen, J. F. W.; Cherney, M. M.; Garen, C.; Zalatoris, J. J.; Rao-Naik, C.; Dunn, B. M.; Martzen, M. R.; Peanasky, R. J.; James, M. N. G. *Nat. Struct. Biol.* **2000**, *7*, 653−657.

(43) Mahoney, N. M.; Rozwarski, D. A.; Fedorov, E.; Fedorov, A. A.; Almo, S. C. *Nat. Struct. Biol*. **1999**, *6*, 666−671.

(44) Steinberg, I. Z.; Berger, A.; Katchalski, E. *Biochim. Biophys. Acta* **1958**, *28*, 647.

(45) Lin, L.-N.; Brandts, J. F. *Biochemistry* **1980**, *19*, 3055−3059.

(46) Liu, Z.; Chen, K.; Ng, A.; Shi, Z.; Woody, R. W.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2004**, *126*, 15141−15150.

(47) Kakinoki, S.; Hirano, Y.; Oka, M. *Polym. Bull.* **2005**, *53*, 109−115.

(48) Isemura, T.; Okabayashi, H.; Sakakibara, S. *Biopolymers* **1968**, *6*, 307−321.

(49) Johnston, N.; Krimm, S. *Biopolymers* **1971**, *10*, 2597−2605.

(50) Rabolt, J. F.; Wedding, W.; Johnson, K. W. *Biopolymers* **1975**, *14*, 1615−1622.

(51) Clark, D. S.; Dechter, J. J.; Mandelkern, L. *Macromolecules* **1979**, *12*, 626−633.

(52) Asher, S. A.; Mikhonin, A. V.; Bykov, S. *J. Am. Chem. Soc.* **2004**, *126*, 8433−8440.

(53) Ireta, J.; Neugebauer, J.; Scheffler, M.; Rojo, A.; Galván, M. *J. Am. Chem. Soc.* **2005**, *127*, 17241−17244.

CT050182T