Modeling of spectral signatures using ab-initio calculations

Jeffrey W. Mirick*

ABSTRACT

Advances in quantum chemistry in the 1990's have resulted in improvements in the modeling of molecules and in calculating energies, transition structures, and modes of vibration, resulting in the capability to approximate vapor-phase infrared spectral signatures of chemicals using ab-inito calculations. These approaches provide a means of deriving signatures of vapor-phase chemicals, or gases, whose spectra may be too difficult to measure through traditional means. The purpose of this paper will be to review the limitations and accuracies in the modeling of spectral signatures using different methods and basis sets. Comparisons will be made between calculated signatures and laboratory measured signatures.

Keywords: ab-initio, spectral signature, modeling, molecules, LWIR, sarin, isopropyl alcohol, ethyl alcohol

1. INTRODUCTION

Why model chemicals? The main reason for molecular modeling is to approximate chemical properties and to predict chemical behaviors, or characteristics, prior to performing expensive or dangerous measurements. Some of the properties that can be approximated include transitions structures, energy barriers, modes of vibration, molecular structures, reaction paths between two chemicals, etc. The key word is APPROXIMATION. Approximations are made in every step of the calculations. In many instances the approximated spectrum, which results from a calculation of the vibration mode, is in good agreement with the measured results, however, there are exceptions. In this paper I will investigate the accuracies and limitations in using ab-initio calculations to approximate the spectral signatures of molecules by comparing the results to lab spectra measured by Pacific Northwest National Laboratory (PNNL) [1, 2].

There have been many books written on the subject of molecular modeling.[3, 4, 5, 6] Before the age of computers, the modeling of molecules was done by hand. The calculations were long, complex, and labor-intensive and were limited to atoms and the simplest of diatomic molecules. The beginning of the computer age brought more computational power and with the computational power came more complex algorithms and equations to solve the multi-body problem with molecules. Today, larger and more complex molecular systems are being modeled and chemical properties approximated with greater accuracies than ever before. Computer applications and complex algorithms were developed to take advantage of the increase in computational power. Computer software was developed for the chemists to use in the modeling of molecules. The most common software on the market is "Gaussian" [7], which is the software used in this paper. The General Atomic and Molecular Electronic Structure System [8] (GAMESS, a general ab-initio quantum chemistry package) software package is another type of software used in the modeling of molecular electronic structures and is available over the Internet

2. THEORY

The starting point for ab-initio calculations involving atoms and molecules is the time-independent Schrödinger equation. This equation was developed by Erwin Schrödinger in 1925 and describes the total energy of atomic, or molecular, systems in terms of the nuclear and electronic coordinates. The equation can be written as

$$H^e_{total}\Psi = E\Psi \tag{1}$$

*dr2stp@bluelaze.com; http://www.bluelaze.com; phone 1 703 877-5017

where

$H^e_{total} = electronic\ Hamiltonian$

$E = total\ enerrgy$

$\Psi = wavefunction$

The total energy of a molecular system, or the Hamiltonian, is the sum of the electronic kinetic energy and the various potential energy (V) terms resulting from the interactions between the nuclei and the electrons, the electron-electron interaction, and the nuclear-nuclear interactions. The electronic Hamiltonian can be written as

$$H^e_{total} = T + V = T^e + V_{electron-electron} + V_{electron-nucleus} + V_{nucleus-nucleus}$$

$$= -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \sum_i^n\sum_j^n \frac{e^2}{r_{ij}} - \sum_i^n\sum_\alpha^N \frac{e^2 Z_\alpha}{r_{i\alpha}} + \sum_\alpha^N\sum_\beta^N \frac{e^2 Z_\alpha Z_\beta}{R_{\alpha\beta}} \tag{2}$$

where
$\hbar = planck's\ constant$
$m_e = electron\ \ mass$
$e = electon\ \ charge$
$\nabla_i^2 = Laplace\ \ Operator$
$Z = Atomic\ \ Number$
$r, R = electron\ /\ nuclei\ \ position\ \ vectors\ \ respective\ ly$
$n, N = number\ \ of\ \ electrons\ /\ atoms\ \ respective\ ly$

There are no exact solutions to the time-independent Schrodinger's equation, and approximations must be made to molecular systems. There are three basic approximations made in molecular modeling; The Born Oppenheimer Approximation, the Hartree-Fock (HF), or Main Field approximation, and expressing Molecular Orbitals as Linear Combination of Atomic Orbitals (LCAO).

