

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Simulated Raman spectral analysis of organic molecules

Jian Zhao, Lu Lu, Zachary Tyler, Kam Kong, Qi Lu, et al.

Jian Zhao, Lu Lu, Zachary Tyler, Kam Kong, Qi Lu, Jun Ren, "Simulated Raman spectral analysis of organic molecules," Proc. SPIE 10526, Physics and Simulation of Optoelectronic Devices XXVI, 105262I (23 February 2018); doi: 10.1117/12.2291176

SPIE.

Event: SPIE OPTO, 2018, San Francisco, California, United States

Simulated Raman Spectral Analysis of Organic Molecules

Jian Zhao^a, Lu Lu^a, Zachary Tyler^a, Kam Kong^a, Qi Lu^a, Jun Ren^{a*}

^aDelaware State Univ., 1200 Dupont Hwy., Dover, DE USA 19901

ABSTRACT

Raman spectroscopy is commonly used in chemistry and biology. As vibrational information is specific to the chemical bonds, Raman spectroscopy provides fingerprints to identify the type of molecules in the sample. In this paper, we simulate the Raman spectrum of organic and inorganic materials by GAMESS and GAUSSIAN on our high-performance cluster. By using MPI (message passing interface), we are able to run the codes in parallel. From our simulations, we will set up a database that allows search algorithms to quickly identify N-H and O-H bonds in different materials.

Keywords: Raman Spectroscopy, Molecular Dynamics, Computational Spectroscopy

1. INTRODUCTION

The invention of laser in the 1960s solved the main difficulty in Raman spectroscopy of low sensitivity, and made Raman spectroscopy a practical and common tool in many chemistry and biology studies [1-3]. As vibrational information is specific to the chemical bonds, Raman spectroscopy provides fingerprints to identify the type of molecules in the sample. Nonetheless after nearly 50 years of development, the sensitivity of Raman spectroscopy is still a major factor that limits various potential applications [4]. Increasing laser power may increase the generated Raman signal, yet there is an upper limit in the sensitivity enhancement factor beyond which the molecular bonds and the sample could be damaged [5, 6]. For many applications, fast and accurate identification of a weak Raman signal from its background is a prerequisite for the technique [7, 8].

In the meantime, computation methods and capacities have advanced rapidly in recent decades and play more and more important roles in many areas of science, technology, engineering and mathematics (STEM) research. Supercomputing now has become an essential tool for scientific research, because a significant and growing portion of STEM problems can only be accessed through computational methods, and very often insights and understandings can merely be gained through computed representations and visualizations. Simulations of Raman spectroscopy offer the possibility of computing and visualizing the impacts of molecular structural modifications on Raman spectrum. Furthermore, successful comparisons between simulated and experimental Raman spectra bring coherent directions in identifying and renovating methods to continuously improving the sensitivity and operability of Raman spectroscopy in different STEM research.

In this study, we simulate the Raman Spectrum of organic and inorganic materials by General Atomic and Molecular Electronic Structure System (GAMESS) and Gaussian, two computational codes that perform several general chemistry calculations. We run these codes on our CPU-based high performance cluster (HPC). Through the message passing interface (MPI), a standardized and portable message-passing system which can make the codes run in parallel, we are able to decrease the amount of time for computation and increase the sizes and capacities of systems simulated by the codes. From our simulations, a database will be set up that allows our search and machine learning algorithms to quickly identify N-H and O-H bonds in different materials. Through our studies, our goal is to analyze and identify the spectra of organic matter compositions from meteorites, and compare these spectra with terrestrial biologically-produced amino acids and residues.

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

2. METHODOLOGY

2.1 Geometry Optimization

Our simulations begin with geometry optimization, a process that finds the balance position for atoms in the molecule with minimum potential energy [11-12]. Typical results of our optimized bond lengths and bond angles for Methane (CH_4) using Hartree-Fock theory [9] with 3-21G, ACCD, and Sadlef basis sets are listed in Table I. Good consistencies can be found between our results and the result from NIST Computational Chemistry Comparison and Benchmark Database [10].

