# Pyrrolidine Puckering and Imide *cis-trans* Isomerization of Ac-X-Pro-NHMe Dipeptides

**Seong Jun Han†, Jong Suk Jhon, and Young Kee Kang***

*Department of Chemistry, Chungbuk National University Cheongju, Chungbuk 361-763, Korea*
*†C&C Research Labs., 146-141 Annyung-ri, Taean-up, Hwasung-goon, Kyunggi-do 445-970, Korea*
*Received May 19, 1997*

Because of the conformational restriction in the pyrrolidine ring, proline (Pro) has no amide hydrogen for use as a donor in hydrogen bonding, and the configuration of atoms of the proline residue adjacent to the α-carbon of the preceding residue is quite similar in both conformations with *cis*- and *trans*-imide X-Pro bonds (Figure 1).[1] This allows the energy difference between the *cis* and *trans* conformers to be less for imide bonds than that for amide bonds. Therefore, the Pro residue has a relatively high intrinsic probability of having the cis peptide bond preceding the proline as compared with other amino acids.[2,3] The *cis* peptide bonds are found mainly in β-bends and a specific structural role for *cis* imide bonds has been suggested.[2,3] The pyrrolidine of the Pro residue is a five-membered ring and its puckering has often been described using the concept of the pseudorotation.[4-8] The ring may adopt two distinct puckered conformations that are almost equally favorable.[4,5] They distinguished by the displacement of the $C_\alpha$ and $C_\beta$ atoms from the mean plane of the ring and are referred to as 'down' and 'up', respectively (Figure 2).[9]

Although many experimental and theoretical studies have been made on either the puckering of prolyl rings[4-6,10-12] or the *cis-trans* isomerization of imide bonds[2,3,13-18] of proline-containing peptides, only a few studies have been established for the ring puckering of *cis*- and *trans*-proline residue by the analysis of X-ray crystal structures of peptides and proteins,[19] by conformational energy calculations on proline-containing peptides,[20-22] and by *ab initio* molecular orbital calculations on the proline dipeptide.[7] In particular, from $^1$H NMR studies on linear and cyclic oligopeptides, it has been known that the *cis : trans* ratio can be influenced by the nature of the residues preceding and following the proline.[23,24] As a further step in understanding the relationship between puckerings and imide bonds of the Pro residue depending on the nature of the residues preceding the proline, we herein report preliminary results of a conformational study on Ac-X-Pro-NHMe dipeptides.

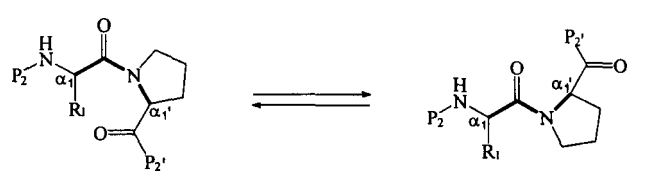


**Figure 1.** Configurations of *cis* and *trans* imide bonds of the proline residue.

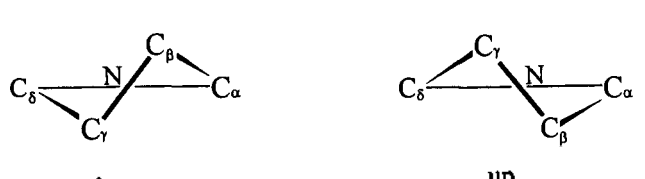


**Figure 2.** Down and up puckerings of the pyrrolidine ring.

*To whom correspondence should be addressed.

## Computational Methods

The nomenclature and conventions used follow the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature.[25] The calculations were carried out on terminally blocked dipeptides Ac-X-Pro-NHMe, where X includes five types of residues, *i.e.*, aliphatic (Ala and Val), nonpolar (Gly and Pro), aromatic (Phe and Tyr), polar (Asn and Ser), and charged (Lys and Asp) side chains. Conformational energy calculations were carried out using the ECEPP/3 (Empirical Conformational Energy Program for Peptides, version 3),[26] in which the total conformational energy is the sum of the electrostatic energy, the nonbonded energy, and the torsional energy. The hydrogen bond energy is included in the nonbonded energy component. In addition, the internal conformational energy of each pyrrolidine ring, depending on the puckering and *cis-trans* imide bond, was added to the total energy.[26] A quasi-Newton algorithm SUMSL (Secant-type Unconstrained Minimization problem SoLver)[27] was used for energy minimization. All torsion angles of the peptide backbone, side chains, and end groups of each dipeptide were allowed to move during minimization. In particular, the torsion angle $\phi$ of the proline was fixed at $-68.8°$ and $-53.0°$ for the *down* and *up* puckerings, respectively, due to the ring constraint.[26] Each conformation of the dipeptides was denoted in terms of a conformational letter code of Zimmerman *et al.*,[28] which was assigned to each residue specifying its location on a $\phi$-$\psi$ map.