The Born-Oppenheimer approximation states that the nucleus of an atom remains fixed in space while the fast moving electrons spins around the nucleus. This results in a decoupling of the electronic and nuclei contributions in the Hamiltonian. This approximation can be made due to the fact that the mass of the nucleus is much larger than the mass of an electron. In other words, the nucleus and the electronic energy can be treated separately. The equations can be written as

$$\Psi_{total} = \Psi^e(r:R) \times \Psi^n(R)$$

$$H^e(r:R) \times \Psi^e(r:R) = E^e(R) \times \Psi^e(r:R) \tag{3}$$
$$H^n(R) \times \Psi^n(R) = E^n(R) \times \Psi^n(R)$$

The second approximation is the HF approximation where the correlated motion of electrons is neglected; therefore the energy calculated in the HF approximation is an upper-limit. The exact solution to the non-relativistic time-independent Schrödinger equation will always be less than the HF energy. The best energy that can be achieved using the HF method

is called the HF limit. The Pauli-Exclusion principle forbids more than one electron from having the same quantum state, and occupying the same space. This is not taken into account in the HF analysis. The force of repulsion is greater then what is predicted in the HF analysis. The energy difference between the HF limit and the true energy, or exact solution of a non-relativistic, time-independent Schrödinger equation is called the correlation energy. There are several methods that have been adapted to approximate a correction for this energy difference. The method used in this paper, except where noted, is the Becke's 3-parameter hybrid method with the exchange-correlation functional developed by Perdew-Wang in 1991 (abbreviated B3PW91) [9, 10, 11, 12]. This method is a class of density functional hybrid method which uses the HF exchange energy, Local Charge Density, and General Gradient Approximation in calculating the Exchange-Correlation Energy (E_{xc}), developed by Perdew-Wang. For density functional methods [13] the general equation is the sum of the kinetic energy (T) and the potential energy (V) of all the interaction terms as a function of charge density (ρ), which is

$$E = T + V_{en}(\rho) + V_{ee}(\rho) + V_{en}(\rho) + V_{en}(\rho) + E_{xc} \tag{4}$$

This method has been shown to provide reasonable good agreement with experiment [3].

The third approximation is that the Molecular Orbitals can be expressed as Linear Combination of Atomic Orbitals. The Atomic Orbitals can be expanded in a set of Gaussian function. This set is called a basis set. The basis set is the basic ingredient that leads to a mathematical representation of the Atomic Orbitals and which is centered on the nuclei in the molecule. Improvements in approximations can be made by using larger basis sets, however, the larger basis sets and greater accuracy requires more computational power and time.   These are the three important approximations in optimizing the structure of molecules and calculating their vibrational frequencies and modes.

The number of vibrational modes for a given molecule depends on the number of degrees of freedom and the number of atom in the molecule. Each atom within a molecule has 3 degrees of freedom; x, y, and z direction. The number of vibrational modes for a non-linear molecule is 3N-6; where N is the number of atoms in the molecule.   For example the molecule sarin (Table 1) has 18 atoms, or 3N-6 = 48 modes of vibration. The vibrational mode representing the strongest spectral feature at 1019 cm^{-1} is dominated by the C-O-P stretch/compression. For the spectral feature at about 850 cm^{-1}, the dominant vibrational motion is the P-F stretch. The hot bands and overtones, as well as the rotational structures, are not included in the approximations.