Table I: Optimized geometrical parameters of methane (CH_4) using Hartree-Fock theory [9] with 3-21G, ACCD, and Sadlef basis sets; the optimized geometrical parameters are compared with experimental data [10]

	3-21G	ACCD	Sadlef	Expt.
Bond lengths (Å)				
C1-H2	1.0828455	1.0893858	1.0900152	1.092±0.014
C1-H3	1.0828455	1.0893858	1.0900152	
C1-H4	1.0828455	1.0893858	1.0900152	
C1-H5	1.0828455	1.0893858	1.0900152	
Bond angles (°)				
H2-C1-H3	109.4734021	109.4727727	109.4728897	110.62±4.47
H2-C1-H4	109.4701303	109.4704498	109.470386	
H2-C1-H5	109.4701303	109.4704498	109.470386	
H3-C1-H4	109.4701303	109.4704498	109.470386	
H3-C1-H5	109.4701303	109.4704498	109.470386	
H4-C1-H5	109.4734021	109.4727727	109.4728897	

2.2 Basis Set

A basis set is a set of mathematical representations (also called “basis functions”) that consists of combinations of functions to characterize molecular orbitals. Most of these functions are typically atomic orbitals centered on atoms but theoretically can be any function [13-14]. In our simulations, we compared results from the two ab-initio approximation methods, including the Density Functional Theory (DFT) and the Hartree-Fock Theory (HF). We tested 3-21G, ACCD, and Sadlej basis sets [15-17] with HF Theory and B3LYP and APFD basis sets [18-20] with DFT.

3. RESULTS AND DISCUSSION

3.1 Methane (CH_4)

Our first simulated sample is Methane (CH_4), the simplest alkane and a main constituent of natural gas. The Carbon hydrogen (C-H) stretching bonds in CH_4 are sensitive to their local chemical environment, and hence can be utilized as a useful marker to study hydrocarbon at liquid/gas interfaces [21, 22]. Methane is formed by one C atom and four H

atoms. There are total of nine different normal vibration modes, as illustrated in Figure 1. The nine vibrational modes shown in Fig. 1 are one symmetric C-H stretching (ν_1), two degenerate bendings (ν_2), three degenerate antisymmetric C-H stretchings (ν_3), and three degenerate bending (ν_4).

