Chemistry Theses   Department of Chemistry

8-10-2021

An Experimental Benchmark Set for Oscillator Strength Calculations of Small to Medium-sized Organic Molecules

Astrid Tarleton

atarleton1@student.gsu.edu

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_theses

An Experimental Benchmark Set for Oscillator Strength Calculations of Small to Medium-sized

Organic Molecules

by

Astrid Sophia Tarleton

Under the Direction of Samer Gozem, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

UV-visible absorption bands are typically characterized by their wavelengths and their intensities. Absorption intensities in experiments are typically reported as extinction coefficients, while in computational chemistry they are more often reported as oscillator strengths (f). The two quantities are related, although this relation is complicated by various broadening effects in experiments that are not usually accounted for in computations. While quantum mechanical methods for the computation of f (f_{comp}) are widely available, experimental oscillator strengths (f_{exp}) are rarely reported. In this study, we describe a protocol to systematically fit and integrate experimental UV-visible spectra. We then apply this protocol to derive f_{exp} for one-hundred small to medium-sized organic molecules. The corresponding f_{comp} are then obtained with density functional theory (DFT). By expressing experimental and computed absorption strengths using a common unit, we can quantify the accuracy of f_{comp}.

INDEX WORDS: Oscillator strength, UV-visible absorption spectroscopy, Transition dipole
moment, Gaussian curve fitting, Electronic transition, Benchmarking

An Experimental Benchmark Set for Oscillator Strength Calculations of Small to Medium-sized

Organic Molecules

by

Astrid Sophia Tarleton

Committee Chair: Samer Gozem

Committee: Kathryn Grant

Gregory Poon

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2021

DEDICATION

I would like to dedicate this work to my mother who has provided me unconditional

support and guidance throughout my academic and professional career.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1 INTRODUCTION

The Schrödinger equation is foundational to our understanding of quantum mechanics. In its simplest form, it is mathematically expressed as the following eigenvalue equation: $\hat{H}\psi = E\psi$. $\hat{H}$ is the Hamiltonian *operator* composing all kinetic and potential energy contributions of a system (molecule, atom, or electron). E is an energy value and is quantized if the *wavefunction* is quantized.[1] Under the *Born-Oppenheimer approximation*, the molecular wavefunction (ψ) of a system may be separately defined in electronic and nuclear terms.[1-3] A qualitative approximation of ψ is expressed as the product of three parameters: the *electronic wavefunction (ψ_o)*, the *vibrational wavefunction (χ)*, and the *electronic spin (S)* (Equation 1.1).[2]

$$\psi \sim (\psi_o)(\chi)(S)$$

Equation 1.1: Expression of the Molecular wavefunction.

1.1   Quantum computational methods

While exact solutions exist for the non-relativistic time-independent Schrödinger equation for hydrogen and hydrogen-like atoms, exact solutions for molecules are not tractable. Instead, modern quantum computational methods aim to solve the Schrödinger equation (or some equivalent of it) approximately for molecules to describe their dynamics and quantify their energies. With a basic understanding of the Schrödinger equation, a qualitative discussion of widely used quantum computational methods is presented in this section. Appendix A provides definitions of some of the technical terms shown in italics in this section.

Firstly, most modern computational methods solve the Schrödinger equation by describing the *molecular orbital*s of a system using a *basis set* composed of Gaussian-type orbitals (GTOs), which mean to resemble an *atomic orbital*s. This is a common approach used as an alternative to

the use of Slater-type orbitals in their basis sets. Slater-type orbitals (STOs) are better at reproducing the shape of the true atomic orbital but are more computationally demanding to compute.[4] GTOs are advantageous because they permit simpler integration of functions and these orbitals generate a more efficient computer algorithm.[4]

One of the earliest breakthroughs for approximating the solution to the Schrödinger equation is the Hartree-Fock method (HF).[5] In HF, the potential energy component of $\hat{H}$ in the Schrödinger equation is modified to include the nuclear attraction felt by the electron and the *average electron correlation* contributions to the system, instead of a explicit pair-wise interaction.[1, 5] The procedure for this method is rooted in the *orbital approximation*.

The adapted operator named the Fock operator solves the one-electron orbital problem for each electron in the field of other electrons in the molecule.[1, 6] One-electron orbitals are solved iteratively in this way until the results converge indicating the E is optimized. The results of HF calculations can be improved by using Post-Hartree-Fock methods. These methods use HF as a starting point and then account for electron-electron correlation energy using either a perturbative or the *variational principle* approach.[6] While the Post-HF methods provide a more accurate approximation of electron-electron correlation energy than HF methods, the former are highly time-consuming and sometimes difficult to execute in large molecular systems.[7]

Another widely used method is the Density Function Theory (DFT). Unlike the HF and Post-HF methods, DFT uses the *electron density* (P(r)) rather than Ψ to obtain E.[1] DFT operates under the assumption that the energy of the molecule is a function of the electron density. The electron density is expressed as P(r)= $\Sigma_{m,occupied}$ |(ψ_m(r))|^2, where m is the number of occupied orbitals and ψ_m(r) is the function holding information on the position of the electron in an occupied orbital.[1] This expression indicates P(r) is a function of ψ_m(r), where ψ_m(r) is also a function.