There are additional approximations that are also used in calculating the spectral signatures of molecules. The vibrational frequencies that are calculated are based on small, harmonic oscillations, of the atoms in the molecule. However, the vibrations of the atoms in molecules do not behave like a harmonic oscillator, but instead behave like an anharrmonic oscillator. The anharmonic frequency is usually slightly less than the harmonic frequency. To correct for this, the harmonic frequency is multiplied by a factor, which is usually less than one. The corrected vibrational frequency then becomes

$$f_{corrected} = factor \times f_{harmonic} \tag{5}$$

A factor of 0.985 is used for the B3PW91 methods. A factor of approximately 0.92 is used for the HF method. Once the vibrational frequencies and relative amplitudes are calculated a broadening function can be applied to the lines. The width of the function is arbitrary, but is chosen to give the lines a shape that resembles that of spectral features. .

In addition to the vibration of molecules, molecules also rotate. The rotational structure of spectral features is seen in many of the lab-measured spectra as "feet" that extend outward from the vibrational band center frequencies. The rotational Q-branch feature, which looks like spikes in many of the measured spectra, should correspond to the calculated vibrational frequencies. The degree of spread in the rotational structure is dependent on the rotational constant (B), which in turn is inversely proportional to the Moment of Inertia. If the rotational structure has a narrow spread, than the other broadening effects become dominant and the rotational structure is merged into one spectral

feature centered on the vibrational frequency. This "narrowness" of the rotational structure occurs most frequently with larger molecules where the Moments of Inertia tend to be larger and the resulting rotational constant tends to be smaller, which results in a better agreement between measured and the calculated spectrum. Although the rotational constants are calculated for every molecule, the rotational structures are still too complicated to model and they do not contribute much to the overall identification or chemistry of the molecule.

3.   RESULTS

The input parameters to the Gaussian software include an approximate molecular structure (coordinates of the atoms), method, basis set, and the type of calculations. The type of calculations includes a full optimization of the molecular structure. In a full optimization the position of the atoms are adjusted in the molecule after each iteration until a minimum energy structure is found. The optimization is followed by a frequency calculation where the second derivative of the energy is approximated, from which the modes of vibration can be calculated. Only the vibrational modes whose frequencies are between 700 and 1500 cm^{-1} are displayed and compared with the PNNL spectral library. In several molecules there are vibrational modes that extend outside this spectral range and are not shown.

