Extensive exploration of the conformational landscapes of neutral and terminally blocked prolines in the gas phase: A density functional theory study

Yizhou Yang and Zijing Lin

Abstract
Proline is an important amino acid that plays unique roles in the structures of peptides and proteins. The conformations of proline are searched by a thorough method, generating 3888 trial structures optimized at the B97D/6-311++G** level. A total of 23 conformations are found and their structural and energetic data are presented. All the proline conformers exhibit a coplanar feature for four of the five pyrrolidine ring atoms. The coplanar rule reduces the cost of the conformational search by a factor of 40. The theoretical composition-weighted infrared spectrum provides a good explanation of the experimental results. A conformational search of capped proline yields seven unique conformers, all with trans C-termini peptide planes. The trans C-termini rule further cuts by half the cost of the conformational search of proline-containing peptides. The theoretical composition of the cis N-termini peptide bonds at room temperature is 5.5%, agreeing with the experimental estimations of 3%–10%.

Keywords
cis peptide plane, computational search methods, conformational analysis, pyrrolidine, vibrational spectra

Date received: 1 April 2022; accepted: 14 June 2022

Corresponding author:
Zijing Lin, Hefei National Research Center for Physical Sciences at Microscale & CAS Key Laboratory of Strongly-Coupled Quantum Matter Physics, Department of Physics, University of Science and Technology of China, Hefei, P.R. China.
Email: zjlin@ustc.edu.cn
Introduction

Among the 20 proteogenic natural amino acids (AAs), proline (P) is unique in forming the secondary structure of peptides and proteins due to its important pyrrolidine ring. In addition to being required for protein biosynthesis, proline plays critical roles in cellular bioenergetics,1–5 osmoregulation,6,6 stress protection,7–10 and cellular signaling processes.11,12 Studies over the last two decades have also established the specialized role of proline in cancer metabolism.13,14–16 Recent discoveries of the broad effects of proline metabolism on cancer cell growth and survival have implicated proline metabolic enzymes as potential targets for therapeutic intervention.17–21

Proline is also involved in many antimicrobial peptides (AMPs). AMPs are small peptides with a broad spectrum of antibiotic activities against bacteria, fungi, protozoa, and viruses, and sometimes exhibit cytotoxic activity toward cancer cells. AMPs containing predominantly either tryptophan or proline can kill microorganisms by targeting intracellular pathways and are, therefore, a promising source of next-generation antibiotics. Proline-rich AMPs interact with the 70S ribosome and disrupt protein synthesis. They can also target the heat shock protein in target pathogens, and consequently lead to protein misfolding.22

The special biological functions of proline are intimately related to the conformations it may adopt. Because of the existence of a pyrrolidine in its ring side chain, proline is less flexible than other AAs. Comparing with other AAs, an amidogen H atom is replaced by a C atom in proline, and it is difficult for proline to form a hydrogen bond (H-bond) with the C–O of the fourth residue at the C-terminal or the nitrogen on an adjacent peptide chain. As a result, proline tends to destroy the α-helix or the β-fold and ends up at the end of a protein secondary structure.

Besides, due to the absence of the N–H bond and the associated spatial repulsion of amide H, the cis conformation may appear in a peptide containing proline, while it is basically forbidden in other peptide bonds. The cis populations of proline in water and chloroform have been established by the experiments of Delaney and Madison.23 They found about 25% cis peptide bond isomers in water and 15%–20% cis in chloroform. Park and Kang24 found that their density functional theory (DFT) results were consistent with experiments. For gas phase proline, the fraction of cis X–P peptide bonds is typically in the range of 3%–10%25,26 (X stands for any AA), and may be up to 40% for some aromatic AA–P peptide bonds.27

Considering the fundamental importance of proline as an elemental AA and its unique features in peptide and protein structures, a detailed knowledge about the conformations of natural and capped proline is desirable. Capped proline refers to a proline capped by a peptide bond at both its N- and C-termini and mimics proline in a peptide or protein better than in natural proline. As illustrated by a number of studies,28–30 the structural properties of natural and capped proline are substantially different. A conformational study of capped proline is essential for understanding the behavior of proline in peptides and proteins. As experimental approaches such as infrared (IR) spectroscopy, mass spectrometry ultraviolet (UV) spectroscopy, and electric dipole moment measurements generally provide only indirect structural information, computational explorations of proline conformations are of special importance.

Sapse et al.31 carried out the earliest ab initio study of proline conformations at the RHF/STO-3G and RHF/6-31G levels. Ramek et al.32 conducted a search on the potential energy surface (PES) at the RHF/6-311++G** level and found 10 proline conformers. They also determined the reaction paths and potential energy barriers connecting the PES stationary points. Császár and Perczel33 obtained 12 proline conformers and their relative energies at correlated levels of electronic structure theory. Meanwhile, Stepanian et al.34 identified 15 stable minima on the PES of proline and reassigned the vibrational spectra of unionized proline obtained from matrix-isolation IR spectroscopy. Lesarri et al.35 configured a special laser-ablation system to obtain the rotational parameters of two plausible gas phase proline conformers and calculated their structures by an iterative-fit procedure. Czinki and Császár36 found 18 proline conformers at the DFT(B3LYP)/6-311++G** level of theory. The search of proline conformations by Czinki and Császár is the most thorough so far. However, their search did not consider the variations of the dihedral angles (DAs) for the main chain Ca–C and C–OH bonds, and some energy minima may be easily missed in their search. Recently, Ten and Shcherbakova37 performed B3LYP/6-311++G (d, p) calculations of the IR spectra of three proline conformers and attempted to attribute the experimental spectrum to two of the three proline conformers. However, the three conformers do not include the global minimum of the conformation ensemble of Czinki and Császár.36 Not surprisingly, the comparison between the computational result of Ten and Shcherbakova37 and the experimental spectrum is of low quality.