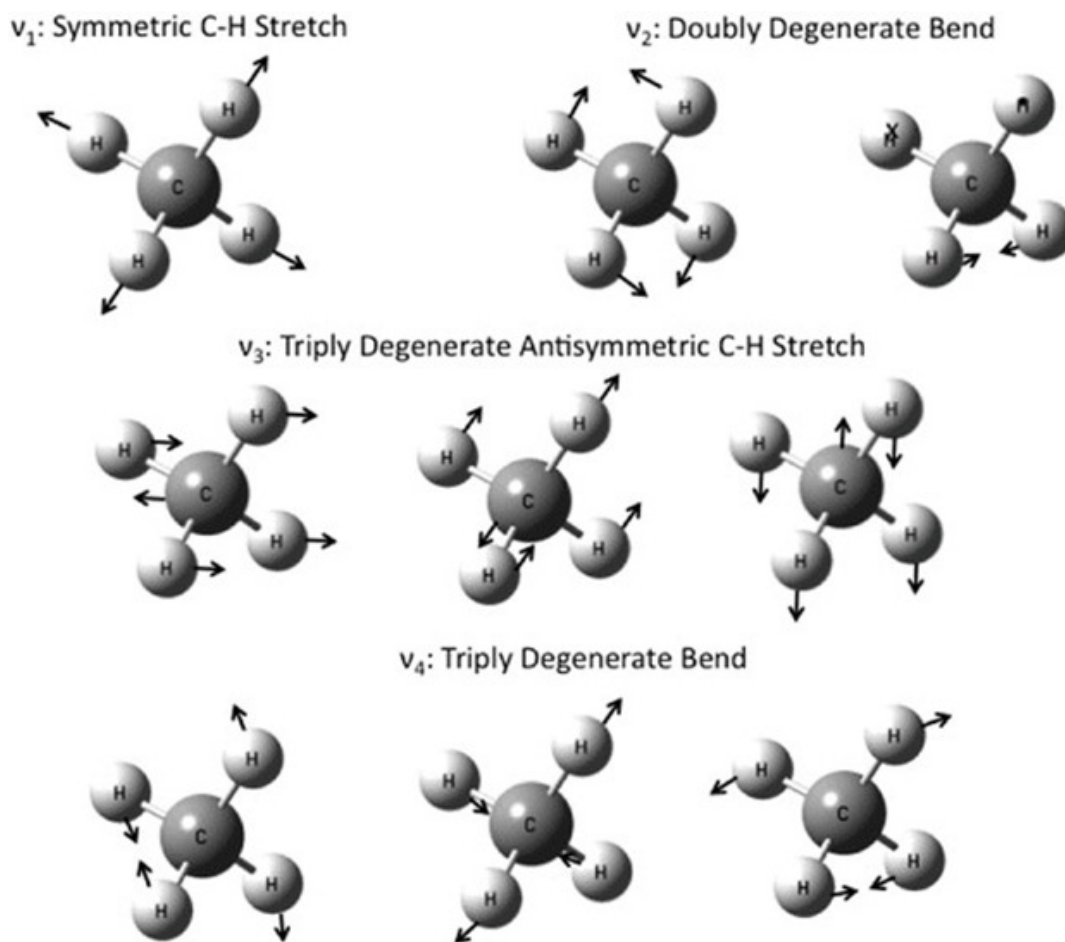


Figure 1. Illustration of the nine normal vibrational modes in CH_4 molecule.

In the HF method, deviation of the system from being a harmonic or double harmonic oscillator was not taken into account in the vibrational computations [23]. Hence the calculated frequency for each basis set was overvalued when comparing with experimental data. To correct this anharmonic effect, we scaled down our calculated frequencies with a scale factor f , where

$$f = \frac{\sum (\gamma_i \times \omega_i)}{\sum \omega_i^2} \quad (1)$$

Here γ_i is the measured vibrational frequencies [24, 25], and ω_i is the calculated vibrational frequencies. For 3-21G, ACCD, and Sadlej basis sets, the scale factor f is estimated as 0.9085, 0.9255, and 0.9238, respectively.

Figure 2 shows our simulated Methane Raman spectrum using the GAMESS package. The basis sets employed are 3-21G, ACCD, and Sadlej, respectively. Figure 2 shows that the simulated spectra using the ACCD and the Sadlej basis sets are similar, and comparable with results from the 3-21G basis set at high frequency regions above $2,000 \text{ cm}^{-1}$. Comparing with experimental data [24, 25], we found that results from the ACCD and Sadlej basis sets agree better with measured data [24, 25].

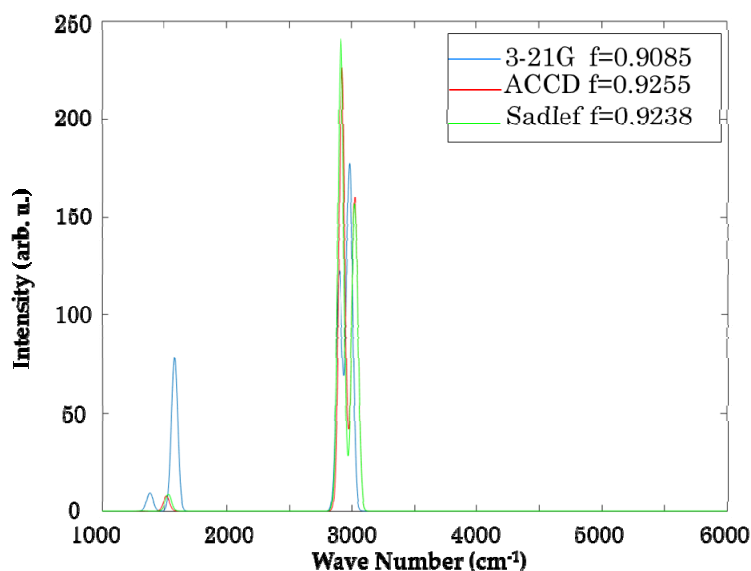


Figure 2. Simulated Methane Raman spectrum using GAMESS package. The basis sets applied are 3-21G, ACCD, and Sadlej in HF theory, respectively.

We also simulated the Methane Raman spectrum using the Gaussian package. Figure 3 shows the result using the 3-21G basis set in the HF Theory. Comparing to the spectrum from 3-21G in Figure 1, we found consistencies at high frequencies, i.e. from the C-H stretching bond, but stronger contributions at lower frequencies, i.e. from the C-H bending bond.