3.1 Comparisons of Different Methods and Basis Sets

The structure of sarin is shown in Table 1. Figure 1 shows the lab-measured spectrum of sarin (red) and calculated spectrum of sarin (blue) using the following methods and basis sets; HF/6-31g, HF/6-311g(d), B3PW91/6-31g, and B3PW91/6-311g(d). The method and basis sets that show the best agreement between measured and calculated is B3PW91/6-311g(d) , the bottom plot. This is expected since the B3PW91 method includes an approximation for the electron correlation, and the large basis set with a polarization term more accurately describes the Molecular Orbitals. The agreement between measured and calculated spectra is based on the position of the center frequency and the relative amplitudes of the spectral features in the calculated spectrum. The method and basis set that will be used for the remainder of the molecular systems will be B3PW91/ 6-311+g(3df,2p). This should be sufficient for reasonable agreement between calculated and measured spectrum for most molecules. However, there are some molecules where modeling will not work.
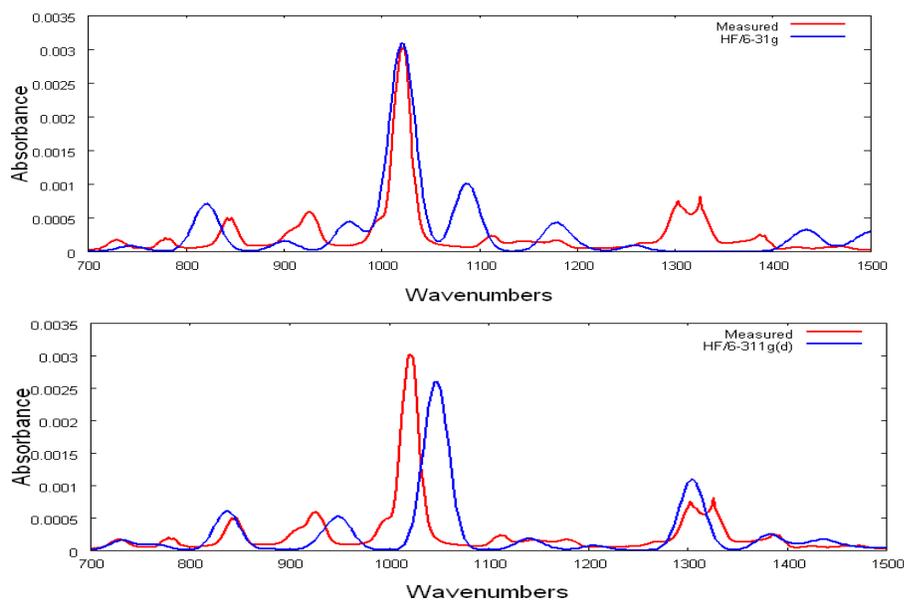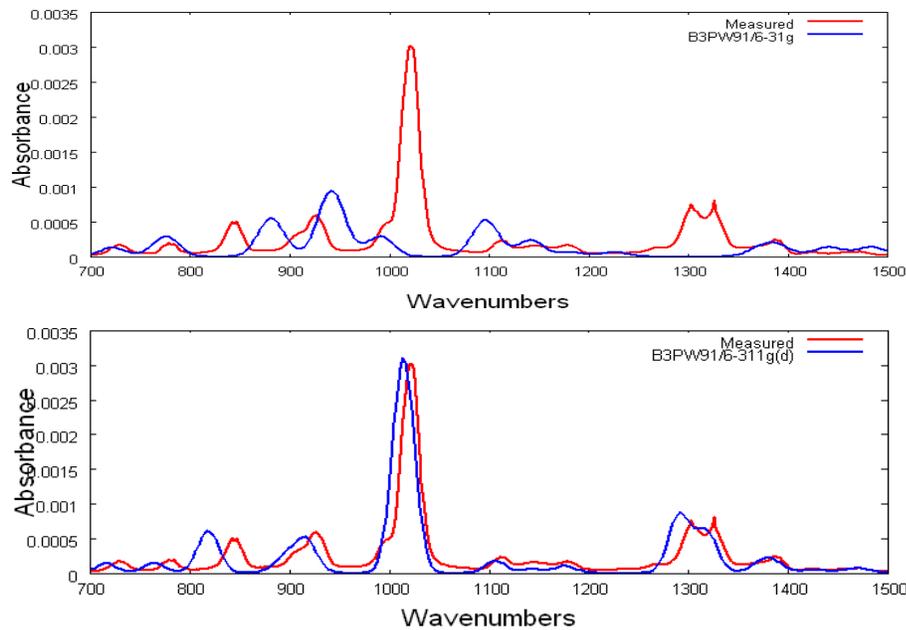
Figure 1: Different plots comparing lab-measured spectrum (red) of sarin with the calculated spectrum (blue) using different methods and basis sets.   The top plot (a) is using HF/6-31g; (b) HF/6-311g(d), (c) B3PW91/6-31g ; (d) B3PW91/6-311g(d).

3.2 Sarin and Chlorosarin

The molecular structures and the calculated spectrum of sarin and chlorosarin are shown in Table 1. A comparison between the calculated spectrum and measured spectrum of sarin are in good agreement. The spectral feature at the low frequency end of the spectrum for sarin has a slightly larger error in frequency.   The calculated spectrum has the spectral feature at about 820 cm^{-1} where the lab-measured value is about 840 cm^{-1}. The vibrational mode corresponding to this frequency is dominated by the stretching of the P-F bond. The large amplitude feature at 1019 cm^{-1} in sarin is dominated by the vibrational mode corresponding to the stretching and compression of the P-O-C bonds.

