Ab Initio Molecular Orbital Theory—A tool for THz Spectroscopic Investigation

Inke Jones, Tamath J. Rainsford, Samuel P. Mickan, and Derek Abbott
Centre of Biomedical Engineering and Department of Electrical & Electronic Engineering, The University of Adelaide, SA 5005, Australia

ABSTRACT
Terahertz spectroscopy, which investigates the electromagnetic spectrum of samples between 0.1 and 10 THz, allows not only for exploration of molecular structures but also of molecular dynamics. One difficulty in performing THz spectroscopy is that the data can be noisy and difficult to interpret. Ab initio molecular modelling has recently become more and more useful in the prediction of, for example, molecular structures, dynamic states and isomeric forms. Since the structure of biomolecules is closely related to their functionality there are broad ranging applications in biomedicine, for example in DNA sensing. An a priori knowledge of the expected THz spectra allows for improved experimentation. There is a growing and recognised need for THz spectroscopic databases to be created and made available along with classifiers that are able to effectively detect a specific substance. We show, for a specific example, the 9-cis and all-trans retinal isomers, how ab initio molecular orbital calculations and quantum chemical modelling programs, such as Gamess, can aid in this endeavour.

Keywords: THz spectroscopy, ab initio molecular modelling, retinal isomers

1. INTRODUCTION
Various rotational, vibrational and translational modes of molecules are within the terahertz (T-ray) range (0.1-10 THz or 3-333 cm$^{-1}$ in wavenumbers). Since these modes are unique to a particular molecule it is possible to obtain a ‘terahertz fingerprint’ allowing for the identification of chemical substances.

Figure 1 shows the electromagnetic spectrum including the ‘Terahertz (THz) gap’.

Terahertz generation and detection techniques have only recently become available. Originally, T-rays were generated and detected by employing conventional techniques borrowed from microwave and millimeter

Further author information: (Send correspondence to I. Jones)
E-mail: ijones@eleceng.adelaide.edu.au
technologies. Since the advent of solid-state Continuous Wave (CW) and ultrafast pulsed lasers, used with biased semiconductors and nonlinear crystals, there have been significant advances in THz technologies. Current research is focusing on the actual applications of such systems and is driving their development towards devices that are low cost, compact and easy to use.

Most molecules have dense and distinctive absorption spectra at THz frequencies, which has led to much interest in THz spectroscopy. Using THz transmission or reflection spectroscopy, samples ranging from gases to solids can be characterised at THz frequencies. The density of molecular twisting and bending modes in the THz band provides a wealth of information about the composition and state of the samples.

Being able to accurately differentiate between various isomeric and polymorphic structures is essential for a number of applications in biomedicine and pharmaceuticals as well as for security applications. For example Taday et al. have shown that it is possible to differentiate between the different isomers that are present in paracetamol (2-acetamidophenol, 3-acetamidophenol and 4-acetamidophenol) by comparing their THz spectra. Such applications are further supported by the fact that plastic and polyethylene packaging is transparent to THz. This allows for better environmental control of samples and also for the direct inspection of packaged materials for quality control monitoring as THz spectroscopy can be applied, for example, while tablets are in blister packets. Thus THz can be used to monitor changes that are related to time, pressure, and temperature without influencing the phase or chemically changing the pharmaceutical.

Pulsed THz techniques are based on ultrafast lasers, where the femtosecond regime allows for the possibility of studying molecular dynamics. The ability of THz to follow dynamical changes at the molecular level has caught the attention of many researchers. For example Upadhya et al. used a time-resolved broadband (100 GHz - 12 THz) THz spectroscopy system to study the far-infrared vibrational modes of crystalline saccharides, and to understand the dynamics of both inter- and intramolecular interactions. We have begun to explore the potential of THz for distinguishing between various isomeric forms of a molecule, studying conformational changes in proteins and other biomolecules. In particular we are exploring the various isomers of retinal. A great deal is already known about retinal’s isomers and its related proteins. A knowledge of the protein dynamics is widely useful to researchers in areas from molecular simulation in HIV research, optical memory devices and artificial vision through to novel technologies such as nano-mechanical switches. Currently, very few THz studies of retinal have been carried out. The work of Walthier et al. has shown that it is possible to identify the different isomeric structures of retinal using THz spectroscopic data. This result indicates that it is worth carrying out further studies in order to look at these lower frequencies more comprehensively and also shows that THz is a useful regime for our ultimate aim of studying retinal dynamics.