Computational studies of the conformations of capped proline are less common than that of natural proline. Yuan et al.38 provided a comprehensive conformational analysis of the 20 capped AAs. In this outstanding work, they identified five unique conformations for capped proline. Unfortunately, no connection was made between the computed conformations and the unique structural characteristics of proline in peptides.

In this study, the PESs of natural and capped prolines are thoroughly explored and the structures of all energy minima are analyzed. The equilibrium distributions of the conformational ensemble at various temperatures are presented. In addition, the theoretical and experimental IR spectra of natural proline are compared. The ensemble of computed conformations of capped proline is used to explain quantitatively the compositions of the cis–trans structures of proline in peptides.

Computational methods

To examine the characteristics of proline conformers comprehensively, the conformational space of natural proline is explored through a systematic variation of all its seven bond rotational degrees of freedom, as indicated in Figure 1(a). Details about the bond rotations and the values of the
DAs used are summarized in Table 1. By considering all possible combinations of the DAs, a total number of $4 \times 2 \times 3 \times 3 \times 3 \times 3 \times 2 = 3888$ trial structures are generated and used for exploring the PES of proline. The trial structures are optimized at the B97D/6-311++G** level of theory, with the default convergence criteria of Gaussian09, for example, maximum force $< 0.00045$, root mean square (RMS) of force $< 0.0003$, maximum displacement $< 0.0018$, RMS of displacement $< 0.0012$. The B97D/6-311++G** level of theory is chosen as it has been shown to provide the required computational accuracy for AAs and peptides.\textsuperscript{39,40} The set of unique conformations thus obtained are verified to be true energy minima by frequency calculations as well as further geometry optimizations at the B97D/6-311++G** level of theory, with the default convergence criteria of Gaussian09, for example, maximum force $< 0.00045$, root mean square (RMS) of force $< 0.0003$, maximum displacement $< 0.0018$, RMS of displacement $< 0.0012$. The B97D/6-311++G** level of theory is chosen as it has been shown to provide the required computational accuracy for AAs and peptides.\textsuperscript{39,40} The set of unique conformations thus obtained are verified to be true energy minima by frequency calculations as well as further geometry optimizations at the B97D/6-311++G** level. Both the electronic energies and the free energy corrections at the B97D/6-311++G** level are used to determine the equilibrium conformation distributions at various temperatures.\textsuperscript{41} The harmonic approximation of molecular vibrations is used, that is, the conformer structures and their vibrational modes are temperature independent.

The vibrational frequencies of the proline conformations are calculated at the B97D/6-311++G** level. Based on the scaling factors suggested for B3LYP/aug-cc-pvdz\textsuperscript{37} and through benchmarking tests, the B97D/6-311++G** frequencies are scaled with a scaling factor of 0.97 for the OH, NH, and CH stretching modes, and with a scaling factor of 1.01 for all other vibrations. The infrared spectrum for the conformational ensemble is calculated from the equilibrium compositions and smoothed with a 5-point Savitzky–Golay smoothing function.\textsuperscript{42}

The conformational search of capped proline is conducted in a way much more efficient than that for proline. Based on analysis and testing of the obtained proline conformers, only two of the five pyrrolidine ring bond rotations are necessary, as indicated in Figure 1(b). Besides, the bond rotations $b(C3–C4–O5–H)$ and $h(H–N8–C3–C4)$ in proline are now revised, respectively, to $b(C3–C4–N5–C12)$ and $h(C11–C9–N8–C3)$ due to the changed C- and N-termini in the capped proline. The number of trial structures required for the PES exploration of capped proline is thus reduced to 96, and can be further reduced to 48, as discussed in the “Results and discussion” section. The trial structures of capped proline are also optimized at the B97D/6-311++G** level.

All the trial structures used in this study are generated with our in-house developed software written in Python consisting of 11 subroutines and 1576 lines of computing code. All the DFT calculations are carried out using the Gaussian-09 software package.\textsuperscript{43}

### Results and discussion

#### Conformations of proline

A total of 23 conformers are found through the PES exploration. All the conformers are unique as assessed from their energy–dipole moment combinations and verified by final visual inspection. The relative electronic energies ($\Delta E$), zero-point vibrational energies ($\Delta E_{0}$), Gibbs free energies ($\Delta G_{0}$) at room temperature ($T_{0}$), rotational constants and permanent dipole moments of all 23 proline conformers are shown in Table 2. As seen in Table 2, all the 23 conformers

---

**Table 1.** Bond rotations and the dihedral angles used in the generation of the trial structures for the conformational search of proline.

| Bond rotation | Dihedral angle |
|---------------|---------------|
| $a$ (N8–C3–C4–O5) | $-60^\circ, 0^\circ, 60^\circ, 180^\circ$ |
| $b$ (C3–C4–O5–H) | $0^\circ, 180^\circ$ |
| $c$ (C7–N8–C3–C4) | $-90^\circ, 90^\circ, 180^\circ$ |
| $d$ (C1–C2–C3–C4) | $-60^\circ, 60^\circ, 180^\circ$ |
| $e$ (C7–C1–C2–C3) | $-30^\circ, 30^\circ, 0^\circ$ |
| $f$ (N8–C7–C1–C2) | $-30^\circ, 30^\circ, 0^\circ$ |
| $g$ (C3–N8–C7–C1) | $-30^\circ, 30^\circ, 0^\circ$ |
| $h$ (H–N8–C3–C4) | $-120^\circ, 120^\circ$ |