The molecule chlorosarin is very similar to the structure of sarin, with the fluorine atom being replaced by the chlorine atom. Since the molecules are similar in structure, the calculated spectrum is also similar, as seen in Table 1. There is no spectrum of chlorosarin available; however, its spectrum can be approximated.   There is one big difference between the signatures of chlorosarin and sarin and that is with the frequency corresponding to the vibrational mode dominated by the P-Cl stretch. For sarin the spectral feature that corresponded to the P-F stretch was approximated at 820 cm^{-1}, however for chlorosarin, the spectral feature corresponding to the P-Cl stretch has been shifted to lower frequencies, centered at approximately 515 cm^{-1}.   This is expected since the mass of chlorine is larger than fluorine. The atomic mass of fluorine and chlorine are 19 u and 35.45 u respectively.

3.3 Isopropyl Alcohol

The structure and calculated spectrum of isopropyl alcohol are shown in Table 1.   There is fair agreement between the measured (red) and calculated spectrum.   The relative amplitudes are slightly off, but the center frequencies of the spectral feature are in good agreement with the measured spectrum.

3.4 Cyclohexane

There is good agreement between the lab-measure and calculated spectra of cyclohexane (Table 1). The rotational structures in the vibration bands are visible in the lab-measured spectra. The calculated frequency of the band center is nearly centered on the Q-branches, or band centers, and the relative amplitudes are in good agreement.

3.5 Cyclohexanone

The calculated spectrum for cyclohexanone is in good agreement with the lab-measured spectrum (Table 1). The rotational structure is less defined for this molecule, due to the smaller rotational constant (Table 2) and the larger bandwidth of the broadening, however, the Q-branches do appear as spike centered on the calculated frequencies.

3.6 Diisopropyl Ether

There is good agreement between the calculated spectrum and the lab-measured spectrum (Table 1). The calculated vibrational frequencies are centered on each spectral feature in the lab–measured spectrum. There are no visible rotational structures in this spectrum do to the larger size of the molecule, or smaller rotational constant.

3.7 Phosphorous Oxychloride

The number of vibrational modes for phosphorous oxychloride is 9, since there are 5 atoms in the molecule. The single and highest frequency mode at about 1350 cm^{-1} corresponds to the P-O stretch in the double bond. This is the only spectral feature visible in this spectral range (see Table 1). The calculated vibrational frequency is in very good agreement with the lab-measured spectrum. There is some rotational structure visible due to the narrowness of the broadening even though the rotational constants are small. The Q-branch spike is visible in the lab-measured spectrum.

3.8 Trifluoromethyl Sulfur Pentafluoride

Although the general shape of the spectrum is in good agreement with the lab-measured spectrum (Table 1), the spectral features in the calculated spectrum appeared to be biased toward the low frequency side of the lab-measured spectral features. The bias is caused by using an anharmonic factor that was slightly too small. The relative amplitudes of the spectral features are in excellent agreement. The rotational structure is not visible except for the spectral feature at about 750 cm^{-1}.

3.9 Ethanol

The calculated spectrum of ethanol is in good agreement with the lab-measured spectrum (Table 1). There are Q-branch spikes in the lab-measured spectrum and rotational structures are also observed. The relative amplitudes are in reasonable agreement.

3.10 Ammonia

The modeling of spectral signatures does not work for every molecule. For ammonia the vibrational (or band center) frequency is calculated for the ammonia molecule to be approximately 1020 cm^{-1} (Table 1). This corresponds to the symmetric stretching of the N-H bonds. However, the nitrogen atom does undergo quantum mechanical "tunneling" which results in a lab-measured spectrum having multiple spectral features. The calculated frequency for this vibration mode is centered on the multiple lines in the lab-measured spectrum.

Table 1: The table below shows the molecular structure and comparisons between the lab-measured spectrum and the calculated spectrum for various chemicals. The molecular optimization and frequency calculations were performed using B3PW91/6-311+g(3df,2p).