One difficulty in performing THz spectroscopy is that the data can be noisy and thus difficult to interpret. An a priori knowledge of the expected THz spectra allows improved experimentation. Molecular modelling, especially ab initio molecular orbital theory in combination with quantum mechanics and molecular mechanics (QM/MM), has recently become a useful tool in the support of spectroscopic data as well as in gaining information about intermolecular structure, interactions and dynamics. Modelling allows for the visualisation and study of the conformation and structure of a molecule, the study of mechanisms of reactions, and comparison of molecules with each other. Modelling the cis-trans isomerization of the retinal molecule can potentially provide information about the structure of intermediates as well as vibrational frequencies and dynamics. We use ab initio molecular orbital calculations to find frequencies in the THz range.

2. ULTRAFAST THZ SPECTROSCOPY

Ultrafast THz generation and detection uses the ultra-broadband nature of femtosecond (fs) optical laser pulses to reach the THz region of the spectrum. Pulsed T-ray radiation consists of ultrashort pulses, with a bandwidth spanning the range from approximately 0.1 to 10 THz (3-333 cm⁻¹), covering the THz Gap. These femtosecond laser pulses are fired at a photoconductor or a crystal, and thereby generate THz electromagnetic transients that can be detected. Through a beam splitter and a synchroniser, the laser pulses are forced to strike the THz generator and detector with a known phase coherence. A time-dependent waveform proportional to the
THz field amplitude and containing the frequency response of the sample can then be produced by scanning
the time delay and sampling the signals on the detector.

**Figure 2. Illustration of a transmission THz-TDS system.** The ultrafast laser beam is split into pump and
probe beams. The pump beam is incident on the THz emitter to generate THz pulses and the THz pulses are collimated
and focused on the target using parabolic mirrors. After transmission through the target the THz pulse is collimated
and refocused on the THz detector. The optical probe beam is used to gate the detector and measure the instantaneous
THz electric field. A delay stage is used to offset the pump and probe beams and allow the THz temporal profile to be
iteratively sampled. Whilst this example shows a THz system in transmission mode, it should be noted that reflective
mode measurements are also well established. Figure adapted from Ferguson and Zhang (2002).27

The two main applications in which THz techniques are involved are THz spectroscopy and THz imaging.
Terahertz time-domain spectroscopy (THz-TDS) and related THz technologies, especially THz wave (T-ray)
imaging modalities provide spectroscopic information, such as functional imaging28 and has the potential to
impact on an almost limitless number of interdisciplinary fields including communications,29 imaging,30–32
medical diagnosis,33 health monitoring,34 environmental control,35 and chemical9 and biological identifi-
cation.27, 36, 37 T-rays do not subject biological tissue to harmful ionizing radiation, because they maintain
low-photon energy (4 meV at 1 THz),38 in comparison to typical X-ray photon energy that is in the order
of keV. In addition to imaging, THz systems can also provide spectroscopic information such as unique
rotational, vibrational and translational responses of materials and therefore enable molecular fingerprinting
with T-rays.28 The potential of THz spectroscopy has been realized for a variety of applications. A number of
researchers have measured the THz spectra of various substances of interest in security and defence,39 medical
research,40 and molecular science.41, 42

3. THZ MOLECULAR MODELLING ON THE EXAMPLE OF RETINAL ISOMERS

Researchers have long been interested in the mechanism of the photocycle and hence the retinal molecule.
Visual pigments, which are found in both vertebrates and invertebrates, are members of a protein class called
“G-protein coupled receptor proteins.” In vertebrates, rhodopsin, also known as visual purple, is a photore-
ceptor protein bound to its chromophore, i.e. retinal. In the vertebrate retina, rhodopsin is responsible for
the primary events in the detection of light.43, 44 It catalyses the only light sensitive step in vision and is found,
for example, in the rod cells of the vertebrate visual system.13 Bacteriorhodopsin is a photosynthetic pigment
similar to rhodopsin and is found in the purple membrane of halophilic bacteria. It has photoreactions similar to those observed in rhodopsin and functions as an energy transducer or proton pump.