---

**Figure 1.** Planar structure and atom numbering for: (a) proline (P), with all the rotational degrees of freedom indicated (see Table 1 for more details); (b) capped P, with the bond rotations necessary for the conformational search indicated (see the main text for more details).
Table 2. Relative electronic energies (ΔE, kcal mol⁻¹), relative ZVPEs (ΔVE₀, kcal mol⁻¹), relative Gibbs free energies at room temperature (ΔG₀, kcal mol⁻¹), rotational constants (GHz) and permanent dipole moments (D, Debye) of the proline conformers determined at the B97D/6-311+G** level of theory.

| Conformer | ΔE | ΔVE₀ | ΔG₀ | Rotational constants | D  |
|-----------|----|------|-----|---------------------|----|
|           |    |      |     | x       | y       | z       |    |
| 1         | 0.00 | 0.00 | 0.00 | 3.65   | 1.66   | 1.39   | 5.77 |
| 2         | 0.69 | 0.07 | 0.50 | 3.87   | 1.59   | 1.26   | 5.85 |
| 3         | 2.49 | −0.16| 1.43 | 3.87   | 1.57   | 1.29   | 1.50 |
| 4         | 2.57 | −0.14| 1.72 | 3.80   | 1.57   | 1.37   | 1.78 |
| 5         | 2.61 | −0.12| 1.49 | 3.98   | 1.49   | 1.31   | 1.94 |
| 6         | 3.87 | −0.08| 2.95 | 3.44   | 1.73   | 1.40   | 1.35 |
| 7         | 3.98 | −0.17| 3.02 | 3.46   | 1.71   | 1.48   | 1.79 |
| 8         | 4.00 | −0.17| 3.12 | 3.92   | 1.54   | 1.33   | 2.08 |
| 9         | 4.33 | −0.17| 2.87 | 4.19   | 1.38   | 1.28   | 2.64 |
| 10        | 4.39 | 0.01 | 3.97 | 4.01   | 1.45   | 1.38   | 2.20 |
| 11        | 4.49 | −0.36| 3.49 | 3.89   | 1.54   | 1.31   | 2.24 |
| 12        | 4.71 | −0.02| 4.26 | 3.99   | 1.46   | 1.39   | 2.12 |
| 13        | 4.86 | −0.02| 4.54 | 3.61   | 1.61   | 1.51   | 2.51 |
| 14        | 4.97 | −0.42| 3.89 | 4.27   | 1.44   | 1.15   | 2.19 |
| 15        | 5.12 | −0.30| 4.09 | 3.96   | 1.49   | 1.39   | 2.44 |
| 16        | 5.36 | 0.02 | 5.01 | 4.17   | 1.48   | 1.15   | 4.63 |
| 17        | 6.58 | −0.31| 4.82 | 3.84   | 1.57   | 1.28   | 3.69 |
| 18        | 6.64 | −0.32| 5.87 | 3.75   | 1.58   | 1.37   | 3.69 |
| 19        | 7.77 | −0.44| 7.14 | 3.97   | 1.51   | 1.32   | 4.58 |
| 20        | 8.43 | −0.17| 7.41 | 3.87   | 1.51   | 1.42   | 3.11 |
| 21        | 9.21 | −0.56| 8.10 | 4.24   | 1.45   | 1.14   | 4.86 |
| 22        | 9.87 | −0.34| 8.83 | 3.81   | 1.54   | 1.36   | 5.26 |
| 23        | 9.92 | −0.16| 9.49 | 3.57   | 1.63   | 1.50   | 5.06 |

The pyrrolidine ring structure and an efficient method of searching proline conformations

Examining the structures of all proline conformers, as partly shown in Figure 2, a universal feature is observed.
That is, four of the five pyrrolidine ring atoms are very close to existing in a common plane. Denoting the five pyrrolidine ring atoms as A, B, C, D, and E, where A, B, C, and D lay approximately in a plane and E is the non-coplanar atom. The deviation of the four atoms from forming an ideal plane may be characterized with \( \gamma = \text{abs}(DA(A–B–C–D)) \), referred to below as the non-planeness. The procedure of finding A, B, C, D, E, and \( \gamma \) is straightforward. The five DAs concerning the five pyrrolidine ring atoms, \( DA(C1–C2–C3–N8) \), \( DA(C2–C3–N8–C7) \), \( DA(C3–N8–C7–C1) \), \( DA(N8–C7–C1–C2) \), and \( DA(C7–C1–C2–C3) \), were computed and compared. The four atoms yielding the smallest absolute DA value are said to be coplanar atoms, and the corresponding DA is referred to as the non-planeness, \( \gamma \). For example, \( \gamma = \text{abs}(DA(C2–C3–N8–C7)) \) if it is the smallest among the five DAs. The four atoms, C2, C3, N8, C7, are coplanar, and are referred to as A, B, C, D, respectively. The other pyrrolidine ring atom, C1, is the non-coplanar atom, E.

The non-planeness and the distances (\( d \)) between E and the mean ABCD plane are shown in Table 4 for all 23 proline conformations. Here \( d \) is calculated as the average of the distance between E and the plane (ABC) and the distance between E and the plane (BCD). The sign of \( d \) is positive if E and C4 are on the same side of the mean ABCD plane, and negative otherwise.