Structure	Calculated and lab spectra
Sarin	
Chlorosarin	No lab spectrum available
Isopropyl alcohol	
Cyclohexane	
Cyclohexanone	
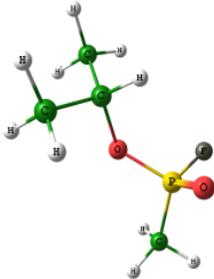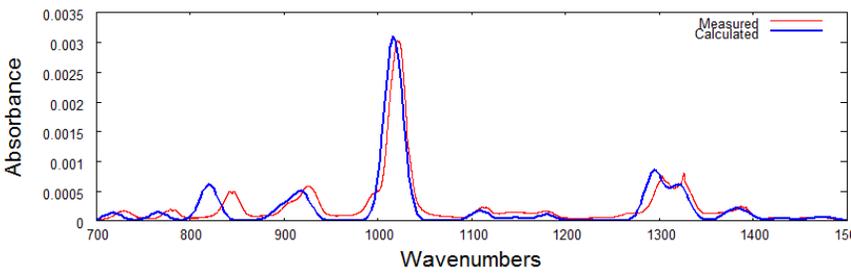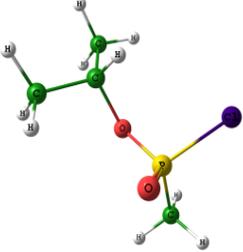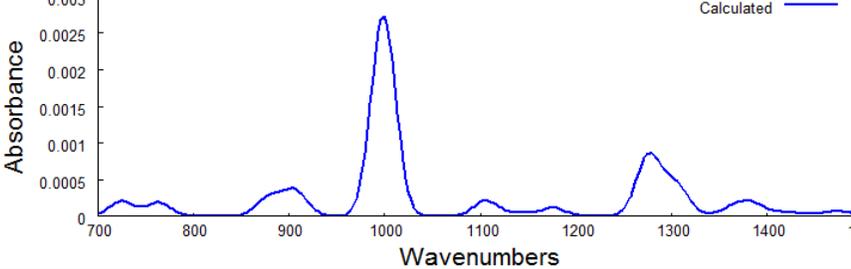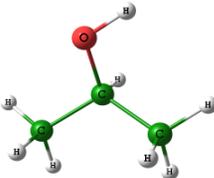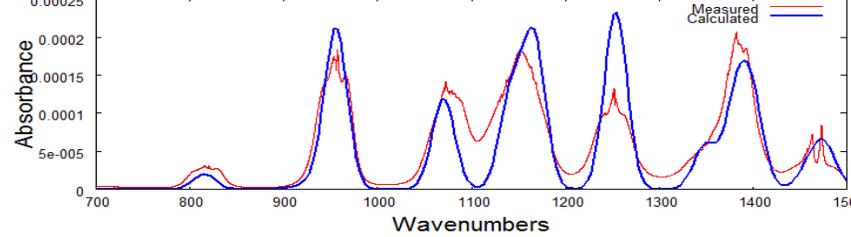
Diisopropyl ether	
Phosphorous oxychloride	
Trifluoromethyl sulfur pentafluoride	
