Where is the Hidden Intramolecular H-bonding Vibrational Signal in Proline?

James Langford,[a] Yuzhe Zhang,[a] Zehua Chen,[a] and Yang Yang*[a]

The frequency of the hydrogen bonded O-H stretch vibration in proline has sparked controversy. Employing constrained nuclear electronic orbital methods developed by our group, we provide a clear assignment that the vibrational frequency drops to near 3000 cm^{-1} as a result of the strong hydrogen bond with significant nuclear quantum effects in proline.

Hydrogen bonding interactions greatly influence the structure and function of proteins, impacting critical properties such as protein folding and enzyme activity.[1–3] Recently, nuclear quantum effects (NQEs) have been shown to play a crucial role in modulating the properties of hydrogen bonds, particularly within the active sites of enzymes.[4–8]

Given that amino acids and short peptides function as the fundamental building blocks of proteins, it is essential to develop a comprehensive understanding of the hydrogen bonding interactions within these systems as well as the influence of NQEs on these hydrogen bonds. Nevertheless, the impact of NQEs on the hydrogen bonding of amino acids and short peptides remains largely unexplored, representing an important direction for research.

Among the naturally occurring amino acids, proline assumes a captivating role owing to its distinctive structure.[9] Unlike other amino acids, proline's side chain is covalently bonded to the amidic nitrogen, thereby forming a characteristic pyrrolidine loop. With this
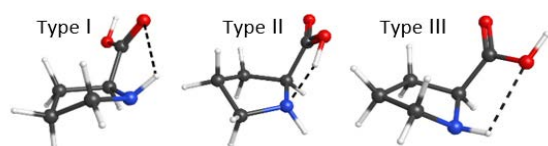

Figure 1. Geometries of proline conformers, with hydrogen bonding interactions depicted by dashed lines.

distinctive structure, proline often has a remarkable impact on protein secondary structure, including locally disrupting alpha helices,[10,11] terminating both alpha helices and beta sheets,[9,11–13] and forming unique kind of protein secondary structure known as a polyproline helix.[9,14,15] Furthermore, proline catalyzes a wide range of reactions including the aldol and Mannich reactions.[16–18]

As with other amino acids, proline can adopt multiple conformational isomers, each with different patterns of intramolecular hydrogen bonding. In general, these conformational isomers can be categorized into three types based on their intramolecular hydrogen bonding patterns: as shown in Figure 1, Type I features an N−H⋯O=C interaction, Type II features an N⋯H−O interaction, and Type III features an N−H⋯O-H interaction.[19,20] In other words, Types I and III feature hydrogen bonded N-H stretches and free O-H stretches while Type II features a free N-H stretch and a hydrogen bonded O-H stretch. In general, Type I and Type III conformers are highly similar in their vibrational spectra, and we may collectively call them Type I/III. The energy ordering of these three types of conformers can vary depending on the amino acid, but for proline, past experimental and theoretical investigations show that the Type II conformer is the lowest in energy.[20–26]

For all other amino acids studied so far, both Type I/III and Type II conformers can be clearly identified through IR spectroscopy with their characteristic O-H stretching peaks appearing near 3560 cm^{-1} and 3200 cm^{-1}, respectively.[27–34] However, for proline, although strong evidence exists for the presence of both the Type I/III and Type II conformers including a peak splitting in the C=O stretch region,[20,35] only the Type I/III O-H stretch frequency can be unequivocally identified experimentally with a sharp peak appearing around 3560 cm^{-1} with no peak distinguishable from the baseline appearing near 3200 cm^{-1}.[20,35,36] This raises a perplexing question: where is the hydrogen bonded O-H vibrational signal in the Type II proline isomer?

Adamowicz and co-workers originally proposed that the hydrogen bonded O-H stretch peak appears in the C-H stretching region and is thus obscured by the C-H stretch peaks.[20] This assignment was determined by the experimental difference spectrum result between proline and doubly deuterated proline (proline-D_2, with N-H and O-H deuterated), which shows a peak at 3025

cm[-1].[20] Computationally, scaled density functional theory (DFT) harmonic analysis by Adamowicz and co-workers[20] as well as Yang and Lin[26] additionally supported this assignment. However, pure DFT-based harmonic analysis is unable to capture anharmonicity and NQEs, and the use of empirical scaling factors to account for these effects is *ad hoc* and not fully reliable.