All minima of the Ac-X-NHMe with the relative energy $\triangle E \leq 5$ kcal/mol reported by Vásquez *et al.*[29] and all minimum-energy conformations of Ac-Pro-NHMe[26] were combined to generate starting conformations for minimization. Because of large number of low energy conformations for terminally blocked Asn and Lys residues, their minima with $\triangle E \leq 3$ kcal/mol were included only for starting points. Because the relative energy of the conformation tCu, *i.e.*, the *trans*-up puckered conformation with a letter code C, is found to be not too high (1.93 kcal/mol) and the con-

formation DtCu of the Ac-Ala-Pro-NHMe has relatively a low conformational energy of 1.88 kcal/mol,[26] the conformation tCu of Ac-Pro-NHMe is additionally included in generating starting conformations. Lysine and Aspartic acid were taken to be in their uncharged forms. All non-Gly residues were taken in the L-configuration. The normalized statistical weight of each conformation was calculated according to the Boltzmann equation suggested by Zimmerman and Scheraga[30] at 298 K. The populations α% and β% for the α-helical and β-sheet structures were computed by adding statistical weights of the conformations denoted by the letter code A and the letter codes C and F, respectively.

## Results and Discussion

Table 1 lists a summary of the numbers of starting and minimized conformations for terminally blocked single residues and dipeptides. The decrease in the number ($n$) of low energy conformations for each X-Pro dipeptide compared to the number ($m$) obtained by combining low energy conformations of Ac-X-NHMe and Ac-Pro-NHMe indicates that conformational spaces of both X and Pro residues are restricted by interresidue interactions, as discussed later in detail.

In addition, the calculated statistical weight for each conformational letter code, defined in the previous section, of X and Pro residues is included in Table 1. The calculated results show that (1) the conformations C and A, which are preferentially probable for Ac-X-NHMe and of which the former has a $C_7$ intramolecular hydrogen bond,[29] become less feasible for dipeptides except for the Gly dipeptide, (2) the relative stability of the conformations E and F is significantly increased, and (3) the conformation F of the Pro-Pro dipeptide is dominantly stable. These conformational features can be understood by the steric conflicts between the $C_\delta H_2$ group attached to the imide nitrogen and the NH and $C_\beta H_2$ groups of the X residue, as described in Figure 5 of ref 3. The dominant stability of the conformation A for the Gly residue of the Gly-Pro dipeptide was also predicted by Schimmel and Flory from conformational energy calculations.[31] In particular, the unstability of the conformation A for X residues is accord with a well-known fact that the proline oftens plays a role as an α-helix breaker in peptides and proteins, which is mainly ascribed to the lack of the H atom used for hydrogen bonding.[1] Nevertheless, Pro residues are frequently seen in *N*-termini of α-helices and in the middle of α-helices by distorting the helical conformation locally.[19] On the other hand, the statistical weight of Pro is almost constant for both Ac-Pro-NHMe and dipeptides, which indicates that the conformation of the Pro residue is not significantly affected by the nature of its preceding residues. Therefore, the conformation of the X resi-