Ethyl alcohol	
Ammonia	
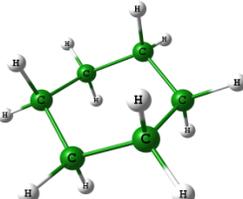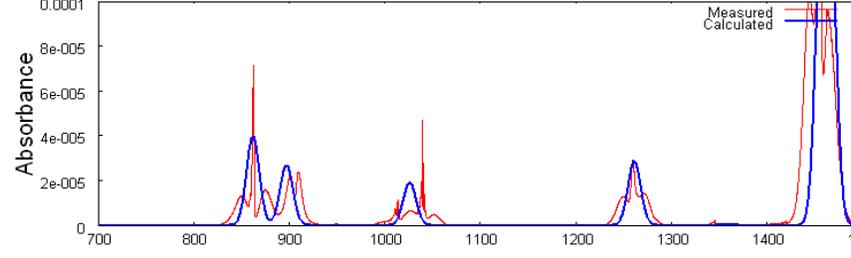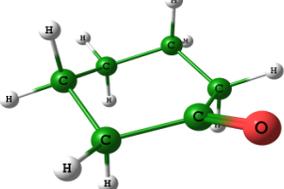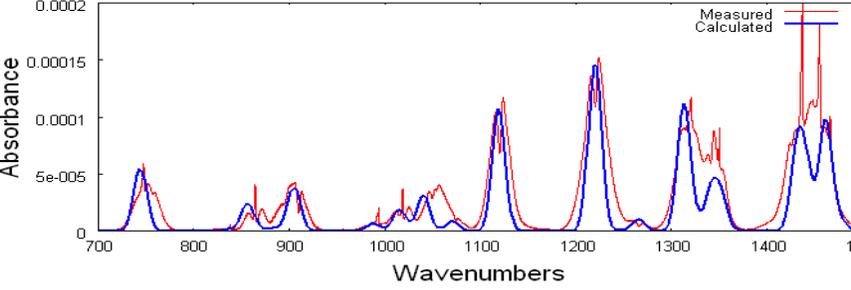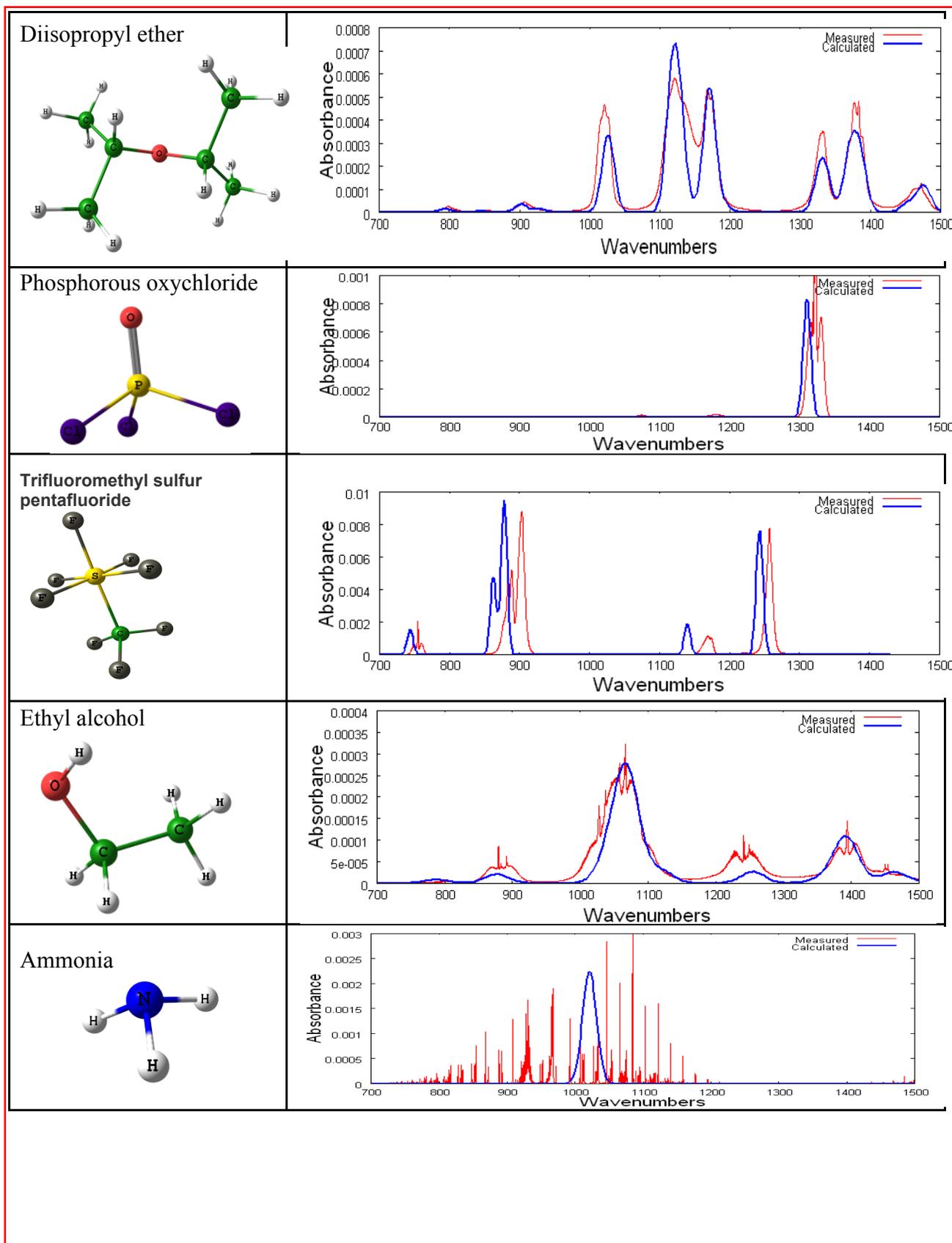
Table 2: The table below shows the rotational constants of each of the molecules modeled in this paper. Note that the molecules with the smaller moments of inertia have the larger rotational constants.