Recently, using the adiabatically switched semiclassical initial value representation (AS-SCIVR), a method that incorporates NQEs,[25,37,38] Conte and co-workers proposed a new assignment of the hydrogen bonded O-H stretch frequency of Type II proline at 3329 cm[-1],[25] which is near the previously assigned N-H stretching region and significantly outside the C-H stretching region.[20,35] Conte and co-workers disputed the past 3025 cm[-1] assignment from the experimental difference spectrum, noting that this peak is very weak.[25] However, their new peak assignment is not completely convincing either, as the experimental spectra show only a weak and broad signal in the 3300-3400 cm[-1] region previously identified as N-H stretches,[20,35] and it would be unusual if this weak signal could correspond to both O-H and N-H stretches. Furthermore, while AS-SCIVR is inherently capable of describing NQEs, to our best knowledge, its ability to accurately predict the vibrational spectra of hydrogen bonding systems has not yet been thoroughly established with a limited set of molecules studied so far.[37,38] Therefore, further theoretical investigations are needed to locate the hidden hydrogen bonded O-H stretching mode in Type II proline.

In this communication, we will respond to this challenge and provide a clear assignment of this O-H stretch mode employing both the constrained nuclear-electronic orbital (CNEO) theory[39–41] and the vibrational second-order perturbation theory (VPT2). VPT2 is a well-established method for anharmonic vibrational analysis,[42,43] while the CNEO theory, which was recently developed in our group, can incorporate NQEs accurately and efficiently in both quantum chemistry calculations[39–41] and molecular dynamics simulations.[44–49] The key difference between CNEO methods and conventional Born-Oppenheimer-based methods is that CNEO treats nuclei quantum mechanically while retaining the classical molecular geometry picture through constraints on the expectation values of nuclear position operators.[39,40] Recently, both CNEO harmonic analysis[41] and CNEO molecular dynamics (CNEO-MD)[44] have successfully been applied to a series of molecular systems with challenging hydrogen motion and/or hydrogen bonding characters,[41,44,47–49] demonstrating their promise for tackling the challenge presented here with proline. Here, with both CNEO harmonic and VPT2 analyses, we will provide strong evidence supporting the original assignment of the Type II hydrogen bonded O-H stretch as appearing close to the C-H stretching region,

and with CNEO-MD, we will further validate this assignment by reproducing the experimental difference spectrum. Finally, we will provide the rationale for the 200 cm[-1] redshift for the proline Type II hydrogen bonded O-H stretch as compared to other amino acids.

In this work, VPT2 calculations were performed with the Gaussian 16[50] package and CNEO density functional theory (CNEO-DFT) calculations were performed with our in-house version of PySCF.[51–53] The PBE0[54–56] electronic functional was adopted in all calculations as it has been shown to give accurate vibrational results for hydrogen-bonded systems.[47,48] The aug-cc-pVDZ basis set was used for electrons,[57] while the PB4-D basis was used for protons.[58] For CNEO-MD simulations, we use the direct *NVE* scheme that our group recently introduced for efficient spectrum calculations.[48] In the direct *NVE* scheme, we begin by assigning k_BT of energy to each vibrational mode obtained from the CNEO-DFT harmonic analysis. Then, we randomly generate 50 starting configurations with different initial velocities for *NVE* simulations. IR spectra are obtained by taking the average of the Fourier transformed dipole derivative autocorrelation functions.

Table 1. O-H Stretching Frequencies (cm[-1]) in Proline

	Expt.[a]	CNEO-DFT[b]	VPT2[b]	AS-SCIVR[b]	DFT[b]	Scaled DFT[b]
Type I	3559	3601	3605	3522	3793	-
Type II	3025 (disputed)	2958	3016	3329	3336	3071

a. Experimental results are obtained from Ref 20.
b. All computational values calculated with PBE0/aug-cc-pVDZ, except for the AS-SCIVR results, which were obtained from Ref 25 calculated with B3LYP-D3/aug-cc-pVDZ.