**Table 1.** Conformational Properties of Ac-X-Pro-NHMe[a]

| X | number of conformations[b] | | | | | X[c] | | | | | | | | | | Pro[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $m$ | $n$ | <1 | <3 | <5 | A | C | D | E | F | A* | D* | E* | F* | H* | A | C | F |
| Ala | 70 | 58 | 9 | 27 | 35 | | | 0.495 | 0.292 | 0.182 | 0.030 | | | | | 0.492 | 0.374 | 0.134 |
| | | | | | | (0.135) | (0.507) | (0.082) | (0.159) | (0.079) | (0.010) | | | | | (0.533) | (0.366) | (0.102) |
| Val | 90 | 79 | 6 | 29 | 50 | | 0.173 | 0.342 | 0.419 | 0.011 | 0.055 | | | | | 0.558 | 0.319 | 0.123 |
| | | | | | | (0.285) | (0.356) | (0.034) | (0.301) | (0.021) | (0.003) | | | | | | | |
| Gly | 70 | 61 | 12 | 37 | 55 | 0.369 | 0.001 | 0.131 | 0.103 | 0.001 | 0.073 | 0.165 | 0.096 | 0.059 | 0.001 | 0.607 | 0.294 | 0.099 |
| | | | | | | (0.051) | (0.390) | (0.035) | (0.049) | | (0.051) | | | | | | | |
| Pro | 100 | 80 | 4 | 18 | 31 | | 0.001 | | | 0.999 | | | | | | 0.398 | 0.457 | 0.145 |
| | | | | | | (0.533) | (0.366) | | | (0.102) | | | | | | | | |
| Phe | 180 | 143 | 3 | 19 | 58 | | 0.013 | 0.034 | 0.769 | 0.184 | 0.001 | | | | | 0.108 | 0.864 | 0.029 |
| | | | | | | (0.087) | (0.077) | (0.059) | (0.682) | (0.072) | (0.002) | | | | | | | |
| Tyr | 310 | 249 | 6 | 33 | 92 | | 0.004 | 0.040 | 0.827 | 0.129 | 0.001 | | | | | 0.104 | 0.863 | 0.033 |
| | | | | | | (0.067) | (0.064) | (0.079) | (0.732) | (0.039) | | | | | | | | |
| Asn | 390 | 266 | 8 | 41 | 123 | | 0.053 | 0.036 | 0.634 | 0.276 | 0.001 | | | | | 0.619 | 0.211 | 0.169 |
| | | | | | | (0.196) | (0.287) | (0.032) | (0.385) | (0.076) | (0.006) | | | | | | | |
| Ser | 460 | 399 | 4 | 84 | 191 | | 0.003 | 0.182 | 0.550 | 0.244 | 0.021 | | | | | 0.641 | 0.251 | 0.108 |
| | | | | | | (0.236) | (0.400) | (0.023) | (0.201) | (0.052) | | | | | | | | |
| Lys | 1780 | 1562 | 39 | 379 | 852 | 0.001 | 0.013 | 0.264 | 0.524 | 0.149 | 0.049 | | | | | 0.554 | 0.353 | 0.093 |
| | | | | | | (0.170) | (0.559) | (0.034) | (0.159) | (0.044) | (0.012) | | | | | | | |
| Asp | 430 | 362 | 20 | 58 | 146 | | 0.044 | 0.009 | 0.526 | 0.420 | 0.001 | | | | | 0.655 | 0.294 | 0.051 |
| | | | | | | (0.289) | (0.223) | (0.043) | (0.283) | (0.140) | (0.006) | | | | | | | |

[a] The values in parenthesis correspond to those of Ac-X-NHMe taken from ref 29, except for Pro, Tyr, and Ser taken from ref 26. [b] $m$ and $n$ correspond to the number of starting and optimized conformations. The quantities in the third to fifth columns are the number of optimized conformations with relative energies less than 1, 3, and 5 kcal/mol, respectively. All low energy conformations of single residues with $\Delta E_{tot}$ 5 kcal/mol were combined to generate $m$ starting conformations, while for Asn and Lys residues low energy conformations with $\Delta E_{tot}$ 3 kcal/mol were used. The conformation tCu of Ac-Pro-NHMe was also included in generating starting conformations. See the text in detail. [c] Statistical weights of X and Pro. Letter codes were defined according to the work of Zimmerman *et al.* (ref 28). The populations for the conformation G of Ac-X-NHMe are 0.028, 0.000, 0.000, 0.000, 0.021, 0.018, 0.018, 0.078, 0.022, and 0.015 for X residues listed in the first column from top to bottom.

**Table 2.** Statistical Populations of Ac-X-Pro-NHMe (%)

| | X | *trans* | | *cis* | | total | PDB[a] *cis*% | |
|---|---|---|---|---|---|---|---|---|
| | | down | up | down | up | *cis*% | *b* | *c* |
| aliphatic | Ala | 78.7 | 17.2 | 3.9 | 0.2 | 4.1 | 3.3 | 2.5 |
| | Val | 61.8 | 36.7 | 1.4 | 0.2 | 1.6 | 2.6 | 2.7 |
| nonpolar | Gly | 60.9 | 34.9 | 3.6 | 0.6 | 4.2 | 7.8 | 8.8 |
| | Pro | 82.9 | 10.5 | 6.5 | 0.1 | 6.6 | 7.6 | 10.8 |
| aromatic | Phe | 96.0 | 3.0 | 1.0 | 0.1 | 1.1 | 9.6 | 6.4 |
| | Tyr | 95.8 | 3.2 | 0.9 | 0.1 | 1.0 | 25.0 | 19.1 |
| polar | Asn | 78.0 | 21.2 | 0.7 | 0.1 | 0.8 | 5.1 | 5.3 |
| | Ser | 76.3 | 21.0 | 2.4 | 0.3 | 2.7 | 11.0 | 10.3 |
| charged | Lys | 75.8 | 21.7 | 2.3 | 0.1 | 2.4 | 5.2 | 5.8 |
| | Asp | 81.4 | 16.7 | 1.8 | 0.1 | 1.9 | 3.6 | 1.8 |
| total av. | | 78.8 | 18.6 | 2.5 | 0.2 | 2.7 | 8.1 | 7.4 |
| PDB analysis[d] | | 40.8 | 50.0 | 8.2 | 1.0 | 9.2 | | |