What all of these pigments such as rhodopsin and related bacteriorhodopsin have in common is that they all comprise seven transmembrane helices and are composed of a retinylpolyene chromophore bound to a parent opsin protein. The retinal chromophores occur in different isomeric forms depending on the physical function of the protein. In rhodopsin (for vertebrates), 11-cis retinal is covalently bound to opsin and the photoisomerized all-trans form is eventually released, which then must be reconverted to 11-cis retinal. However, in certain invertebrates such as insects however, the active form of rhodopsin, metarhodopsin, is thermally stable and is thus photoisomerized back to 11-cis retinal.

During the photocycle, changes occur to the retinal chromophore as well as to the entire protein. Changes in the retinal binding pocket cause structural modification of the retinal molecule. The cis-trans isomerization is the fastest known biological photochemical reaction (~200 fs). Vibrational coherence of the isomerization may be responsible for the speed of this reaction. The initial excited state dynamics along the C$_{11}$=C$_{12}$ torsion may direct the correct distortion of the chromophore. Indeed, studies with chromophores, which lack intramolecular steric interactions, isomerize on a slower time scale (400-600 fs) and produce a smaller quantum yield.

Vibrational coherence has previously been observed in the reactant state of the photosynthetic reaction centre of bacteria prior to ultrafast electron transfer. This has led to a series of experimental studies of the role of vibrational coherence in chemical and biological reactions, which since then has been observed in various biologically important molecules.

Electron microscopy and X-ray crystallography have provided a great deal of information about structure and dynamics of retinal. One approach to studying the nature of structural changes is to use conditions that prolong the life of the structural intermediates, such as lower temperature, adding in additives, and the use of genetically altered variants of the protein. Spectroscopic studies have characterized vibrational modes and are useful in predicting possible structures of intermediates as well as providing information about molecular mechanics. Different information can be gained about a molecule depending on which part of the spectrum is being explored. The region that has long been of most interest for chemical analysis is the mid-infrared region (4,000 cm$^{-1}$ to 400 cm$^{-1}$), which corresponds to changes in vibrational energies within molecules. However, information about dynamics is contained within the far-infrared part of the spectrum (400 cm$^{-1}$-10 cm$^{-1}$), especially below 200 cm$^{-1}$.

It is rarely possible to identify an unknown compound using IR spectroscopy alone. Being able to interpret the terahertz spectrum of a molecule and assign its low frequency vibrational modes would allow for more precise identification. Terahertz spectroscopy is a potentially useful modality for explaining intermediates and dynamics. Ultrafast changes in absorption spectra can be seen with pump-probe experiments using a THz pulse duration of less than a picosecond. Walther et al. have used THz time domain spectroscopy to measure the FIR spectra of 3 isomers of retinal: all-trans, 13-cis and 9-cis. Their observations show that there are distinct differences between the low frequency vibrational spectra of the three isomers. Also by comparing the three isomers they were able to deduce the approximate localization of the different vibrational modes within a molecule. However, they only measured the spectra between 0.3 and 3 THz and therefore only observed a small number of vibrational modes. A broader view of the THz spectrum of these isomers, for example between 0.1 and 6 THz, would give better insight into the molecular dynamics and would further enable localization of the modes within the molecules.

### 3.1. Ab initio molecular modelling

Modelling techniques allow any chemical species to be studied in detail. For example, in our case, calculations can be performed on the reactive intermediates that are difficult to investigate experimentally. Information can be gained about the reactive transition structures and excited states—sometimes it is only possible to obtain this information by calculation. Complete sets of data can be obtained and calculations can be carried out on structures artificially constrained to allow individual interactions to be assessed. There is a wide range
of molecular modelling and quantum chemistry software available and different researchers have made use of different techniques depending on their research goals.

The *ab initio* approach to chemistry is that the Schrödinger equation leads to the direct quantitative prediction of chemical phenomena using only Planck’s constant, the speed of light and the masses and charges of electrons and nuclei. Of course very few problems are tractable and so approximate mathematical models of the Schrödinger equation for which the solutions may exist are used. There are two distinctly different approaches to approximating solutions of Schrödinger’s equation. The first is to solve the problem at the highest level of theory. This is only possible for very small systems such as the hydrogen molecule. The second approach is to solve a theoretical model for which a number of characteristics must hold true: The model should be unique, well defined, continuous, unbiased, and relative errors should increase in proportion to the size of the molecule. It should also yield a total energy that is an upper bound to that which would result from an exact solution of the full Schrödinger equation. Such a model should also be implementable on a computer with minimal computational effort. The most commonly used models of this class are those which are based on molecular orbital theory: the approximate treatment of electron distribution and motions, which uses one-electron functions or orbitals to approximate the full wave equation. A many-electron wave function is constructed from the molecular orbitals in the form of a determinant. Such models have been validated through systematic comparison with experimental data. Individual molecular orbitals are expressed as linear combinations of a finite set of one-electron functions known as basis functions. The choice of basis set determines the level of accuracy and depends on two components (a) the size of the basis set and (b) the treatment of electron correlation. In general the choice of basis set will always be a compromise.