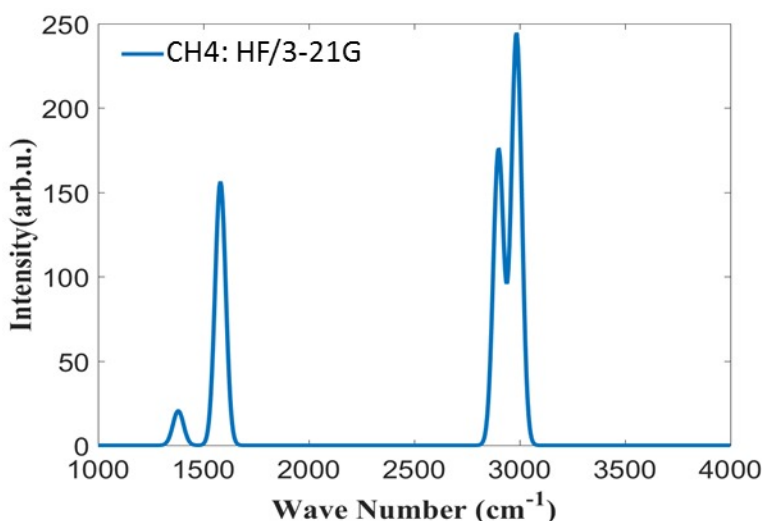


Figure 3. Simulated Methane Raman spectrum using Gaussian package. The basis sets applied is 3-21G.

Results from using the B3LYP and APFD basis sets in the DFT are shown in Figure 4 and Figure 5, respectively. To correct the anharmonicity in the vibrational computation, a scale factor of 0.989 was used in our simulation [26]. Comparing to the results from the 3-21G basis set in the HF Theory, as shown in Figure 2 and Figure 3, the results from the APFD basis set in the DFT agree better with measured data [24, 25]. In general, different from HF Theory which assumes that many-electron wavefunction can be expressed as a single determinant (a Slater determinant), DFT takes into account the electron density and can provide more consistent result with experimentally determined molecular frequencies [27]. Although we found that for Methane, the B3LYP result is less consistent with measured data than the 3-21G basis set in the HF Theory.

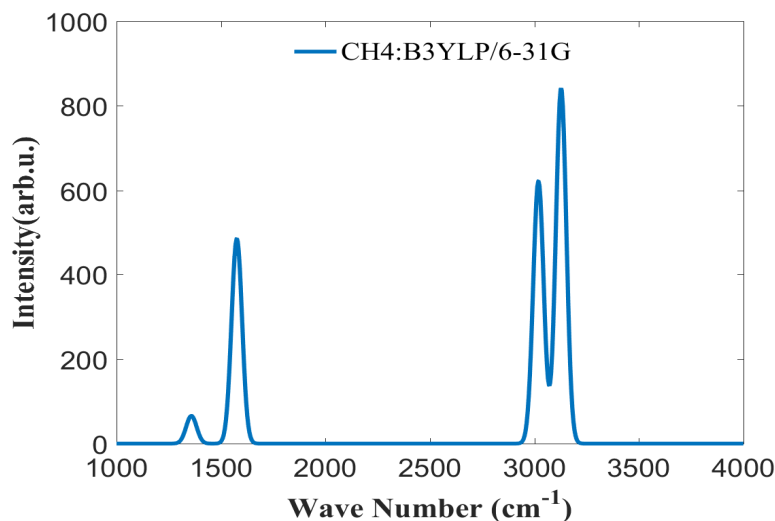


Figure 4. Simulated Methane Raman spectrum using Gaussian package. The basis sets applied is B3LYP/6-31G.

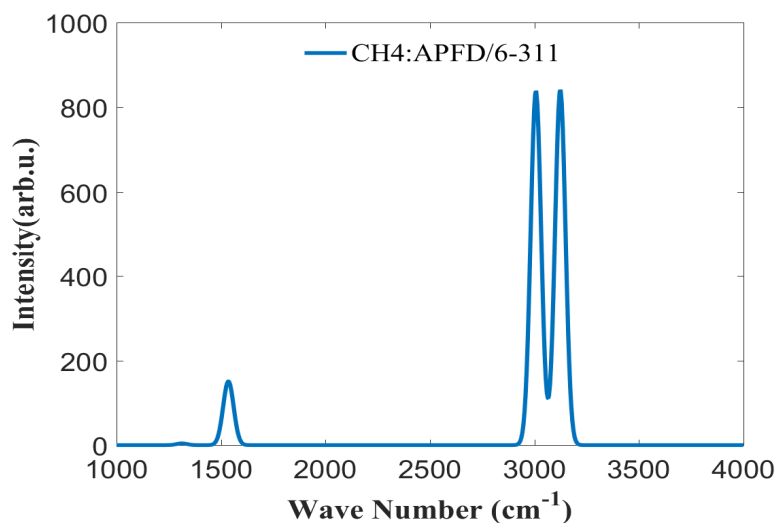


Figure 5. Simulated Methane Raman spectrum using Gaussian package. The basis sets applied is APFD/6-311.