As seen in Table 4, the non-planeness is generally quite small, with a maximum of only 11.4° for conformer 8 and an average of 6° for all 23 conformers. The coplanar rule is, therefore, a fairly good approximation. In addition, it is noted that C2 is always one of the coplanar atoms, as shown in Table 4. Besides, the \( d \) values are all positive except for those of conformers 2 and 6.

The coplanar feature means that the five bond rotations, \( (c, d, e, f, g) \), about the pyrrolidine ring are inter-dependent, and we are left with only two independent bond rotations. Therefore, the eight bond rotations \( (a, b, c, d, e, f, g, h) \) shown in Figure 1(a) can be reduced to five independent bond rotations \( (a, b, c, d, h) \) in Figure 1(b). The number of full combinations of the rotational degrees of freedom is thus reduced to \( 4 \times 2 \times 3 \times 3 \times 2 = 144 \). The number of trial structures for the conformational search can, in fact, be further reduced. Notice that the bond rotation \( c \) (\( d \)) corresponds to the swing of C7 (C1) relative to the N8–C2–C3 plane. Specifically, \( c = -90^\circ, -120^\circ, \) and \(-150^\circ\) indicate that C7 is positioned above, on, and below the N8–C2–C3 plane, respectively. Meanwhile, \( d = 90^\circ, 120^\circ, \) and \( 150^\circ\) correspond to the fact that C1 is positioned above, on, and below the N8–C2–C3 plane, respectively. The four-atom coplanar rule can be satisfied if either C1 or C7 is on the N8–C2–C3 plane. Specifically, \( c = -90^\circ, -120^\circ, \) and \(-150^\circ\) indicate that C7 is positioned above, on, and below the N8–C2–C3 plane, respectively. Meanwhile, \( d = 90^\circ, 120^\circ, \) and \( 150^\circ\) correspond to the fact that C1 is positioned above, on, and below the N8–C2–C3 plane, respectively. The four-atom coplanar rule may satisfy the five bond rotations \( (c, d, e, f, g) \), about the pyrrolidine ring are inter-dependent, and we are left with only two independent bond rotations. Therefore, the eight bond rotations \( (a, b, c, d, e, f, g, h) \) shown in Figure 1(a) can be reduced to five independent bond rotations \( (a, b, c, d, h) \) in Figure 1(b). The number of full combinations of the rotational degrees of freedom is thus reduced to \( 4 \times 2 \times 3 \times 3 \times 2 = 144 \). The number of trial structures for the conformational search can, in fact, be further reduced. Notice that the bond rotation \( c \) (\( d \)) corresponds to the swing of C7 (C1) relative to the N8–C2–C3 plane. Specifically, \( c = -90^\circ, -120^\circ, \) and \(-150^\circ\) indicate that C7 is positioned above, on, and below the N8–C2–C3 plane, respectively. Meanwhile, \( d = 90^\circ, 120^\circ, \) and \( 150^\circ\) correspond to the fact that C1 is positioned above, on, and below the N8–C2–C3 plane, respectively. The four-atom coplanar rule can be satisfied if either C1 or C7 is on the N8–C2–C3 plane. Specifically, \( c = -90^\circ, -120^\circ, \) and \(-150^\circ\) indicate that C7 is positioned above, on, and below the N8–C2–C3 plane, respectively. Meanwhile, \( d = 90^\circ, 120^\circ, \) and \( 150^\circ\) correspond to the fact that C1 is positioned above, on, and below the N8–C2–C3 plane, respectively. The four-atom coplanar rule can be satisfied if either C1 or C7 is on the N8–C2–C3 plane. Specifically, \( c = -90^\circ, -120^\circ, \) and \(-150^\circ\) indicate that C7 is positioned above, on, and below the N8–C2–C3 plane, respectively. Meanwhile, \( d = 90^\circ, 120^\circ, \) and \( 150^\circ\) correspond to the fact that C1 is positioned above, on, and below the N8–C2–C3 plane, respectively.

Table 3. Equilibrium compositions (%) of proline conformers.

| Conformer | 98 K | 198 K | 298 K | 398 K | 498 K | 548 K |
|-----------|------|-------|-------|-------|-------|-------|
| 1         | 97.3 | 80.4  | 59.2  | 39.6  | 26.1  | 20.1  |
| 2         | 2.7  | 17.6  | 25.6  | 24.1  | 19.5  | 16.5  |
| 3         |      | 0.8   | 5.3   | 10.6  | 13.6  | 13.9  |
| 4         |      | 0.5   | 3.2   | 6.8   | 9.0   | 9.6   |
| 5         |      | 0.7   | 4.8   | 10.2  | 13.7  | 14.2  |
| 6         |      |       | 0.4   | 1.5   | 2.7   | 3.6   |
| 7         |      |       | 0.4   | 1.4   | 2.6   | 3.5   |
| 8         |      |       | 0.3   | 1.2   | 2.2   | 3.0   |
| 9         |      |       | 0.5   | 2.1   | 4.4   | 5.8   |
| 10        |      |       | 0.1   | 0.3   | 0.7   | 1.1   |
| 11        |      |       | 0.2   | 0.7   | 1.5   | 2.1   |
| 12        |      |       | 0     | 0.2   | 0.6   | 0.9   |
| 13        |      |       | 0     | 0.2   | 0.4   | 0.7   |
| 14        |      |       | 0.1   | 0.4   | 1.0   | 1.5   |
| 15        |      |       |       | 0.4   | 0.9   | 1.4   |
| 16        |      |       |       | 0.1   | 0.2   | 0.4   |
| 17        |      |       |       |       | 0.7   | 1.2   |
| 18        |      |       |       |       |       | 0.2   |
| 19        |      |       |       |       |       | 0.1   |
| 20        |      |       |       |       |       |       |
| 21        |      |       |       |       |       |       |
| 22        |      |       |       |       |       |       |
| 23        |      |       |       |       |       |       |
The pyrrolidine ring remains unchanged with the formation of a peptide bond at the N- and/or C-terminus of P. The four-atom coplanar rule of the pyrrolidine ring is also expected to hold for proline in peptides. The bond rotations for the conformational searches of proline and capped proline are, therefore, basically the same, except that the bond rotation $h$ (H–N8–C3–C4) for proline is changed to $t$ (C11–C9–N8–C3) for capped P, as shown in Figure 1 and Table 5.