Typical engines used for *ab initio* calculations are Gamess\(^1\) and Gaussian.\(^{59,60}\) Here, *ab initio* calculations themselves can be divided into quantum mechanics and molecular mechanics/dynamics. In general, quantum methods are slow but potentially very accurate, and require huge memory and CPU resources, which limit their application to relatively small molecules. On the other hand they provide full simulations of all the properties and fine behaviour of the molecules. For bigger molecules, methods with a reasonable accuracy requiring less computation time are needed. Molecular Mechanics and Molecular Dynamics (MM/MD) methods treat atoms as spheres with charge, and bonds as springs. They do not consider independent subatomic particles, and as such are more limited in their utility.

Such calculations begin with specifying the bond lengths and angles. Often information about symmetry can greatly reduce the time required for integral evaluation. Programs such as Gaussian and Gamess have incorporated many standard basis sets although non-standard sets can be specified in detail for each atom if required. First, integrals are calculated and a guess at the wavefunction made. Execution proceeds into a programmed loop where the self consistent field (SCF) equations are solved for the total energy and wavefunction. First derivatives of the energy with respect to the displacements in the nuclear coordinates are evaluated. If the wavefunction (gradient of the energy) is below some preset limit then the originally specified geometry represents within some limit a stationary point on the potential energy surface. The optimisation procedure then terminates. Otherwise the original geometry is varied and a new calculation of integrals, SCF and energy gradient follows.

Once the correct molecular geometry is obtained, other calculations such as the vibrational frequencies, IR and Raman spectra, excitation energies and related properties of excited states may be calculated. A number of other more advanced quantum mechanical methods have been used recently.\(^{23,26,61}\) Examples include multiconfiguration self-consistent field theories, density functional theory and the consistent force-field method. We have chosen to use *ab initio* techniques, in the first instance, which are well documented, easy to use and widely available.

### 3.2. Modelling approaches to retinal

In 1997, Gervasio *et al.*\(^{23}\) published the first report on *ab initio* calculations on low frequency modes of a retinal isomer. The infrared and Raman spectra of all-\textit{trans}-retinal were obtained at room temperature and at 15 K. Aside from these experiments, *ab initio* calculations of vibrational frequencies based on density functional
theory were carried out. Frequencies between 17 and 334 cm$^{-1}$ were recorded. Frequencies and normal modes were also calculated using the density functional approach B3-LYP. This showed that many of the vibrational modes are located on the ring or chain fragment of molecule and that the chain torsional modes are of primary interest for the photoisomerization process.

More recently, Morari et al.$^{25}$ used \textit{ab initio} and vibrational self-consistent field (VSCF) computations to investigate the vibrational normal coordinates of the protonated Schiff base (PSB) of 11-cis retinal. These studies focused on the normal coordinates modes that involve the central C=C bond, which plays a significant role in the isomerization process. The calculations were performed at the Restricted Hartree-Fock (RHF) level with Pople’s N-31G split valence basis set. Light atom polarization functions were also used (RHF/6-31G*). Anharmonicity correction were taken into account by using the correlation-corrected vibrational SCF (VSCF) method. Vibrational frequencies of 270, 328 and 998 cm$^{-1}$ were reported with relatively large corrections. It was found that those vibrational bands contribute significantly to the rotation around the angles most important in promoting the isomerization process.

3.3. THz modelling of retinal isomers

The calculations of the structure and vibrational frequencies of all-trans and 9-cis retinal were carried out with \textit{ab initio} methods using the GAMESS-UK electronic structure package$^1$ and the visualization programs Molden$^{62}$ and GaussView.$^{59, 60}$ The calculations included optimization of the molecular structures initially using the minimal basis set STO-3G (Slater-Type Orbital), later Pople’s split valence basis set 6-31G** with light and heavy atom polarization functions. A low gradient convergence tolerance was used, which was increased with each optimization step. The frequency calculations were carried out by calculating the Hessian using the same basis set and tolerance as in the last optimization. The modelled and optimised structures of all-trans and 9-cis retinal are shown in Figure 3 and 4.