3.2 Water (H₂O)

Our next simulated sample is water (H₂O). We selected H₂O because the intermolecular coupling between H₂O molecules plays an important role in various chemical and biological processes [28, 29], and this unique properties of water can be studied by measuring the Raman spectra in the OH stretching bond.

Water molecule has total of three fundamental vibrational modes, as illustrated in Figure 6, including one symmetric stretching (ν_1), one symmetric bending (ν_2), and one anti-symmetric stretching (ν_3). From our simulations, we compared simulated Raman spectra of water using 3-21G, ACCD, and Sadlej basis sets with HF Theory in GAMESS and 3-21G basis set with HF Theory and B3LYP and APFD basis sets with DFT in Gaussian. We found that the simulated spectrum from the HF/3-21G basis set, shown in Figure 7, produced the best comparison with the experimental data [30, 31].

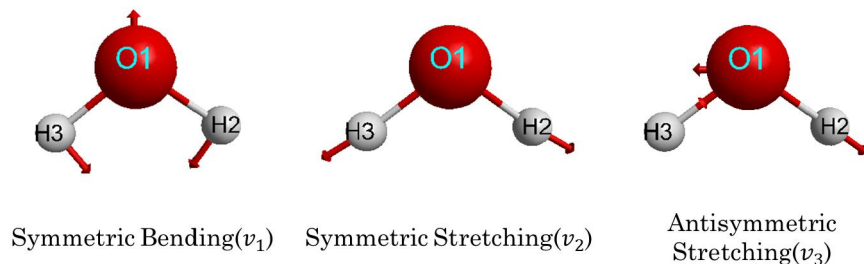


Figure 6. Fundamental vibrational modes of water (H_2O).

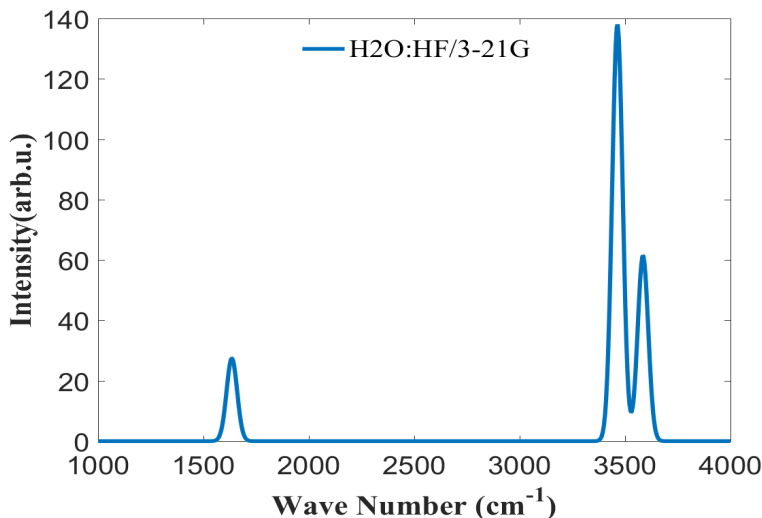


Figure 7. Simulated water Raman spectrum using Gaussian package. The basis sets applied is HF/3-21G.

The experimentally measured O-H stretching band of liquid water [30, 31] covers a range of more than 400 cm^{-1} , much broader than our simulated results on single molecule. This is mainly due to the coupling of H_2O molecules in an aqueous environment through the intermolecular hydrogen bonds [32, 33].

To simulate this effect, we added a water molecule in our setup and ran the geometry optimization process to locate the minimum potential energy point. The intermolecular action between water molecules occurs normally in liquid water, where water molecule couples with other water molecules to form a cluster. We start from water dimer geometry since a water dimer would add reasonable load on the computation time. Additionally, by comparing the simulated Raman spectrum of a single water molecule to that of a water dimer, the effects of molecule-molecule interaction and clustering can be illustrated on Raman spectrum.