The hydrogen bonded O-H vibrational frequencies for the Type II proline are provided in Table 1. As mentioned above, Adamowicz and co-workers assigned the experimental frequency to be at 3025 cm[-1], as observed as a wide and relatively faint peak in the difference spectrum. Without NQEs, unscaled DFT harmonic analysis predicts the frequency to be 3336 cm[-1], which is around 300 cm[-1] higher than the experimental assignment. When the empirical scaling factor as used by Adamowicz and co-workers for B3LYP was applied here to PBE0, the frequency sharply drops to 3069 cm[-1] and becomes closer to the difference spectrum result of 3025 cm[-1]. However, the empirical nature of this scaling factor is hard to justify and often needs to be varied with vibrational modes as well as functional choices. The AS-SCIVR result by Conte is 3329 cm[-1], which, interestingly, is similar to the unscaled DFT harmonic result. Although the unscaled harmonic result is above 3300 cm[-1], we find that with VPT2, the frequency significantly drops to 3018 cm[-1], supporting the assignment by Adamowicz

and co-workers. With the CNEO-DFT harmonic analysis, the frequency is expected to be 2958 cm^{-1}, which is even lower and also qualitatively supports the experimental assignment. Note that it has been shown in previous studies that with NQEs but no additional anharmonic effects, CNEO-DFT harmonic analysis already can account for most of the errors in conventional DFT harmonic analysis.[41,44,47,48] Hence, this 300 cm^{-1} shift relative to the DFT harmonic result is mainly due to the inclusion of NQEs through the underlying effective potential energy surface by CNEO-DFT.
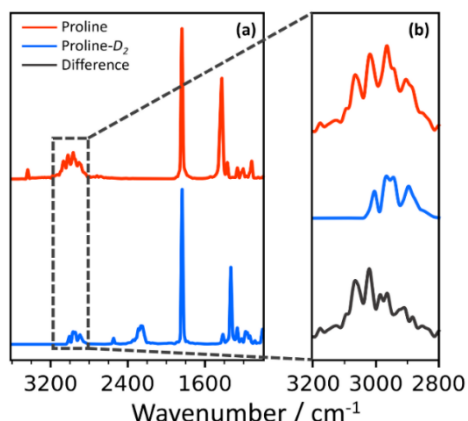

Figure 2. (a) IR spectra of Type II proline and proline-D_2 calculated with CNEO-MD. (b) Inset of 2800-3200 cm-1 region, including the difference spectrum of proline and proline-D_2.

In order to obtain more accurate CNEO results, we additionally performed CNEO-MD simulations. In modes with relatively small mode coupling effects such as free OH stretches, CNEO-MD and CNEO harmonic analysis tend to give essentially the same results, whereas for hydrogen bonded modes which often have strong mode coupling effects, these two methods can generate larger differences.[48] In general, CNEO-MD will be more accurate because of the incorporation of mode coupling effects with thermal motions.[47,48] The simulated IR spectra by CNEO-MD for Type II proline and proline-D_2 are provided in Figure 2(a). For both molecules, a significant band appears in the simulated spectra beginning near 2800 cm^{-1}. However, for proline-D_2, this band ends at around 3000 cm^{-1} whereas for regular proline a tail extends to about 3200 cm^{-1}. This tail is consistent with previous experimental observations and suggests that although being weak, it is related to the O-H stretch motion. The original experimental assignment can be further supported by the simulated CNEO-MD difference spectrum in Figure 2(b). This difference spectrum shows a broad band centered at around 3025 cm^{-1}, matching excellently with the experimental difference spectrum,[20] and notably contrasting with the highest peak of the regular proline spectrum, which here appears slightly below 3000 cm^{-1}. Therefore, together with the CNEO harmonic and VPT2 analyses

above, these results provide an unequivocal assignment of the peak position for the Type II hydrogen bonded O-H stretch close to the C-H stretch region near 3000 cm^{-1}. Note that by accounting for anharmonicity and mode coupling effects, CNEO-MD blueshifts the peak position of the CNEO Hessian result from 2958 cm^{-1} to around 3025 cm^{-1}. This shift for hydrogen-bonded O-H stretches has been observed in a study of water cluster systems by our group.[48] In general, the mode coupling between the hydrogen-bonded O-H stretch and low-frequency soft modes provides a notable peak shift as well as a large peak broadening.

Given that the Type II hydrogen bonded O-H stretch has been unequivocally assigned, now the remaining question is: why is this frequency significantly redshifted to around 3000 cm^{-1} in contrast to the ~3200 cm^{-1} results observed for other amino acids? We comment that this is in fact a simple organic chemistry question: Because secondary amines are more nucleophilic than primary amines, the corresponding N···H−O hydrogen bond in proline will be stronger than those in other amino acids. This stronger hydrogen bond will correspondingly weaken the O-H bond more and thus lead to a lower O-H vibrational frequency. This rationale can be further supported by the optimized geometries of Type II proline and another Type II amino acid, and here we use glycine for a comparison. With CNEO-DFT, the O-H and N-H distances are 1.031 Å and 1.711 Å for proline, and 0.986 Å and 1.924 Å for glycine. Similar results can also be obtained from DFT geometry optimizations with a longer O-H bond and shorter N···H hydrogen bond for proline than for glycine, demonstrating the stronger N···H interaction in proline relative to glycine.