![Figure 3. The all-trans (A) and 9-cis (B) retinal isomers visualized in Molden,$^{62}$ optimized with Gamess.$^1$](image-url)
Figure 4. The calculated THz spectra of all-trans and 9-cis retinal in comparison to each other and to experimental data. The spectra cover the region of approximately 0.1 to 6 THz (≈ 3-200 cm$^{-1}$). The intensities are shown in arbitrary units, where the highest intensity of the full spectrum of each molecule was equated with 1.0 and all other frequencies were normalised to this value. The reference experimental data is included only in consideration of the frequency modes, not their absolute intensities as only relative values were provided by the previous study.

The calculated vibrational spectra of all-trans and 9-cis retinal in the range from 0.1 to 6 THz (0-200 cm$^{-1}$) are plotted in Figure 5. Table 1 shows a comparison of our results with previous data. The frequency lines in Figure 5 show that there is good agreement between our calculated vibrational modes and previous experimental data. They also show that there are significant differences in the THz spectra of all-trans and 9-cis retinal in frequency shifts as well as in relative intensities. For example, 9-cis retinal shows a peak at 24.40 cm$^{-1}$ and all-trans retinal does not. Also, all-trans retinal has a peak at 48.31 cm$^{-1}$ is missing in the spectrum of 9-cis retinal. Furthermore, the intensities of the different peaks common in both spectra show significant differences.
Table 1. Calculated low frequency vibrational modes (wavenumbers in cm\(^{-1}\)) of all-trans and 9-cis retinal compared to experimental results

<table>
<thead>
<tr>
<th></th>
<th>all-trans retinal</th>
<th>9-cis retinal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ref.(^{23})</td>
<td>This work</td>
</tr>
<tr>
<td>46</td>
<td>-</td>
<td>15.59</td>
</tr>
<tr>
<td>47</td>
<td>-</td>
<td>33.27</td>
</tr>
<tr>
<td>48</td>
<td>-</td>
<td>34.16</td>
</tr>
<tr>
<td>53</td>
<td>46</td>
<td>48.31</td>
</tr>
<tr>
<td>54</td>
<td>53</td>
<td>-</td>
</tr>
<tr>
<td>61</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>64</td>
<td>66</td>
<td>64.39</td>
</tr>
<tr>
<td>68</td>
<td>69</td>
<td>-</td>
</tr>
<tr>
<td>89</td>
<td>90</td>
<td>84.42</td>
</tr>
<tr>
<td>109</td>
<td>-</td>
<td>107.75</td>
</tr>
<tr>
<td>130</td>
<td>-</td>
<td>111.74</td>
</tr>
<tr>
<td>152</td>
<td>-</td>
<td>132.46</td>
</tr>
<tr>
<td>163</td>
<td>-</td>
<td>149.31</td>
</tr>
<tr>
<td>172</td>
<td>-</td>
<td>164.31</td>
</tr>
<tr>
<td>184</td>
<td>-</td>
<td>173.55</td>
</tr>
<tr>
<td>194</td>
<td>-</td>
<td>196.61</td>
</tr>
</tbody>
</table>

4. CONCLUSION

Through \textit{ab initio} modelling we have found THz frequencies for two isomers of retinal: all-trans retinal and 9-cis retinal. Previous data in this low frequency range (0.1-10 THz) is scarce, however, our data fits well with the existing data. Our results also show that in the THz frequency range there are significant differences in the spectra of the all-trans and 9-cis isomers, which suggest proceeding with experimentation at these lower frequencies as well as further modelling of other isomers such as 11-cis and 13-cis retinal as well as various intermediate structures. From frequency versus intensity plots of our data it would be interesting to explore differences between the spectra of these isomers in the 10-40 cm\(^{-1}\) range experimentally.

5. ACKNOWLEDGEMENT

The authors would like to thank Trevor D. Lamb, John Curtin School of Medical Research, Australian National University, Canberra, for his useful comments on the manuscript and Magdalen Addicoat, School of Chemistry & Physics, the University of Adelaide, for useful discussions on molecular modelling. Funding from the Australian Research Council (ARC) and the Sir Ross and Sir Keith Smith Fund is gratefully acknowledged.

REFERENCES