Raman spectral analyses of amino acids in life processes

Lianxin Xin, Jie Hou, Aleem Sayles, Jian Zhao, Aristides Marcano, et al.

Lianxin Xin, Jie Hou, Aleem Sayles, Jian Zhao, Aristides Marcano, Hui Xia, Jun Ren, "Raman spectral analyses of amino acids in life processes," Proc. SPIE 10885, Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics, 108850O (20 February 2019); doi: 10.1117/12.2509883

Event: SPIE BiOS, 2019, San Francisco, California, United States

Raman Spectral Analysis of Amino Acids in Life Processes

Lianxin Xin[a], Jie Hou[a], Aleem Sayles[a], Jian Zhao[a], Aristides Marcano[a], Hui Xia[a], Jun Ren[a*]
[a]Delaware State Univ., 1200 Dupont Hwy., Dover, DE USA 19901

ABSTRACT

Amino acids are the basic "building blocks" of peptides and proteins and play important roles in the physiological processes of all species. In this study, we simulated the Raman spectrum of Glycine, Tyrosine and Phenylalanine using General Atomic and Molecular Electronic Structure System (GAMESS) and Gaussian, two computational codes that perform calculations of electronic and vibrational properties of molecules. Through our work, strong bands with N-H and O-H bonds and with benzyl ring were pinpointed and identified. Our work presents insights into the importance of intermolecular bonding of amino acids in the life and physiological processes, including metabolism, signal transduction, and neurotransmission etc.

Keywords: Raman Spectroscopy, Amino Acids, Quantum Chemical Computational

1. INTRODUCTION

Raman spectroscopy provides information on sample molecular composition and structure by inelastic scattering of incident light with the sample and the resulting frequency shifts by the energy of characteristic vibrations of the molecules, it has been one of the most important tools widely used in biochemical and biomedical fields [1-4]. Rapid and accurate identification of Raman spectra is essential in most fields [5, 6]. Advances in data analysis and modeling techniques such as multivariate analysis have enabled extraction of extensive information from Raman spectra to identify chemical structure and composition of complex systems such as biological materials [7]. Amino acids are basic building blocks of proteins, experimental study and theoretical simulations have been extensively carried out on the Raman spectra of amino acids during the past several decades. Vibrational spectra of amino acids forming the protein have strong dependence on the configuration and interaction with its environment therefore can provide useful information on the conformation and property of the protein or peptide. The purpose of our simulation is to use available computation tools to understand the vibrational properties of the important chemical bonds such as N-H, C-H and benzyl rings in the various environments of different amino acids and peptides or proteins, with the goal to use these information to assist in identifying various relevant compounds by Raman spectroscopy.

In this study, we start out to simulate the Raman spectra of three amino acids: Glycine, Tyrosine and Phenylalanine. The Glycine is the smallest and has the simplest structure among amino acids, Tyrosine and Phenylalanine have benzyl ring structures in the side chain. Aside from the structural features, these species also have biological and physiological significances. For example, Glycine is the main amino acid of animals. It is synthesized by the interaction between serine, threonine, choline and hydroxyproline mainly involved in the liver and kidney organs. It plays an important role in metabolic regulation, antioxidant response and neurological function. Glycine can also be used for the biosynthesis of glutathione, heme, creatine, nucleic acid and uric acid. Glycine is an inhibitory neurotransmitter found in the nervous system, may indirectly help to reduce sleepiness and fatigue caused by accidental sleep deprivation by regulating certain neuropeptides on the rat suprachiasmatic nucleus (SCN) [8, 9]. In addition, glycine is an important component of bile acids secreted into the small intestine and is necessary for digesting dietary fat and absorbing long-chain fatty acids [9]. Tyrosine is important in preventing acute effects of cognitive impairment of the body's stress response [10, 11]. Phenylalanine is an essential amino acid with antidepressant and analgesic effects (natural pain killers) [12]. Phenylalanine is metabolized by tyrosine into acetic acid and fumaric acid. The tyrosine metabolite dopamine is converted into the neurotransmitter adrenaline and norepinephrine. Lack of the neurotransmitter dopamine is a cause of Parkinson's disease and schizophrenia and albinism [13].

*jren@desu.edu; phone 1 302 857-6809; fax 1 302 857-7482; desu.edu

Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics, edited by Gerard L. Coté, Proc. of SPIE
Vol. 10885, 108850O · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2509883

To simulate the vibrational behavior of these amino acids, we used two computational package: Gaussian and GAMESS (General Atomic and Molecular Electronic Structure System). We run these codes on our CPU-based high performance cluster (HPC) as described previously [14]. At this stage, we focus more on qualitative features of calculation results rather than the quantitative accuracy, and compare the results for the three amino acids under the same basis set and algorithm. From our simulation, characteristic region of C-H, N-H, O-H, and benzyl ring can be recognized in the spectrum. In the future, we will apply more complex approximation method and basis set with more kinds of amino acids spectra and compare them with experimental data, so as to analyze and identify Raman spectra more accurately.