In conclusion, by employing CNEO methods to account for NQEs in proline, we have unequivocally assigned the Type II proline hydrogen bonded O-H stretch to a broad band overlapping with the C-H stretch region near 3000 cm^{-1}. CNEO-DFT harmonic analysis and VPT2 qualitatively support this assignment, and the CNEO-MD result further quantitatively recovers the broad O-H stretching band feature in experiment as well as accurately predicts the center position and broadness of the experimental difference spectrum. Furthermore, these results again demonstrate that by incorporating NQEs with CNEO methods, IR spectra of hydrogen bonding systems can be accurately assigned from first principles, suggesting the potential of employing CNEO methods for future investigations of hydrogen bonding interactions in other chemical and biological systems.

JL performed the CNEO-DFT and VPT2 calculations, prepared the initial draft, and assisted with revisions. YZ performed CNEO-MD calculations, created figures, and assisted with revisions. ZC helped with CNEO calculations and assisted with revisions. ZC and YY supervised the project. JL and YY conceptualized the project.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 G. J. Narlikar and D. Herschlag, *Annu. Rev. Biochem.*, 1997, **66**, 19–59.
2 R. E. Hubbard and M. Kamran Haider, in *eLS*, John Wiley & Sons, Ltd, 2010.
3 D. Herschlag and M. M. Pinney, *Biochemistry*, 2018, **57**, 3338–3352.
4 J.-K. Hwang and A. Warshel, *J. Am. Chem. Soc.*, 1996, **118**, 11745–11751.
5 J. Pu, J. Gao and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 3140–3169.
6 L. Wang, S. D. Fried, S. G. Boxer and T. E. Markland, *Proceedings of the National Academy of Sciences*, 2014, **111**, 18454–18459.
7 A. Vardi-Kilshtain, N. Nitoker and D. T. Major, *Archives of Biochemistry and Biophysics*, 2015, **582**, 18–27.
8 T. E. Markland and M. Ceriotti, *Nat Rev Chem*, 2018, **2**, 1–14.
9 A. A. Morgan and E. Rubenstein, *PLOS ONE*, 2013, **8**, e53785.
10 R. Sankararamakrishnan and S. Vishveshwara, *Biopolymers*, 1990, **30**, 287–298.
11 L. Piela, G. Némethy and H. A. Scheraga, *Biopolymers*, 1987, **26**, 1587–1600.
12 S. C. Li, N. K. Goto, K. A. Williams and C. M. Deber, *Proceedings of the National Academy of Sciences*, 1996, **93**, 6676–6681.
13 C. M. Deber and A. G. Therien, *Nat Struct Mol Biol*, 2002, **9**, 318–319.
14 A. A. Adzhubei and M. J. E. Sternberg, *Journal of Molecular Biology*, 1993, **229**, 472–493.
15 A. A. Adzhubei, M. J. E. Sternberg and A. A. Makarov, *Journal of Molecular Biology*, 2013, **425**, 2100–2132.
16 S. Bahmanyar and K. N. Houk, *Org. Lett.*, 2003, **5**, 1249–1251.
17 S. K. Panday, *Tetrahedron: Asymmetry*, 2011, **22**, 1817–1847.
18 R. B. Sunoj, *WIREs Computational Molecular Science*, 2011, **1**, 920–931.
19 R. D. Suenram and F. J. Lovas, *J. Am. Chem. Soc.*, 1980, **102**, 7180–7184.
20 S. G. Stepanian, I. D. Reva, E. D. Radchenko and L. Adamowicz, *J. Phys. Chem. A*, 2001, **105**, 10664–10672.
21 S. Mata, V. Vaquero, C. Cabezas, I. Peña, C. Pérez, J. C. López and J. L. Alonso, *Physical Chemistry Chemical Physics*, 2009, **11**, 4141–4144.
22 R. Hadidi, Božanić, H. Ganjitabar, G. A. Garcia, I. Powis and L. Nahon, *Commun Chem*, 2021, **4**, 1–14.
23 E. Czinki and A. G. Császár, *Chemistry – A European Journal*, 2003, **9**, 1008–1019.