Molecule	Calculated rotational constants (GHZ)
Sarin	2.81285, 1.14235, 1.04103
Chlorosarin	2.02800, 1.02367, 0.87018
Isopropyl alcohol	8.53665, 8.06595, 4.78400
Cyclohexane	4.33652, 4.33429, 2.47497
Cyclohexanone	4.24016, 2.50570, 1.75457
Diisopropyl ether	4.11436, 1.55006, 1.38152
Phosphorous oxychloride	1.99438, 1.99438, 1.45367
Trifluoromethylsulfur pentafluoride	1.81037, 1.08398, 1.08348
Ethanol	34.73164, 9.21491, 8.14795
Ammonia	298.45838, 298.45838, 189.82443

4. CONCLUSION

Density Functional hybrid methods offer an alternative to approximating spectral signatures of many types of molecules when lab measurements are not available. The main points in this paper include the following:

- In general, better approximations of spectral signatures are achieved with smaller rotational constants (larger molecules). The reason for this is that the spread of the rotational structure is less with a smaller rotational constant, which result in spectral features that appear more Gaussian.
- For many molecules, identification can be made by comparing the spectral signature of known molecules with that of unknown molecules. Although the modeling of signatures is an approximation, by comparing the approximate signature with a library of signatures, chemicals can be ruled out as a spectral match, instead of ruling them in.
- The modeling of spectral signatures can be useful in the identification when there are no other lab measured spectral signature available to make identification.

REFERENCES

[1]   Steven W. Sharpe, Timothy J. Johnson, Robert L. Sams, Pamala Chu, George C. Rhoderick, and Paricia A. Johnson, "Gas-Phase Databases for Quantitative Infrared Spectroscopy," Applied Spectroscopy 58(12), 1452-1461, (2004).
[2]   http://nwir.pnl.gov
[3]   Jeffrey W. Mirick, [A Study of Molecular Processes in Calcium Clusters, Sarin, and Calcium Carbonate Clusters], (2002).
[4]   James B. Foresman and Eleen Frisch, [Exploring Chemistry with Electronic Structure Methods, 2nd Edition], Gaussian Inc (1995).
[5]   Frank Jensen, [Introduction to Computational Chemistry], John Wiley & Sons Inc (1999).
[6]   Andrew R. Leach, [Molecular Modeling, Principles and Applications], Addison Wesley Longman Limited (1996).
[7]   http://www.gaussian.com/
[8]   http://www.msg.ameslab.gov/GAMESS/

[9]   A. D. Becke, "Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior," Physical Review A 38, 3098 (1988).

[10]   A. D. Becke, "A New Mixing of Hartree-Fock and Local Density-Functional Theories," J Chem Phys 98, 1372 (1993).

[11]   A. D. Becke, "Density Functional Thermochemistry. III. The Role of Exact Exchange," J Chem Phys 98, 5648 (1993).

[12]   J. Perdew, K. Burke and Y. Wang, "Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System," Physical Review B 54, 16533 (1996).

[13] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," Physical Review B 136, 864 (1964).