2. METHODS

2.1 Energy Minimization (Molecular Geometry Optimization)

In molecular mechanics, the potential energy of various possible conformations can be calculated according to the suitable force field. The conformation with the lowest potential energy is the most stable, such conformation is called geometric optimization conformation. The process of finding the lowest potential energy is called energy minimization [15]. Before each computation, the molecular structure needs to be geometrically optimized to reduce the probability of local minimum during GAMESS and Gaussian computation process. We apply Steepest Descents (SD) method with larger range of the amplitude for optimize in geometry optimization conformation. For force field, we employ the Merck Molecular Force Field (MMFF) 94. This force field defines very complete types of atoms [16], and applicable to the monomolecular system of amino acids we studied.

2.2 Simulation details

2.2.1 Gaussian

Gaussian is a general-purpose software commonly used in quantum computational chemistry, which is good at handling the reaction mechanism, transition state and molecular properties [17]. Our simulation used Density Functional Theory (DFT) by Becke, 3-parameter, Lee-Yang-Parr (B3LYP) approximation method in 6-31G basis set.

2.2.2 General Atomic and Molecular Electronic Structure System（GAMESS)

GAMESS(USA) is another program for ab initio molecular quantum chemistry, it can perform several general computational chemical computations to compute self-consistent field (SCF) molecular wavefunctions, including Hartree-Fock, Density functional theory (DFT), Generalized Valence Bond (GVB) , and Multi-Configuration Self-Consistent Field (MCSCF) [18-20]. For our study, we started from the simplest of Hartree-Fock (HF) approximation method in 3-21G basis set. GAMESS calculation provides another way to verify and compare the results from Gaussian.

2.3 Basis Set

The basis group in quantum chemistry is used to describe the wave functions of a system [21]. The split-valence basis set generated from John Pople's group, typical notations are X-YZG, which means each inner shell orbital is represented by a total of X gaussian functions, and each valence electron orbital is represented by Y gaussian functions and Z gaussian functions [22]. In most simulations, we can switch between different basis sets at will. For us, we use 6-31G basis set for DFT, and 3-21G basis set for HF.

3. RESULTS

Amino acids are indispensable substance for protein. They are composed by C, H, O, N elements. Each of them has an amino group, a carboxylic acid, and a side chain with special properties. All of them have N-H bond and O-H bond, and some of them have a benzyl ring. We summarized the Raman spectrum of several simulated of amino acids and found that the Raman Spectra of N-H bond, O-H bond and benzyl ring have obvious characteristic peaks.

3.1 Glycine

Glycine is the simplest of amino acids, with only one hydrogen as the side chain, aside from the amino and carboxylic acids that are present in all amino acids. Figure 1 shows Glycine molecule in 2-dimensional structure formula and 3-dimensional molecular model.

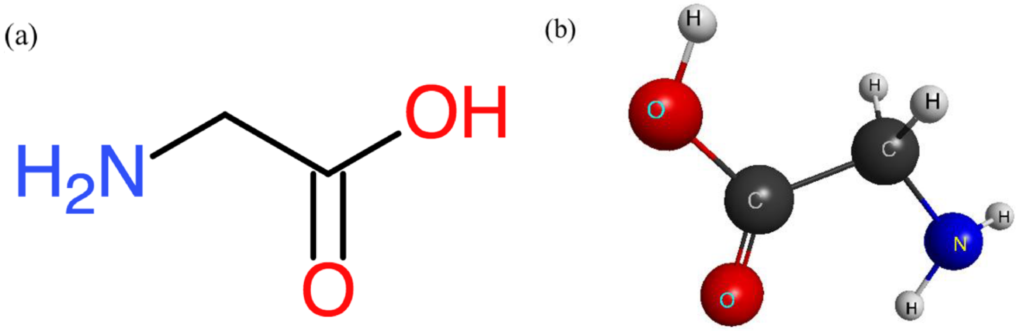

Figure1. Glycine is the simplest amino acid, aside from the amino and carboxylic acids common to amino acids, its side chain only contains an H atom. These are the Glycine's molecular structures in (a) 2-D and (b) 3-D.

The geometrical parameters of Glycine after the geometric optimization using Gaussian and GAMESS are shown in Table 1. In the GAMESS calculation, the H-N-H and H-C-H bonds have asymmetrical bond length for the two arms, therefore labelled as "H1" and "H2".

Table 1: The geometrical parameters of Glycine based on the Gaussian (B3LYP/6-31G) and GAMESS (HF/3-21G) computation. Distances (r) unit is [pm], angles (α) and torsion angles (β) units are [degree].