Consequently, the number of trial structures generated for the conformational search of capped proline is 96, the same as that for natural P.

It should be pointed out that the puckered structures of the proline ring have been well-established in terms of the concept of pseudorotation by DeTar and Luthra. They described the conformational state of the five-membered ring of the proline side chain using only two parameters.
However, the new four-atom coplanar rule of the pyrrolidine ring revealed in this study is more advantageous in reducing the size of the conformational search space.

**Vibrational spectrum of natural proline**

The IR spectra of the four conformers 1, 5, 7, 22 and the weighted average of the 23-conformation ensemble at $T = 548$ K are shown and compared with the experimental result (Figure 3).

Conformers 1, 5, 7, and 22 have different H-bonds (Figure 2), resulting in different frequencies of the related vibrational modes (Figure 3). Due to the strong O–H...N H-bond of conformer 1, its O–H stretching mode, $q(OH)$, shows a large red-shift, as seen in Figure 3. Specifically, the $q(OH)$ frequencies are 3099, 3566, and 3560 cm$^{-1}$ for conformers 1, 5, and 7, respectively, in comparison with 3597 cm$^{-1}$ for conformer 22 with no H-bond. That is, the strong O–H...N H-bond causes a red-shift of about 500 cm$^{-1}$ for $q(OH)$ of conformer 1. The $q(OH)$ frequencies are 1419, 1125, 1120, and 1269 cm$^{-1}$ for conformers 1, 5, 7, and 22, respectively.

Notice that both the blue-shifted O–H bending of conformer 1 and the red-shifted bending of conformer 5 have high IR intensities (Figure 3). They are critically important for the correct interpretation of the experimental results. For example, the vibration for $\nu_{\exp} \approx 1367$ cm$^{-1}$ is missing in the computation of Linder et al.44 who used only one conformer in their calculations. Similarly, $\nu_{\exp} \approx 1110$ cm$^{-1}$ cannot be properly assigned to $\delta(OH)$ in the study by Stepanian et al.34 and in that by Ten and Shcherbakova.37

It is of practical interest to emphasize that conformations with similar H-bonds also produce similar IR spectra, as is well-documented in the literatures,52 and is also seen in Figure 3 for conformers 5 and 7. That is, as far as the IR spectrum is concerned, the ensemble average may be approximated by a mix of a few conformers with a proper coverage of characteristic H-bonds. For example, Table 3 shows that only conformers 1–5 are significant at $T = 548$ K, while Figure 2 shows that the H-bonds in conformers 1
and 2 or in conformers 3–5 are similar. At $T = 548$ K, the combined populations are 36.6% for conformers 1 and 2 and 37.7% for conformers 3, 4, and 5. Using a 50%/50% mix of the IR spectra of conformer 1 (or 2) and conformer 3 (or either 4 or 5) produces roughly the same IR spectrum as the ensemble average shown in Figure 3.

Table 6 compares the experimental and theoretical frequencies for the major IR peaks observed, together with their vibrational mode assignments. Notice that there are several C–H bonds, each with a somewhat distinct stretching frequency, forming a broad $q(CH)$ band. The lowest and highest C–H stretching frequencies are denoted in Table 6 as $q(CH)_L$ and $q(CH)_H$, respectively.

As seen in Table 6, the experimental and theoretical results are in excellent agreement. Very good agreement between the experimental and theoretical IR spectra is also seen in Figure 3, except that the theoretical $q(OH)$ peaks of conformers 1 and 2 are absent in the experiment. The complete absence of the theoretical $q(OH) = 3099/3157$ cm$^{-1}$ in the experimental IR spectrum is quite puzzling when considering the overall high quality theoretical and experimental agreement shown in Figure 3. One possible reason is that the scaling factor for $q(OH)$ associated with a strong H-bond should be smaller than the value of 0.97 assumed above. As noted by Stepanian et al., scaling factor of 0.92 for glycine, alanine, and valine, instead of the value of 0.96 suggested by calculations at the B3LYP/aug-cc-pVZD level. Benchmarked with 0.92 instead of 0.96 for B3LYP/aug-cc-pVZD, the scaling factor here for $q(OH)$ of conformer 1 or 2 should be reduced to 0.93, instead of 0.97 as used above. With the newly adjusted scaling factor, $q(OH)$ is found to be 2971/3027 cm$^{-1}$ for conformer 1 or 2. The revised $q(OH)$ value is close to $q(CH)_H = 2967$ cm$^{-1}$ and would merge with the $q(CH)$ modes to form the experimentally observed broad band of $\nu_{exp} \sim (2800$ cm$^{-1}, 3000$ cm$^{-1})$ shown in Figure 3. The theoretical and experimental IR spectra would, therefore, be in a perfect agreement.