[a] The Brookhaven Protein Data Bank. [b] Taken from ref 2. [c] Taken from ref 3. [d] Taken from ref 19.

due in X-Pro dipeptides appears to depend highly on its following Pro residue by interresidue interactions between the pyrrolidine ring of the proline and the side chain of the X residue.

The statistical populations of Ac-X-Pro-NHMe with *cis*- and *trans*-imide bonds depending on its down and up pyrrolidine puckerings are shown in Table 2. The calculated overall preference is *trans*-down > *trans*-up > *cis*-down > *cis*-up, which is similar to that of ECEPP/3[26] and *ab initio* HF/6-31G**[7] calculations on Ac-Pro-NHMe. The higher population of *trans*-down conformations seems to be ascribed to the significant contribution of the conformation tCd, which is the lowest energy conformation of Ac-Pro-NHMe.[26] However, the relative abundance of *trans*-down conformations appears to be relatively overestimated, compared to that obtained from the analysis of X-ray structures of proteins, since the average abundance ratio of down:up is calculated to be 79 : 19, whereas the corresponding ratio for proteins is 41 : 50.[19] This overestimation can be attributed to the contribution of low energy conformations tAd and tFd, whose conformational energies are relatively low.[26] However, our recent *ab initio* HF/6-31G** calculations on Ac-Pro-NHMe indicate that these conformations are not local minima at high levels of basis sets.[7] In the ECEPP/3, the internal conformational energies of proline for the *trans*-down and *trans*-up conformations are assigned to be 0.473 and 0.950 kcal/mol higher than the *cis*-down conformation.[26] Hence, our calculated results show that the more stability of *trans*-down and *trans*-up conformations is attributed to interresidue interactions between X and Pro residues.

The average *cis* population of Ac-X-Pro-NHMe is calculated to be about 3%, which is somewhat lower than the values obtained from the analysis of X-ray structures of proteins.[2,3,19] The lowest *cis* population is found for aromatic Phe and Tyr residues, whereas their abundances appear to be highest from X-ray structures of proteins. This disagreement may be interpreted as the weak aromatic-proline interactions for isolated dipeptides, because these interactions are known to be effective in proteins from X-ray studies[2,3] and in a pentapeptide $^{+}H_2$-Ser-Tyr-Pro-Tyr-Asp-O$^{-}$ from NMR experiments.[32] In addition, the overestimation of *trans*-down conformations seems to be a crucial factor to de-

**Table 3.** Conformation of Proline in Ac-X-*trans*-Pro-NHMe (%)[a]

| | X | α% | | | | β% | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | down | up | total | PDB[b] | down | up | total | PDB[b] |
| aliphatic | Ala | 33.6 | 14.2 | 47.8 | 29 | 48.5 | 3.7 | 52.2 | 71 |
| | Val | 23.6 | 31.7 | 55.3 | 22 | 39.1 | 5.5 | 44.6 | 78 |
| nonpolar | Gly | 26.9 | 33.2 | 60.1 | 55 | 36.7 | 3.3 | 40.0 | 45 |
| | Pro | 27.6 | 8.0 | 35.6 | 52 | 61.1 | 3.2 | 64.3 | 48 |
| aromatic | Phe | 8.5 | 1.5 | 10.0 | 51 | 88.6 | 1.4 | 90.0 | 49 |
| | Tyr | 8.1 | 1.6 | 9.7 | 38 | 88.6 | 1.6 | 90.2 | 62 |
| polar | Asn | 43.8 | 18.1 | 61.9 | 69 | 34.8 | 3.3 | 38.1 | 31 |
| | Ser | 44.7 | 19.1 | 63.8 | 52 | 33.7 | 2.5 | 36.2 | 48 |
| charged | Lys | 35.5 | 19.1 | 54.6 | 45 | 42.2 | 3.1 | 45.3 | 55 |
| | Asp | 49.2 | 15.8 | 65.0 | 89 | 33.7 | 1.2 | 34.9 | 11 |
| total av. | | 30.2 | 16.2 | 46.4 | 50 | 50.7 | 2.9 | 53.6 | 50 |
| PDB analysis[c] | | 19 | 27 | 46 | | 26 | 28 | 54 | |

[a] α% and β% were computed by adding statistical weights of conformations, of which the former corresponds to the letter code A and the latter to the letter codes C and F shown in Table 1. Statistical weights were normalized only for *trans* conformations. [b] Taken from ref 3. [c] Taken from ref 19.

crease the *cis* populations, as discussed above.