	Gaussian: B3LYP/6-31G	GAMESS: HF/3-21G(H1)	
		H1	H2
r (N-H)	101.3	100.2	100.0
r (C-N)	144.5	144.8	
r (C-H)	109.9	108.6	109.0
r (C-C)	153.5	151.8	
r (C-O)	138.8	135.7	
r (O-H)	98.0	96.3	
r (C=O)	122.4	119.7	
α (H-N-H)	111.1	115	
α (H-N-C)	114.2	113.5	116.4
α (N-C-C)	115.2	109.5	
α (H-C-H)	106.2	107.2	
α (H-C-C)	108.2	106.7	107.8
α (C-C-O)	115.5	114.9	
α (C-O-H)	113.2	115.1	
α (C-C=O)	124.7	124.4	
β (H-N-C-C)	64.7	10.21	147
β (N-C-C-O)	180	174.3	
β (C-C-O-H)	0	-3.3	

3.1.1 Gaussian

Using B3LYP approximation method with 6-31G basis set in the DFT, we calculated the Raman spectrum of Glycine. In order to make the simulated results more consistent with the experimental results and compensate for the anharmonic effect, we used the corresponding harmonic frequency scale factor of B3LYP/3-21G to correct the simulated Raman spectrum [23].

The calculated Raman spectrum of Glycine is shown in Figure 2. The full range spectrum from 0 cm^{-1} to 4000 cm^{-1} as shown in graph (a), includes the so-called fingerprint region (400 cm^{-1}-1800 cm^{-1}) where majority of the Raman bands appear, and the high-wavenumber region (2400cm^{-1}-3800cm^{-1}) which represents the characteristic peaks region [24] as expanded in graph (b).

For Glycine, high-wavenumber region contains the CH_2 bonds symmetric stretching (SS) and anti-symmetric stretching (AS) of CH_2, SS and AS of NH_2, and the O-H bond stretching (S) modes. Positions of these peaks are 3002 cm^{-1}, 3044 cm^{-1}, 3514 cm^{-1}, 3588 cm^{-1} and 3621 cm^{-1} respectively. Among them, the intensity of symmetric stretching of NH_2 and CH_2 modes are stronger than their antisymmetric counterparts, which conforms to the rule that the molecular vibration modes affects the strength of Raman spectrum [25].