Conformations of capped proline

Seven conformers are found for the capped proline, adding only the relatively unimportant conformers 5 and 7 to the existing results of Yuan et al. To facilitate further discussion, the structures and the intramolecular H-bonds of all the seven conformers of capped proline are shown in Figure 4.

As shown in Figure 4, the C-termini peptide plane of capped proline, O6C4N5H, always adopts the trans form. Moreover, the capped C-termini can only serve as H-bond donor, but not as an acceptor. The trans form and H-bond donor restrictions are responsible for the small number of conformations of capped proline. As with natural proline, four of the five pyrrolidine ring atoms of capped proline are coplanar, with a maximum deviation of 11.8° for conformer 2 and an average deviation of 6.6° for all the conformations. Due to the influence of the N-terminus capping group, however, both N8 and C3, instead of C2 for natural proline, are always the coplanar atoms. Besides, C4 and the non-coplanar atom, C1 or C2, are on the opposite side of the four-atom plane for six conformers of capped proline. Only for conformer 2 with C7 as the non-planar atom are C4 and C7 on the same side of the four-atom plane.

As the C-termini peptide plane of all conformations of capped proline unanimously adopt the trans form, the bond rotation $b$ is unnecessary for the conformational search of capped proline. As a result, the number of trial structures required is reduced from 96 to 48. The 48 initial trial structures are verified to reproduce all the seven conformations of capped proline. The reduction in the number of
trial structures is quite important for the conformational searches of proline-containing peptides.

The relative electronic energies, zero-point vibrational energies, Gibbs free energies at $T_0$, rotational constants and permanent dipole moments of all the capped proline conformers are shown in Table 7. As seen in Table 7, the seven conformers fall within an energy range of 6 kcal mol$^{-1}$ when measured by $\Delta G_0$.

The equilibrium distributions of capped proline conformers at temperatures between 98 and 548 K are shown in Table 8. As shown in Table 8, the capped proline exists predominantly as conformer 1 at $T \approx T_0$. The compositions of conformers 2, 3, and 4 increase notably with $T$. Their combined compositions are 2%, 11%, 24%, 35%, and 42% for $T = 198$, 298, 398, 498, and 548 K, respectively.

**Figure 4.** Conformations of capped proline (P). Intramolecular H-bonds are noted with dashed green lines. The H-bond lengths are also indicated.
Cis and trans peptide bond structures of proline-containing peptides

As shown in Figure 4, the C-terminus peptide plane adopts the trans form for all capped proline conformations, but the N-terminus peptide plane may be in either the trans or the cis form. Specifically, the N-termini peptide bonds of conformers 1, 2, and 5 are trans, while those of conformers 3, 4, 6, and 7 are cis. Notice that the N- and C-termini peptide planes are, respectively, the analogue of the X–P and P–X_{np} peptide planes in a general peptide chain. Here, X–P (P–X_{np}) means X (X_{np}) is connected to P and on the N (C)-terminal side of P. X can be any AA residue, while X_{np} can also be any AA residue excluding P. The trans–cis contents of the capped proline conformations can be used to mimic the populations of trans–cis X–P and P–X_{np} peptide planes.

Using the equilibrium distributions shown in Table 8, the proportions of the cis X–P and P–X_{np} peptide bonds at different temperatures can be calculated and are displayed as the blue and orange lines in Figure 5, respectively. As shown in Figure 5, the population of cis X–P peptide bonds increases with the temperature, and is 5.5% at T₀. The computational results agree well with the early observation of Chen and Bovey that there are 5%-10% cis structures for the X–P peptide plane, but essentially nil for the P–X_{np} peptide plane. The theoretical results are also consistent with the recent report by Alderson et al. By using multi-dimen-
sional nuclear magnetic resonance (NMR) to quantify the cis–trans peptide bonds in unfolded proteins at T₀, Alderson et al. found that peptide bonds are always trans, except for 3%-10% cis structures for the X–P peptide bonds. The experiment of Jabs et al. estimated that there are about 0.03% cis peptide bonds for P–X_{np}. Although their results were not exactly the same as our result of 0% cis P–X_{np}, they do agree quite well. The minor theoretical and experimental discrepancy indicates that the environment experienced by the P–X_{np} peptide bond in a long peptide chain may sometimes induce a potential energy barrier for the bond rotation b. Nevertheless, the experiment does show clearly the energy of the cis configuration, when as a local minimum of the PES, is much higher than that of the trans form. The cis form of the P–X_{np} peptide plane can almost always be neglected. A total of 48 trial structures per proline residue suffice for the full conformation search of proline-containing peptides, such as proline-containing AMPs.