24 F. Fathi and H. Farrokhpour, *Chemical Physics Letters*, 2013, **565**, 102–107.
25 G. Botti, C. Aieta and R. Conte, *J. Chem. Phys.*, 2022, **156**, 164303.
26 Y. Yang and Z. Lin, *Journal of Chemical Research*, 2022, **46**, 17475198221110480.
27 S. G. Stepanian, I. D. Reva, E. D. Radchenko, M. T. S. Rosado, M. L. T. S. Duarte, R. Fausto and L. Adamowicz, *J. Phys. Chem. A*, 1998, **102**, 1041–1054.
28 S. G. Stepanian, A. Yu. Ivanov and L. Adamowicz, *Chemical Physics*, 2013, **423**, 20–29.
29 S. G. Stepanian, I. D. Reva, E. D. Radchenko and L. Adamowicz, *J. Phys. Chem. A*, 1998, **102**, 4623–4629.
30 S. G. Stepanian, I. D. Reva, E. D. Radchenko and L. Adamowicz, *J. Phys. Chem. A*, 1999, **103**, 4404–4412.
31 F. Huisken, O. Werhahn, A. Yu. Ivanov and S. A. Krasnokutski, *The Journal of Chemical Physics*, 1999, **111**, 2978–2984.
32 S. Jarmelo, L. Lapinski, M. J. Nowak, P. R. Carey and R. Fausto, *J. Phys. Chem. A*, 2005, **109**, 5689–5707.
33 A. Kaczor, I. D. Reva, L. M. Proniewicz and R. Fausto, *J. Phys. Chem. A*, 2006, **110**, 2360–2370.
34 J. Cz. Dobrowolski, M. H. Jamróz, R. Kołos, J. E. Rode and J. Sadlej, *ChemPhysChem*, 2007, **8**, 1085–1094.
35 I. D. Reva, S. G. Stepanian, A. M. Plokhotnichenko, E. D. Radchenko, G. G. Sheina and Yu. P. Blagoi, *Journal of Molecular Structure*, 1994, **318**, 1–13.
36 R. Linder, M. Nispel, T. Häber and K. Kleinermanns, *Chemical Physics Letters*, 2005, **409**, 260–264.
37 R. Conte, L. Parma, C. Aieta, A. Rognoni and M. Ceotto, *J. Chem. Phys.*, 2019, **151**, 214107.
38 G. Botti, M. Ceotto and R. Conte, *J. Chem. Phys.*, 2021, **155**, 234102.
39 X. Xu and Y. Yang, *J. Chem. Phys.*, 2020, **153**, 074106.
40 X. Xu and Y. Yang, *The Journal of Chemical Physics*, 2020, **152**, 084107.
41 X. Xu and Y. Yang, *J. Chem. Phys.*, 2021, **154**, 244110.
42 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.
43 P. R. Franke, J. F. Stanton and G. E. Douberly, *J. Phys. Chem. A*, 2021, **125**, 1301–1324.
44 X. Xu, Z. Chen and Y. Yang, *J. Am. Chem. Soc.*, 2022, **144**, 4039–4046.
45 Z. Chen and Y. Yang, *J. Phys. Chem. Lett.*, 2023, **14**, 279–286.
46 Y. Wang, Z. Chen and Y. Yang, *J. Phys. Chem. A*, 2023, **127**, 5491–5501.
47 Y. Zhang, X. Xu, N. Yang, Z. Chen and Y. Yang, *The Journal of Chemical Physics*, 2023, **158**, 231101.
48 Y. Zhang, Y. Wang, X. Xu, Z. Chen and Y. Yang, 2023. ChemRxiv: 10.26434/chemrxiv-2023-0lb60
49 X. Xu, *J. Phys. Chem. A*, 2023, **127**, 6329–6334.
50 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P.

Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16 Revision C.01 Gaussian, Inc., Wallingford, CT 2019.

51 theorychemyang/pyscf: Python-based Simulations of Chemistry Framework theorychemyang 2023.

52 Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters and G. K.-L. Chan, *WIREs Computational Molecular Science*, 2018, **8**, e1340.

53 Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Yu. Sokolov and G. K.-L. Chan, *J. Chem. Phys.*, 2020, **153**, 024109.

54 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

55 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

56 C. Adamo and V. Barone, *The Journal of Chemical Physics*, 1999, **110**, 6158–6170.

57 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

58 Q. Yu, F. Pavošević and S. Hammes-Schiffer, *J. Chem. Phys.*, 2020, **152**, 244123.

59 Center for High Throughput Computing, 2006. doi:10.21231/GNT1-HW21