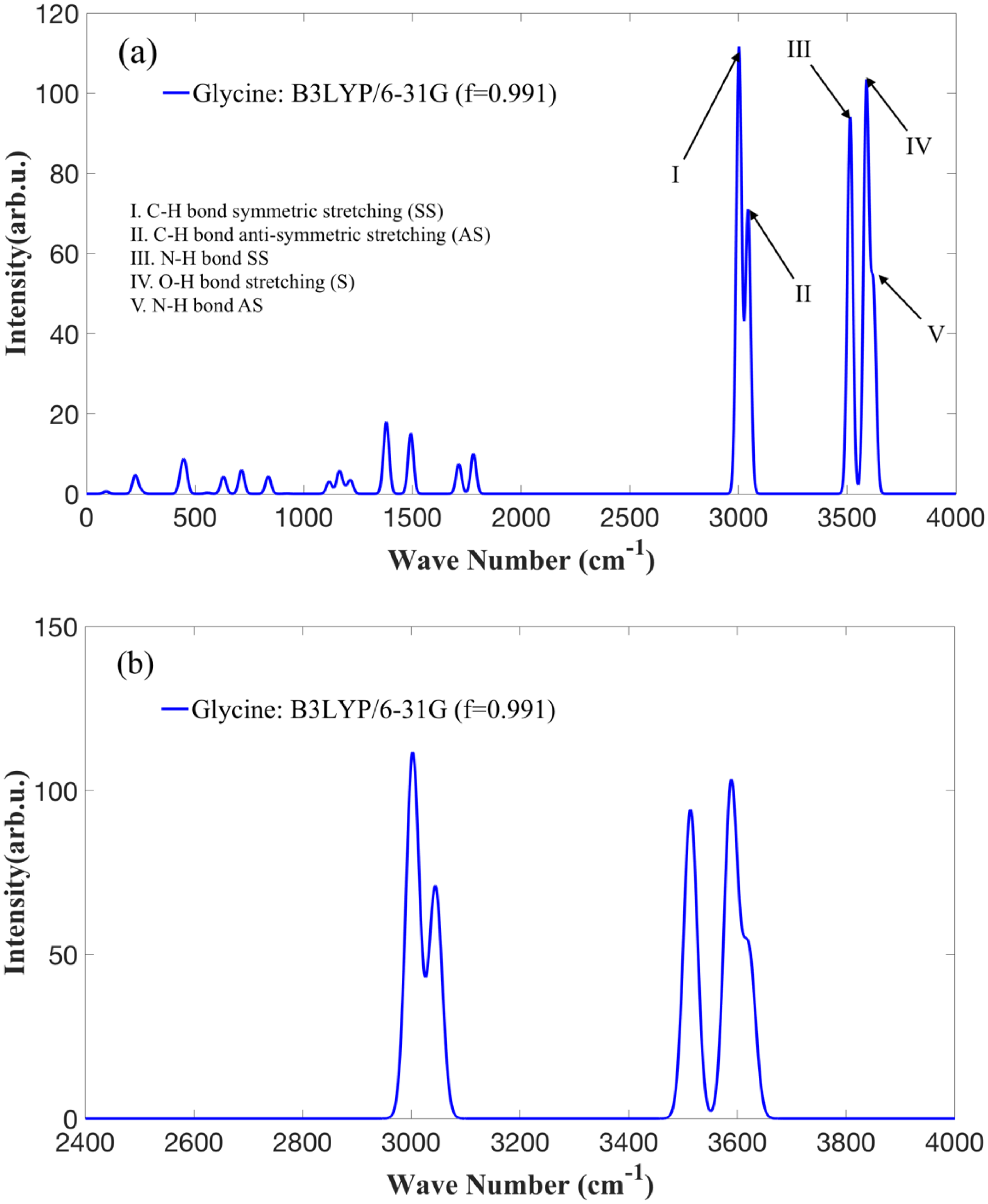

Figure 2. Raman spectrum produced from Gaussian package by executing B3LYP/3-21G method with scale factor 0.991. H-C-H bonds symmetric stretching (SS) and antisymmetric stretching (AS), H-N-H bonds SS and AS, and O-H bond stretching (S) appeared in high wavenumber segment. (a) Whole spectrum, which includes the fingerprint region (400 cm^{-1}-1800 cm^{-1}) and the high-wavenumber region (2400 cm^{-1}-4000 cm^{-1}) expanded in (b).

3.1.2 GAMESS

GAMESS is the second simulation method adopted by us. The result of Raman spectrum is calculated by using GAMESS's special codes for Glycine molecule after optimization. MMFF94 force field with SD method was used for molecular structural optimization. Computing applications HF approximation method in 3-21G basis set. For HF/3-21G method, the scale factor 0.932 is supplementary. Figure 3 is the Raman spectrum of Glycine generated through GAMESS computation by applying HF/3-21G method with its scale factor 0.932 [23].

The results from GAMESS, there are also five characteristic peaks in the high-wavenumber section, namely C-H bonds SS (2938 cm^{-1}) and AS (2970 cm^{-1}), N-H bonds SS (3460 cm^{-1}) and AS (3562 cm^{-1}), and O-H bond S (3666 cm^{-1}). By comparing and analyzing Figure 2 (b) and Figure 3, it can be shown that all characteristic peaks are basically consistent. Due to the differences caused by the different computing methods with different normalization, the entire characteristic peaks area in GAMESS's Raman spectrum have a slight deviation to the short-wave direction than that in Gaussian.

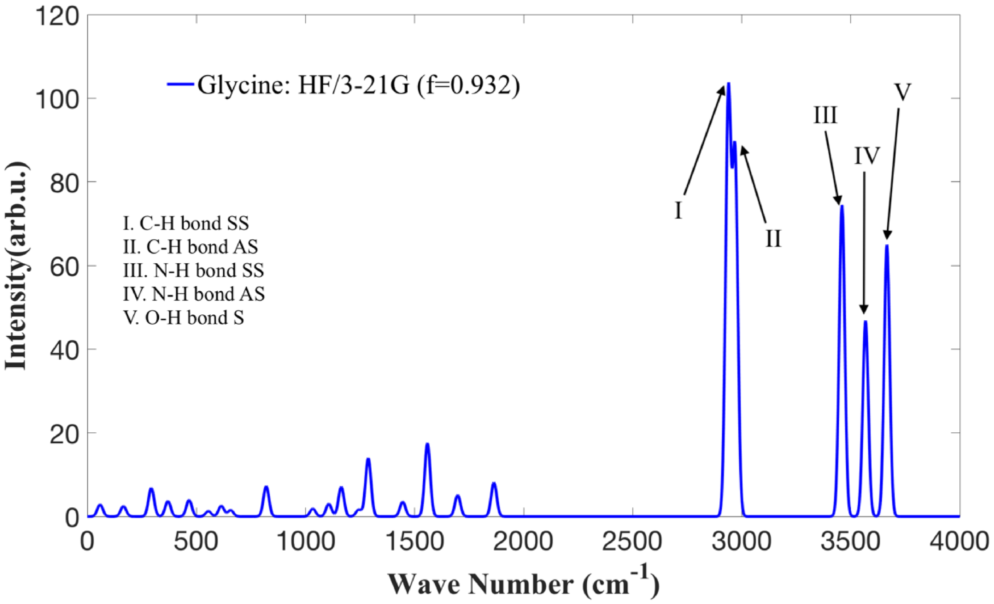

Figure 3. Raman spectrum generated by GAMESS package using HF/3-21G method with scale factor 0.932. Results show that there are C-H bonds SS and AS, N-H bonds SS and AS, and O-H S in the high-wavenumber region, consistent with results from Gaussian.

Our results from two different ways are comparable with the experimental data of C-H bonds, N-H bonds and O-H bond in different solutions [26], showing qualitative consistency.