In Table 3, the populations for secondary structures of α-helix and β-sheet for the proline in Ac-X-*trans*-Pro-NHMe are listed. The analysis of X-ray structures of proteins shows that proline in proteins adopts two distinct conformations, *i.e.*, α-helical and β-sheet structures.[3] The average calculated α% and β% populations are 46% and 54%, respectively, which are in good agreement with the values obtained from X-ray structures of proteins[3,19] in spite of some overestimation for down-puckered conformations, as discussed above. In particular, the higher abundance of β-sheets relative to α-helices for aromatic Phe and Tyr dipeptides is promising, which may indicate that the bulkier side chains of X residues enforce the conformation of Pro residue to be more extended.

From the analysis of the results obtained here, it can be concluded that the conformation of X residues is strongly affected by the following proline and the conformational space of the proline is restricted by the bulkier side chains of X residues. The effects of conformational entropy and hydration on the conformation of X and Pro residues are now being carried out. In addition, the study for the effect of residues following the proline on the conformation of X-Pro sequences is now in progress.

**Acknowledgments.** This work was supported by research grants from the Korea Science and Engineering Foundation (No. 961-0303-027-2). S.J.H. also appreciates the support from the Korea Research Foundation (1995).

## References

1. Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; Freeman: New York, 1993.
2. Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253.
3. MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397.
4. Madison, V. *Biopolymers* **1977**, *16*, 2671.
5. DeTar, D. F.; Luthra, N. P. *J. Am. Chem. Soc.* **1977**, *99*, 1232.
6. de Leeuw, F. A. A. M.; van Kampen, P. N.; Altona, C.; Díez, E.; Esteban, A. L. *J. Mol. Struct.* **1984**, *125*, 67.
7. Kang, Y. K. *J. Phys. Chem.* **1996**, *100*, 11589.
8. Han, S. J.; Kang, Y. K. *J. Mol. Struct. (Theochem)* **1996**, *369*, 157.
9. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
10. London, R. E. *J. Am. Chem. Soc.* **1978**, *100*, 2678.
11. de Leeuw, F. A. A. M.; Altona, C.; Kessler, H.; Bermel, W.; Friedrich, A.; Krack, G.; Hull, W. E. *J. Am. Chem. Soc.* **1983**, *105*, 2237.
12. Sarkar, S. K.; Young, P. E.; Torchia, D. A. *J. Am. Chem. Soc.* **1986**, *108*, 6459.
13. Han, S. J.; Kang, Y. K. *Int. J. Pept. Protein Res.* **1993**, *42*, 518.
14. Schmid, F. X.; Mayr, L. M.; Mücke, M.; Schönbrunner, E. R. *Adv. Protein Chem.* **1993**, *44*, 25.
15. Oka, M.; Montelione, G. T.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 7959.
16. Shin, Y. A.; Han, S. J.; Kang, Y. K. *J. Phys. Chem.* **1993**, *97*, 9248.
17. Choi, S. H.; Yu, J. Y.; Shin, J. K.; Jhon, M. S. *J. Mol. Struct.* **1994**, *323*, 233.
18. Fischer, S.; Dunbrack, R. L., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931.
19. Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. *J. Mol. Biol.* **1992**, *228*, 725.
20. Tanaka, S.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 698.
21. Thomasson, K. A.; Applequist, J. *Biopolymers* **1990**, *30*, 437.
22. Schmidt, J. M.; Brüschweiler, R.; Ernst, R. R.; Dunbrack, R. L., Jr.; Joseph, D.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 8747.
23. Grathwhol, C.; Wüthrich, K. *Biopolymers* **1981**, *20*, 2623
24. Yao, J.; Feher, V. A.; Espejo, B. F.; Reymond, M. T.; Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1994**, *243*, 736.
25. IUPAC-IUB Commission on Biochemical Nomenclature. *Biochemistry* **1970**, *9*, 3471.
26. Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472.
27. Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503.
28. Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
29. Vásques, M.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.
30. Zimmerman, S. S.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 811.
31. Schimmel, P. R.; Flory, P. J. *J. Mol. Biol.* **1968**, *34*, 105.
32. Yao, J.; Brüschweiler, R.; Dyson, H. J.; Wright, P. E. *J. Am. Chem. Soc.* **1994**, *116*, 12051.