Due to a lack of definite and precise measurements, a detailed quantitative theoretical, and experimental comparison cannot be made yet. However, the computational

| Conformer | ΔE  | ΔVE₀ | ΔG₀  | Rotation constants | D   |
|-----------|-----|------|------|--------------------|-----|
|           |     |      |      | x                 |     |
| 1         | 0.00| 0.00 | 0.00 | 1.13              | 0.56|
| 2         | 1.46| 0.25 | 1.64 | 1.16              | 0.83|
| 3         | 2.45| -0.25| 1.74 | 1.14              | 0.79|
| 4         | 3.43| -0.26| 2.77 | 1.14              | 0.79|
| 5         | 4.72| -0.22| 4.08 | 1.16              | 0.79|
| 6         | 5.51| -0.22| 4.84 | 1.27              | 0.75|
| 7         | 6.05| -0.26| 5.38 | 1.28              | 0.74|

Table 7. Relative electronic energies (ΔE, kcal mol⁻¹), relative ZVPEs (ΔVE₀, kcal mol⁻¹), relative Gibbs free energies at T₀ (ΔG₀, kcal mol⁻¹), rotation constants (GHz) and permanent dipole moments (D, Debye) of the capped proline conformers determined at the B97D/6-311++G** level of theory.

| Conformer | 98 K | 198 K | 298 K | 398 K | 498 K | 548 K |
|-----------|------|------|------|------|------|------|
| 1         | 100  | 97.8 | 88.7 | 75.5 | 62.5 | 53.4 |
| 2         | —    | 1.5  | 5.6  | 9.6  | 12.1 | 13.7 |
| 3         | —    | 0.7  | 4.7  | 11.1 | 17   | 20   |
| 4         | —    | 0.1  | 0.8  | 2.9  | 5.7  | 7.9  |
| 5         | —    | —    | 0.1  | 0.6  | 1.5  | 2.7  |
| 6         | —    | —    | —    | 0.2  | 0.7  | 1.5  |
| 7         | —    | —    | —    | 0.1  | 0.4  | 0.9  |

Table 8. Equilibrium contents (%) of the conformers of capped proline.

Using the equilibrium distributions shown in Table 8, the proportions of the cis X–P and P–X_{np} peptide bonds at different temperatures can be calculated and are displayed as the blue and orange lines in Figure 5, respectively. As shown in Figure 5, the population of cis X–P peptide bonds increases with the temperature, and is 5.5% at T₀. The computational results agree well with the early observation of Chen and Bovey that there are 5%-10% cis structures for the X–P peptide plane, but essentially nil for the P–X_{np} peptide plane. The theoretical results are also consistent with the recent report by Alderson et al. By using multi-dimen-
sional nuclear magnetic resonance (NMR) to quantify the cis–trans peptide bonds in unfolded proteins at T₀, Alderson et al. found that peptide bonds are always trans, except for 3%-10% cis structures for the X–P peptide bonds. The experiment of Jabs et al. estimated that there are about 0.03% cis peptide bonds for P–X_{np}. Although their results were not exactly the same as our result of 0% cis P–X_{np}, they do agree quite well. The minor theoretical and experimental discrepancy indicates that the environment experienced by the P–X_{np} peptide bond in a long peptide chain may sometimes induce a potential energy barrier for the bond rotation b. Nevertheless, the experiment does show clearly the energy of the cis configuration, when as a local minimum of the PES, is much higher than that of the trans form. The cis form of the P–X_{np} peptide plane can almost always be neglected. A total of 48 trial structures per proline residue suffice for the full conformation search of proline-containing peptides, such as proline-containing AMPs.

Due to a lack of definite and precise measurements, a detailed quantitative theoretical, and experimental comparison cannot be made yet. However, the computational

Figure 5. Percentages of the equilibrium contents of X–P and P–X_{np} conformations with cis peptide bonds.
results here do provide, for the first time to our knowledge, a very good quantitative explanation of the experimentally observed population of cis X–P peptide bonds.

Conclusion

Through comprehensive PES exploration, a total of 23 proline conformations have been found, among which 5 conformers are newly identified. The 23 conformations span an energy range of 10 kcal mol$^{-1}$, with 5 of them existing less than 3 kcal mol$^{-1}$ above the global minimum. The relative energies, rotation constants, dipole moments, relative ZPVEs, relative Gibbs free energies, and equilibrium distributions of the proline conformations as determined at the B97D/6-311++G** level are presented. Structural analysis reveals a coplanar rule for four of the five atoms of the pyrrolidine ring. The coplanar rule can reduce the computational cost of the conformational search by a factor of 40, while producing the same results as those of the thorough search method. The equilibrium distribution of the conformational set provides a very good explanation of the experimental IR spectrum measured at $T = 548$ K.

The conformations of capped proline are searched by the method incorporating the four-atom coplanar rule of the pyrrolidine ring. In addition to the five conformations known previously, two new conformers of capped proline are found. The seven conformers fall into an energy range of 6 kcal mol$^{-1}$. The capped proline conformations indicate that all P–Xnp peptide planes should adopt the trans form. The trans form restriction can be used to cut by half the cost of the conformational search of the proline-containing peptide, requiring only 48 combinations of bond rotations for a complete PES scan of a proline residue. The X–P peptide bond may adopt either the trans or the cis forms. The equilibrium compositions of the cis X–P peptide bonds for 98 K $\leq T \leq 548$ K have been presented. The theoretical composition of cis X–P peptide bonds at $T_g$ is 5.5%, agreeing well with the experimental estimations of 3%–10%.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful for the financial support from the National Natural Science Foundation of China (grant nos. 12074362 and 11774324) and the computing time at the Supercomputing Center of the University of Science and Technology of China.