3.2 Benzyl ring

Figure 4 illustrates the tree-dimensional molecular structure of Tyrosine and Phenylalanine. Both of these structures possess a benzyl ring in addition to the amine and carboxylic acid.

We simulated Tyrosine and Phenylalanine by B3LYP/6-31G in Gaussian. As Figure 5 shows, the Raman spectra of Tyrosine and Phenylalanine are compared with Glycine, respectively. After using harmonic frequency scale (scale factor is 0.991 [23] for B3LYP/3-21G method), we get results from Tyrosine and Phenylalanine, Table 2 lists the vibration of Tyrosine and Phenylalanine in the high wavenumber region.

Comparing the spectra of Tyr, Phe and Gly, there exist extra peaks between the H-C-H and H-N-H stretching bands, i.e. in the region of 3080-3220 cm^{-1}. We also calculated the spectrum for benzene as shown in Fig 6 to compare. We can conclude that this band in region of 3080-3220 cm^{-1} is the characteristic area of the benzyl ring, which is consistent with experimental data of liquid benzyl ring's characteristic peaks at 3075 cm^{-1} of polarized and 3079 cm^{-1} of depolarized [27].

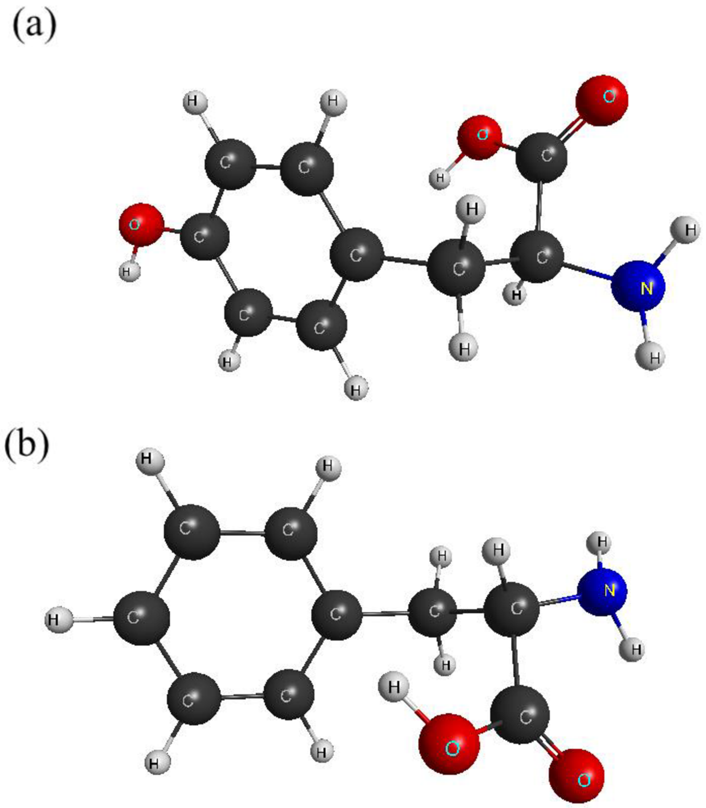

Figure 4. 3-D molecular structures for the only two amino acids with a benzyl ring, (a) Tyrosine ($C_9H_{11}NO_3$) and (b) Phenylalanine ($C_9H_{11}NO_2$).

Table 2. Simulated by Gaussian-B3LYP/6-31G method with scale factor 0.991[23], results about C-H, N-H, O-H bonds and benzyl ring's characteristic peaks in high-wavenumber part of Tyrosine and Phenylalanine.

Vibration mode / Amino Acids	H-C-H	H-N-H	O-H	Benzyl ring
Tyrosine	3014.434 (SS) 3028.806 (AS)	3478.642 (SS) 3581.510 (AS)	3538.684 (S) 3634.145 (S)	3080.061 (S) 3154.209 (S) 3172.681 (S) 3175.151 (S) 3212.065 (S)
Phenylalanine	2930.493 (S) 3017.801 (SS) 3075.402 (AS)	3580.678 (SS) 3717.913 (AS)	3559.502 (S)	3156.410 (S) 3116.145 (S) 3175.415 (S) 3185.078 (S) 3198.752 (S)
Unit: cm^{-1}; Scale factor f=0.991; SS-Symmetric Stretching; AS-Anti-symmetric Stretching; S-Stretching.				