Figure 8 shows our simulated water dimer Raman spectrum using the Gaussian package with HF/3-21G basis set. Comparing to the single water molecule spectrum shown in Figure 7, the O-H stretching bond bandwidth in Figure 8 has increased to $\sim 300 \text{ cm}^{-1}$, and the center peak has shifted toward lower frequency. These modifications bring the simulated water dimer Raman spectrums closer to the measured experimental water spectrum [30, 31].

In our next step, we will continue increasing the number of water molecules in our geometry setup and minimize the total potential energy through geometry optimization. The optimized molecule cluster will be employed in our simulations with different HF and DFT basis sets. We will repeat the process until a convergence of the simulated spectrum can be reached.

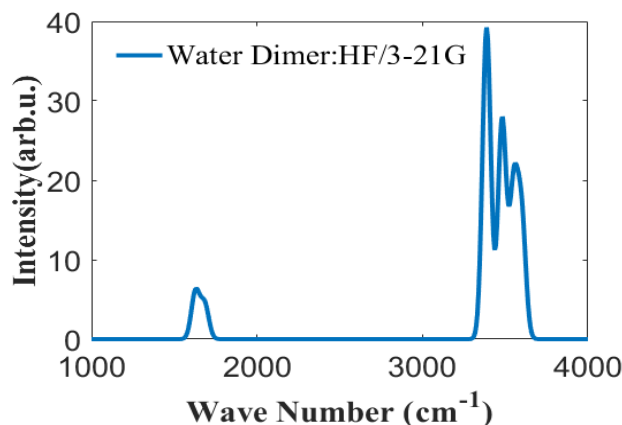


Figure 8. Simulated Water Dimer Raman spectrum using Gaussian package. The basis sets applied is HF/3-21G.

3.3 Glycine ($C_2H_5NO_2$)

Amino acids are the basic “building blocks” of peptides and proteins and play an important role in the physiological processes of all species. To determine whether there are proteins in a material under test, it is necessary to study the Raman spectra of amino acids. Glycine is the simplest amino acid, so we picked Glycine as our first sample. Figure 9 shows a 3-D illustration of the Glycine molecular structure.

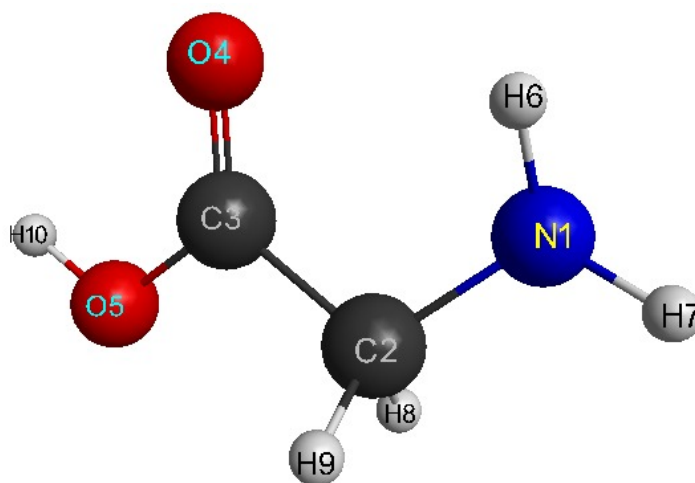


Figure 9. 3-D model illustrates the structure of Glycine ($C_2H_5NO_2$) geometry and structure.

Glycine has 10 atoms, theoretically it has 24 fundamental vibrational modes ($3N-6$, with N as the number of atoms). From published experimental data [34, 35], 13 bands have been located at 497, 602, 697, 893, 1,033, 1,323, 1,410, 1,508, 1,567, 1,667, 2,123, 2,930, and 3,050 cm^{-1} . Furthermore, out of the 13 bands, two bands have been recognized with intense peaks at around 894 and 1,327 cm^{-1} . The 894 cm^{-1} peak is characterized as the NH_2 twisting vibration and the COOH bending vibration. And the other 1,327 cm^{-1} peak is labelled as vibrations from the three functional groups: NH_2 twisting, CH_2 twisting, and COOH bending.

Figure 10 shows our simulated Glycine Raman spectrum using the Gaussian package with DFT/APFD/6-311 basis set. Comparing to the experimental spectrum [34, 35], we found consistencies between our simulation results and the experimental measurements, especially in the spectrum region known as the fingerprint region, where majority of the Raman bands appear. For Glycine, the fingerprint region is between 700 cm^{-1} and 1700 cm^{-1} .