ORCID iDs

Yizhou Yang https://orcid.org/0000-0003-0074-1979
Zijing Lin https://orcid.org/0000-0001-9270-1717

References

1. Adams E and Frank L. *Annu Rev Biochem* 1980; 49: 1005.
2. Phang JM. *Curr Top Cell Regul* 1985; 25: 91.
3. Phang JM. *J Antioxidants* 2019; 30: 635.
4. Tanner JJ. *J Antioxidants* 2019; 30: 650.
5. Christgen SL and Becker DF. *J Antioxidants* 2019; 30: 683.
6. Wood JM. *Annu Rev Microbiol* 2011; 65: 215.
7. Krishnan N, Dickman MB and Becker DF. *Free Radic Biol Med* 2008; 44: 671.
8. Natarajan SK, Zhu W, Liang X, et al. *Free Radic Biol Med* 2012; 53: 1181.
9. Liang X, Dickman MB and Becker DF. *J Biol Chem* 2014; 289: 27794.
10. Zhang L, Alfano JR and Becker DF. *J Bacteriol* 2015; 197: 431.
11. Liu Y, Borchert GL, Donald SP, et al. *J Carcinogenesis* 2005; 26: 1335.
12. Donald SP, Sun XY, Hu CA, et al. *Cancer Res* 2001; 61: 1810.
13. Liu W, Le A, Hancock C, et al. *Proc Natl Acad Sci USA* 2012; 109: 8983.
14. Liu W, Hancock CN, Fischer JW, et al. *Sci Rep* 2015; 5: 1.
15. Liu W, Zabirnyk O, Wang H, et al. *J Oncogene* 2018; 29: 4914.
16. Liu Y, Borchert GL, Donald SP, et al. *Cancer Res* 2009; 69: 6414.
17. Elia I, Broekaert D, Christen S, et al. *Nat Commun* 2017; 8: 1.
18. Ding JF, Kuo ML, Su LL, et al. *J Carcinogenesis* 2017; 38: 519.
19. Cai F, Miao YY, Liu CY, et al. *Oncol Lett* 2018; 15: 731.
20. Olivares O, Mayers JR, Gourivand V, et al. *Nat Commun* 2017; 8: 1.
21. Craze ML, Cheung H, Jewa N, et al. *Br J Cancer* 2018; 118: 258.
22. Mishra AK, Choi J, Moon E, et al. *J Molecules* 2018; 23: 815.
23. Delaney NG and Madison V. *J Am Chem Soc* 1982; 104: 6635.
24. Park HS and Kang YK. *New J Chem* 2019; 43: 17159.
25. Alderson TR, Lee JH, Charlier C, et al. *ChemBioChem* 2018; 19: 37.
26. Cheng H and Bovey F. *Biopolymers* 1977; 16: 1465.
27. Gustafson CL, Parsley NC, Asimgil H, et al. *J Mol Cell* 2017; 66: 447.
28. Andreasen M, Skeby KK, Zhang S, et al. *Biochemistry* 2014; 53: 6968.
29. Tao K, Wang QJ, Zhou P, et al. *Langmuir* 2011; 27: 2723.
30. Chen MS, Zhao DS, Yu YP, et al. *Chem Commun* 2013; 49: 1799.
31. Sapse AM, Mallab-Levy L, Daniels SB, et al. *J Am Chem Soc* 1987; 109: 3526.
32. Ramek M, Kelterer AM and Nikolić S. *Int J Quantum Chem* 1997; 65: 1033.
33. Császár AG and Perczel A. *Prog Biophys Mol Biol* 1999; 71: 243.
34. Stepanian S, Reva I, Radchenko E, et al. *J Phys Chem A* 2001; 105: 10664.
35. Lesarri A, Mata S, Cosciner EJ, et al. *Angew Chem Int Ed* 2002; 41: 4673.
36. Czinkl E and Császár AG. *Chem Eur J* 2003; 9: 1008.
37. Ten GN and Sherbakova NE. *Proc SPIE* 2020: 114580U.
38. Yuan Y, Mills MJ, Popelier PL, et al. *J Phys Chem A* 2014; 118: 7876.
39. Yang KY, Yang B and Lin ZJ. *Chin J Chem Phys* 2015; 28: 161.
40. Ru X, Song C and Lin ZJ. *J Comput Chem* 2016; 37: 1214.
41. Zhang ML, Huang ZJ and Lin ZJ. *J Chem Phys* 2005; 122: 134313.
42. Savitzky A and Golay MJ. Anal Chem 1964; 36: 1627.
43. Frisch M, Trucks G, Schlegel H, et al. Gaussian 09 2009; 32: 5648.
44. Linder R, Nispel M, Häber T, et al. Chem Phys Lett 2005; 409: 260.
45. Politzer P, Murray JS and Lane P. Int Quantum Chem 2007; 107: 3046.
46. Popelier P. J Phys Chem A 1998; 102: 1873.
47. Ling S, Yu W, Huang Z, et al. J Phys Chem A 2006; 110: 12282.
48. Görbitz C. Acta Crystallogr B 1989; 45: 390.
49. Stepanian S, Reva I, Radchenko E, et al. J Phys Chem A 1998; 102: 4623.
50. Arunan E, Desiraju GR, Klein RA, et al. Pure Appl Chem 2011; 83: 1637.
51. DeTar DF and Luthra NP. J Am Chem Soc 1977; 99: 1232.
52. Sandorfy C. In: Hydrogen bonds. Springer, 1984, p. 41.
53. Stepanian S, Reva I, Radchenko E, et al. J Phys Chem A 1998; 102: 1041.
54. Stepanian S, Reva I, Radchenko E, et al. J Phys Chem A 1999; 103: 4404.
55. Jabs A, Weiss MS and Hilgenfeld R. J Mol Biol 1999; 286: 291.