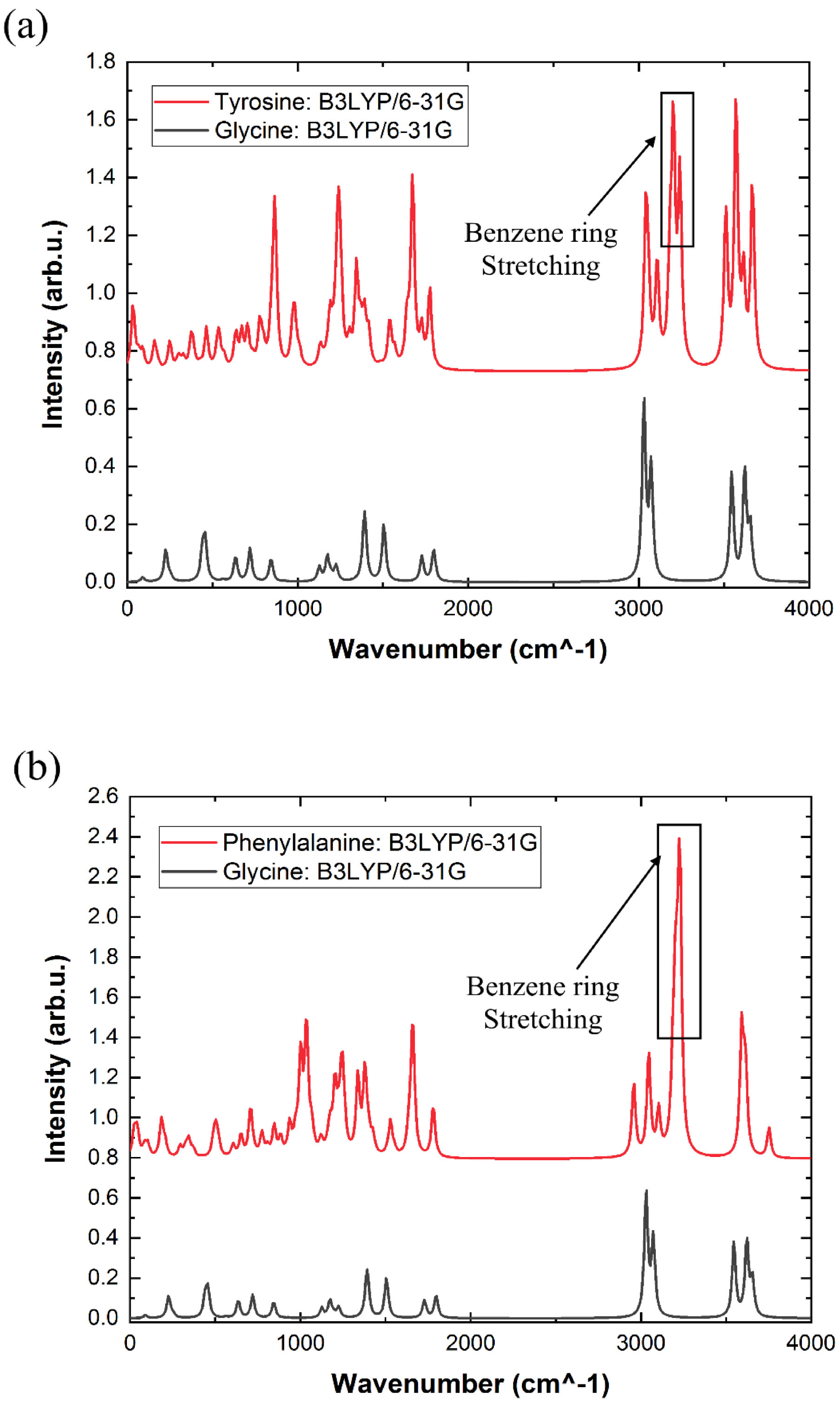

Figure 5. Simulated Tyrosine and Phenylalanine using Gaussian package through B3LYP/6-31G method. Through the contrast of (a) Tyrosine with Glycine, and (b) Phenylalanine with Glycine, clearly to see the benzyl ring stretching area is around 3080 cm^{-1} to 3220 cm^{-1}.