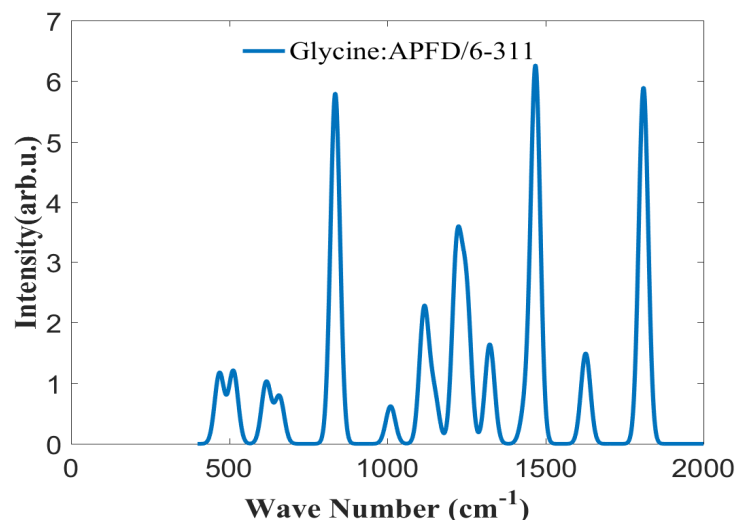


Figure 10. Simulated Glycine Raman spectrum using Gaussian package. The basis sets applied is APFD/6-311

4. SUMMARY

In this work, we simulated Raman spectra for methane (CH_4), water (H_2O), and glycine ($\text{C}_2\text{H}_5\text{NO}_2$), using the DFT and the Hartree-Fock Theory (HF). We tested 3-21G, ACCD, and Sadlej basis sets with HF Theory and B3LYP and APFD basis sets with DFT. Good agreements have been found between our simulation results and the experimental results. Strong bands with N-H and O-H bonds are pinpointed and identified. In the next step, we will expand our current capacity to include more molecules in our simulation, e.g. simulate water molecule cluster and amino acid with larger molecular weight and more complex structures. We will also set up our Raman spectrum database and develop search and machine learning algorithms to quickly identify N-H and O-H bonds in different materials.

REFERENCES

- [1] Schrader, B., [Infrared and Raman Spectroscopy], Schrader, B. ed., VCH Publishers, New York, Chapter 4 (1995).
- [2] Myers, A.B., and Mathies, R.A., [Biological Applications of Raman Spectroscopy], Spiro, T.G. ed., John Wiley and Sons, New York, Vol. II Chapter 1 (1987).
- [3] Morris, M.D., [Applied Laser Spectroscopy], Andrews, D.L. ed., VCH Publishers, New York, Chapter 6 (1992).
- [4] Ruud, K. and Thorvaldsen, A.J., "Theoretical approaches to the calculation of Raman optical activity spectra," *Chirality* 21 Suppl 1, E54–67 (2009).
- [5] Anker, J. N., Hall, W. P., Lyandres, O., Shah, N. C., Zhao, J. et al., "Biosensing with plasmonic nanosensors," *Nat. Mat.* 7, 442-453 (2008)
- [6] Panneerselvam, R., Liu, G. K., Wang, Y. H., Liu, J. Y., Ding, S. Y. et al., "Surface-enhanced Raman spectroscopy: bottlenecks and future directions," *Chem. Comm.* 54, 10-25 (2018).
- [7] Chen, K., Wei, H. Y., Zhang, H. Y., Wu, T., and Li, Y., "A Raman peak recognition method based automated fluorescence subtraction algorithm for retrieval of Raman spectra of highly fluorescent samples," *Anal. Methods* 7, 2770-2778 (2015).
- [8] Austin, L. A., Osseiran, S., and Evans, C. L., "Raman technologies in cancer diagnostics," *Analyst* 141, 476-503 (2016).
- [9] Fischer, C. F., "General Hartree-Fock program," *Computer Physics Communication.* 43, 355–365 (1987).