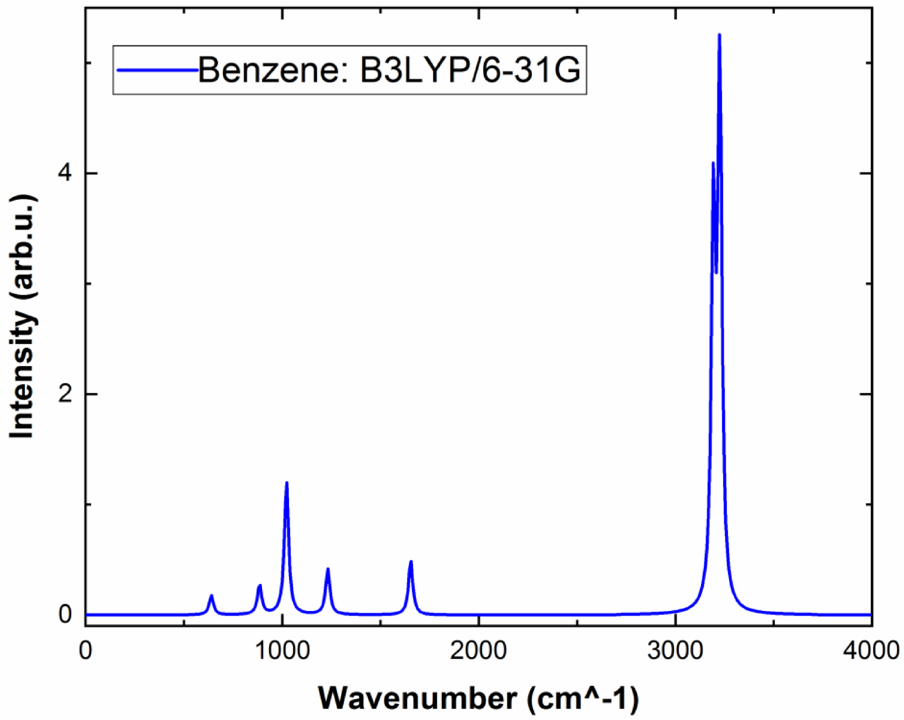

Figure 6. Simulated Benzene using Gaussian package with B3LYP/6-31G method and basis set.

4. DISCUSSION

The experimental data for Raman spectrum of free gas-phase amino acid is very rare due to its very low vapor pressure [28]. We compared our calculation with published work of simulation such as the work in ref [28] for Glycine. While the positions of the vibrational modes have qualitatively similar patterns, there is discrepancy in the details of optimized geometry and the position of vibrational lines. In particular, our optimized geometry with Gaussian B3LYP is close to the ttc in [28] rather than the ttt conformation, which is widely accepted as the most stable conformation for Glycine [29]. As pointed out in ref [28], choice of larger basis set (such as aug-cc-pVTZ used in [28]) would enable B3LYP to localize all stationary points of glycine potential energy surface.

As mentioned in [30], the finger print region of Raman spectra is in general very complex and difficult to use directly for structural determination but are valuable for confirmation. Whereas high wavenumber region of the Raman spectrum can provide valuable information on structure properties, for example, work in [31] studied the dependence of the Glycine CH_2 stretching frequencies on conformation, ionization state and hydrogen bonding.

In our preliminary calculation, we have simulated the characteristic lines of H-C-H, H-N-H, O-H bands in the high wavenumber region for three amino acids, and the benzyl's characteristic peaks in the range between 3080 cm^{-1} and 3220 cm^{-1}. In the two amino acids simulated, the Raman shift changes due to the different substituents of phenyl, indicating the impact of different substituents in phenyl on the Raman spectrum. More detailed study will be needed to confirm the behavior of these bonds under different conformation and environment of amino acids.

5. SUMMARY

In our study, Glycine, Tyrosine and Phenylalanine were calculated on HPC by using the DFT/B3LYP method in Gaussian with 6-31 basis set, and the HF method with 3-21G basis set in GAMESS, showing qualitative consistency. These amino acids carry important roles in the life and physiological processes, and as Raman spectra are sensitively depending on the configuration and conformation arrangements of amino acids, our studies provide valuable insights into identifying protein or peptide and understanding their functions. In future study, we will simulate more amino acids and organic compounds, and adopt more accurate simulation methods and basis set to improve the accuracy of the results.

REFERENCES

[1] Kneipp, K., Kneipp, H., Itzkan, I., Dasari, R. R., and Feld, M. S., "Ultrasensitive Chemical Analysis by Raman Spectroscopy," Chem. Rev. 99, 2957-2975 (1999).

[2] Lawson, E. E., Barry, B. W., Williams, A. C., and Edwards, H. G. M., "Biomedical Applications of Raman Spectroscopy," Journal of Raman Spectroscopy, VOL. 28, 111-117 (1997).

[3] Choo-Smith, L. -P., Edwards, H. G. M., Endtz, H. P., Kros, J. M., and Heule, F. et al., "Medical Applications of Raman Spectroscopy: From Proof of Principle to Clinical Implementation," Biopolymers, 67(1), 1-9 (2002).

[4] Moskovits, M., and Suh, J. S., "Surface Selection Rules for Surface-Enhanced Raman Spectroscopy: Calculation and Application to the Surface-Enhanced Raman Spectrum of Phthalazine on Sliver," J. Phys. Chem. 88,5526-5530 (1984).

[5] Das, R. S., and Agrawal, Y. K, "Raman spectroscopy: Recent advancements, techniques and applications," Vibrational Spectroscopy 57, 163-176 (2011).

[6] Zhang, X. Y., Young, M. A., Lyandres, O., and Van Duyne, R. P., "Rapid Detection of an Anthrax Biomarker by Surface-Enhanced Raman Spectroscopy," J. AM. CHEM. SOC. 127, 4484-4489 (2005).

[7] Hanlon E.B., Manoharan R., Koo T.W., Shafer K.E., Motz J.T., et al., "Prospects for in vivo Raman spectroscopy," Phys Med Biol. 45(2):R1-59 (2000).

[8] Bannai, M., Kawai, N., Ono, K., Nakahara, K., and Murakami, N., "The effects of glycine on subjective daytime performance in partially sleep-restricted healthy volunteers," Front. Neur. 3:61 (2012).

[9] Wang, W., Wu, Z., Dai, Z., Yang, Y., Wang, J. et al., "Glycine metabolism in animals and humans: implications for nutrition and health," Amino acids, 45(3), 463-477 (2013).

[10] McEwen, B. S., "Plasticity of the hippocampus: adaptation to chronic stress and allostatic load," Annals of the New York Academy of Sciences, 933(1), 265-277 (2001)

[11] Mahoney, C. R., Castellani, J., Kramer, F. M., Young, A., and Lieberman, H. R., "Tyrosine supplementation mitigates working memory decrements during cold exposure," Physiology & behavior, 92(4), 575-582 (2007)

[12] National Center for Biotechnology Information. PubChem Compound Database; CID=6140, <https://pubchem.ncbi.nlm.nih.gov/compound/6140> (2019).

[13] Komoda, T. and Matsunaga T., "Metabolic Pathways in the Human Body", Biochemistry for Medical Professionals, 25-63 Academic Press (2015).

[14] Zhao, J., Lu, L., Tyler, Z., Kong, K., Lu, Q. and Ren, J., "Simulated Raman Spectral Analysis of Organic Molecules," Proc. SPIE 10526, Physics and Simulation of Optoelectronic Devices XXVI, 105262I (2018).

[15] Wang, J. M., Cieplak, P., Kollman, P. A., "How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules," Journal of Computational Chemistry, Vol. 21, No. 12, 1049-1074 (2000).

[16] Halgren, T. A., "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94*," Journal of Computational Chemistry, Col. 17, Nos. 5&6, 490-519 (1996).

[17] Valadbeigi, Y., Farrokhpour, H., and Tabrizchi, M., "Theoretical study on the mechanism and kinetics of atmospheric reactions NH2OH + OOH and NH2CH3 + OOH," Physics Letters A 378, 777-784 (2014).

[18] Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S. et al., "General atomic and molecular electronic structure system," Journal of computational chemistry, *14*(11), 1347-1363 (1993)

[19] Summary of GAMESS' Capabilities, <https://www.msg.chem.iastate.edu/gamess/capabilities.html> (2019).

[20] Dykstra, C. E., Frenking, G., Kim, K. S., and Scuseria, G. E., "Advances in electronic structure theory: GAMESS a decade later," Theory and Applications of Computational Chemistry, pp. 1167-1189 (2005).

[21] Schuchardt, K. L., Didier, B. T., Elsethagen, T., Sun, L., Gurumoorthi, V. et al., "Basis Set Exchange: A Community Database for Computational Sciences," J. Chem. Inf. Model. 47, 1045-1052 (2007).

[22] Ditchfield, R., Hehre, W. J., and Pople, J. A., "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules," J. Chem. Phys. 54, 724 (1971).

[23] Alecu, I. M., Zheng, J. J., Zhao Y., and Truhlar, D. G., "Computational Thermochemistry: Scale Factor Database and Scale Factors for Vibrational Frequencies Obtained from Electronic Model," J. Chem. Theory Comput. 6(9), 2872-2887 (2010)

[24] Chau, A. H., Motz, J. T., Gardecki, J. A., Waxman, S., Bouma, B. E. et al., "Fingerprint and high-wavenumber Raman spectroscopy in a human-swine coronary xenograft in vivo," Journal of biomedical optics, 13(4), 040501 (2008).

[25] Lin-Vien, D., Colthup, N. B., Fateley, W. G., and Grasselli, J. G., [The handbook of infrared and Raman characteristic frequencies of organic molecules], Elsevier, Chapter 11 (1991).

[26] Leikin, S., Parsegian, V. A., Yang, W. H., and Walrafen, G. E., "Raman spectral evidence for hydration forces between collagen triple helices," Proceedings of the National Academy of Sciences, 94(21), 11312-11317 (1997).

[27] Chelli, R., Cardini, G., Procacci, P., Righini, R. Califano, S. et al., "Simulated structure, dynamics, and vibrational spectra of liquid phenyl," J. Chem. Phys. 113, 6851 (2000).

[28] Balabin, R.M., "Conformational equilibrium in glycine: Focal-point analysis and ab initio limit," Chem. Phys. Lett. 479, 195-200 (2009).

[29] Ke, H.W., Rao, L., Xu, X. and Yan, Y.J., "Theoretical Study of Glycine Conformers," J. Theoret. Comput. Chem. 7, 889-909 (2008).

[30] Zhu G., Zhu, X., Fan, Q., and Wan, X., "Raman spectra of amino acids and their aqueous solutions," Spectrochimica Acta Part A 78 1187-1195 (2011).

[31] Bykov, S.V., Myshakina, N.S., and Asher, S.A., "Dependence of Glycine CH2 Stretching Frequencies on Conformation, Ionization State, and Hydrogen Bonding," J. Phys. Chem. B, 112, 5803-5812 (2008).