- [10] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. Release 18, October 2016, Editor: Russell D. Johnson III <<http://cccbdb.nist.gov/>>
- [11] Peng, C., Ayala, P. Y., and Schlegel, H. B., "Using Redundant Internal Coordinates to Optimize Equilibrium Geometries and Transition States," *J. Comput. Chem.* 17, 49-56 (1996).
- [12] Niu, S. Q. and Hall, M. B., "Theoretical studies on reactions of transition-metal complexes," *Chem. Rev.* 100, 353-405 (2000).
- [13] Jensen, F., "Atomic orbital basis sets," *WIREs Comput. Mol. Sci.* 3, 273-295 (2013).
- [14] Jensen, G., "Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions," *WIREs Comput. Mol. Sci.* 4, 127-144 (2014).
- [15] Dunning, T. H., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.* 90, 1007-1023 (1989).
- [16] Kendall, R. A., Dunning, T. H., and Harrison, R. J., "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.* 96, 6796-806 (1992).
- [17] Sadlej, A. J., "Medium-size polarized basis sets for high-level correlated calculations of molecular electric properties," *Collection of Czechoslovak Chemical Communications* 53, 1995-2016 (1988).
- [18] Lee, C., Yang, W., and Parr, R. G., "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B* 37, 785-789 (1988)
- [19] Miehlich, B., Savin, A., Stoll, H., and Preuss, H., "Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr," *Chem. Phys. Lett.* 157, 200-206 (1989)
- [20] Austin, A. J., Petersson, G., Frisch, M. J., Dobek, F. J., Scalmani, G. et al., "A Density Functional with Spherical Atom Dispersion Terms," *Chem. Theory Comput.*, 8, 4989-5007 (2012)
- [21] Sebek, J., Pele, L., Potma, E. O., and Gerber, R. B., "Raman spectra of long chain hydrocarbons: anharmonic calculations, experiment and implications for imaging of biomembranes," *Phys. Chem. Chem. Phys.* 13, 12724-12733 (2011).
- [22] Roy, T. K., and Gerber, R. B., "Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications," *Phys. Chem. Chem. Phys.* 15, 9468-9492 (2013).
- [23] Lee, T. J., Martin, J. M., and Taylor, P. R., "An accurate ab initio quartic force field and vibrational frequencies for CH₄ and isotopomers," *J. Chem. Phys.* 102, 254-261 (1995).
- [24] Jourdanneau, E., Chaussard, F., Saint-Loup, R., Gabard, T., and Berger, H., "The methane Raman spectrum from 1200 to 5500 cm⁻¹: A first step toward temperature diagnostic using methane as a probe molecule in combustion systems," *J. Mol. Spectrosc.* 233, 219-230 (2005).
- [25] Magnotti, G., Kc, U., Varghese, P. L., and Barlow, R. S., "Raman spectra of methane, ethylene, ethane, dimethyl ether, formaldehyde and propane for combustion applications," *J. Quant. Spectrosc. Radiat. Transfer* 163, 80-101 (2015).
- [26] Bauschlicher, C., "A comparison of the accuracy of different functionals," *Chem. Phys. Lett.* 246, 40-44 (1995).
- [27] Kent, P. R. C., "Density Functional Theory," < <http://web.ornl.gov/~kentpr/thesis/pkthnode15.html> > (August 1999).
- [28] Ahmed, M., Singh, A., and Mondal, J. A., "Hydrogen-bonding and vibrational coupling of water in a hydrophobic hydration shell as observed by Raman-MCR and isotopic dilution spectroscopy," *Phys. Chem. Chem. Phys.* 4, 2767-2775 (2016).
- [29] Ghosh, N., Singh, A. K., Mondal, J. A., "pH Dependence of Interfacial Water in the Presence of Amino Acid Side Chains Revealed by Heterodyne-Detected Sum-Frequency Generation Spectroscopy," *J. Phys. Chem. C* 120, 23596-23603 (2016).
- [30] Carey, D. M. and Korenowski, G. M., "Measurement of the Raman Spectrum of liquid water," *J. Phys. Chem.* 108, 2669-2675 (1998).
- [31] Walrafen, G. E., "RAMAN SPECTRAL STUDIES OF WATER STRUCTURE," *J. Chem. Phys.* 40, 3249-3256 (1964).
- [32] Bosma, W. B., Fried, L. E. and Mukamel, S., "Simulation of the intermolecular vibrational spectra of liquid water and water clusters," *J. Phys. Chem.* 98, 4413-4421 (1993).
- [33] Auer, B. M., and Skinner, J. L., "IR and Raman spectra of liquid water: Theory and interpretation," *J. Phys. Chem.* 128, 224511 (2008).
- [34] Kumar, S., Rai, A. K., Singh, V. B., and Rai, S. B., "Vibrational spectrum of glycine molecule," *Spectrochimica Acta Part A* 61, 2741-2746 (2005).
- [35] 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).