Hence, P(r) is a functional.[8] This method uses the modified Schrödinger equations, Kohn-Sham equations, which define P(r).[7] P(r) is composed of kinetic, nuclear, and electronic interactions, and the *electron exchange energy*.[6] The exact form of the exchange energy is not known.[6, 9] To compensate for this, computational and theoretical chemists have derived many approximated expressions for the exchange energy to uniquely fit a calculation to produce energy values that correlate well with experimental data.[9] One of the most widely used functionals in computational chemistry is the Becke Lee Yang Par (B3LYP) functional.[7, 10] This is also the method used in this study. B3LYP is a hybrid method that combines the Hartree-Fock exchange contribution (contributed by Becke) with DFT (with the potential developed by Lee, Yang, and Par).[7, 10] Thus, B3LYP accounts for both the quantum mechanical properties inherent to the wavefunction and partly for electron exchange correlation of the electron density.

To extend DFT to the calculation of excited states (which is required in this work), time-dependent DFT (TDDFT) is often used. TDDFT describes the density and energy properties of dynamic multi-electron systems with regards to time-dependent potentials, and uses those properties to extract information about the excitation energies.[1, 11, 12]

1.2   Electronic Spectroscopy

To understand the relationship between the oscillator strength f and experimental UV-visible spectral data, a discussion of electronic spectroscopy proceeds. Electronic spectroscopy is the detection and analysis of electronic transitions in an atom or a molecule induced by applied electromagnetic radiation.[1] In electronic absorption spectroscopy, the electronic transition only occurs when energy of a discrete value absorbed by the excited electron is equal to the energy of discrete value of the applied electromagnetic radiation.[1, 2] These electronic transitions are recorded in an emission or absorption spectrum as the amount (intensity) of radiation with respect to energy

(with energy typically reported as wavenumber ($\tilde{\nu}$) or wavelength (λ)).[1] The intensity of radiation absorbed by the sample is related to the transmittance. Transmittance is the intensity of light detected after passing through the sample (atom or molecule) divided by the intensity of incident light.[1, 13] Closely related to transmission (T) is absorbance (A): $A = -log_{10}(T)$.[13] Both transmittance and absorbance are quantities that are related to details of the experimental setup (concentration of the solvent and path length of light). Often, in electronic spectra, the molar extinction coefficient (ε) is reported, where ε is a measure of the likelihood of an electronic transition to occur at a certain wavelength. Importantly, ε is an intrinsic molecular property. The ε is related to A by the Beer-Lambert law: $A = (\varepsilon)(c)(l)$, where c is the molar concentration of the sample, and l is the optical path length.[13]

In an atomic absorption spectrum, a spectral line indicates that an electron in its initial (ground) energy state has transitioned to an excited energy state. In this case, ε is a measure of the height of the spectral line.[14] In a molecular absorption spectrum, due to electronic, nuclear, and solvent interactions, energy states mix and form broad bands of energy (as shown schematically in Figure 1.2.1).[1, 2, 14, 15] Here, the interpretation of ε as a measure of the probability for electronic transitions to occur requires a more involved discussion.
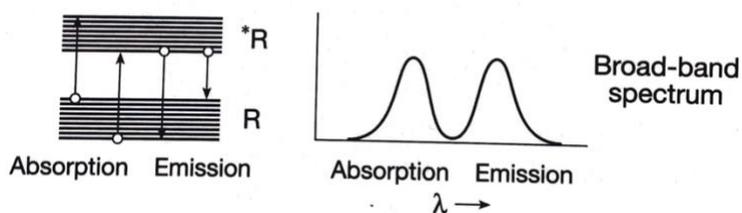
Figure 1.2.1: Illustration of molecular absorption and emission spectra.

Figure 1.2.1 is reproduced with permission from University Science Books.[2]

1.3   Vibronic contributions to broadening

The Franck Condon (FC) principle helps explain the variance of the ε at specific energy values in a molecular spectrum. This principle states that the most probable electronic transitions occur between vibrational wavefunctions which overlap the strongest in the initial and final states.[2, 16] This principle can be expressed mathematically (Equation 1.3.2).[2]

$$P_{i \to f} \propto \left[< \chi_i | \chi_f >^2 \right]$$

Equation 1.3.2: Expression for the Franck-Condon principle.

In Equation 1.3.2, the probability of a transition from a vibrational state of the ground state (i) to a specific vibrational state of the excited state (f) depends on the square of the vibrational overlap integral $< \chi_1 | \chi_2 >$, also known as the FC factor.

The FC principle is rooted in the understanding of timescales of intramolecular motions. The timescale of electronic transitions (10^{-18} s or attosecond timescale) is significantly faster than the timescale of vibrational motion ($>10^{-15}$ s or femtosecond timescale).[2] Therefore, the FC factor is a measure of the direct *vertical* vibrational overlap between the two mixing states (Figure 1.3.2).[2]

In a molecular absorption spectrum, unlike atomic spectra where there are no vibrations, these FC factors end up broadening each electronic transition. In a well resolved spectrum, (e.g., in the gas phase), the absorption spectrum appears as a series of sharp peaks. According to the FC principle, in cold (low-temperature) spectra, the vibronic peak with the greatest ε is a consequence of the lowest vibrational energy level (v=0) in the ground state wavefunction (ψ^o) overlapping most prominently with a higher vibrational energy level (v=4) in the excited state wavefunction (ψ^*) (Figure 1.3.2).[2, 14] Thus, $v = 0 \to 4$ (highlighted in red) is the most probable electronic transition because it has the largest FC factor.
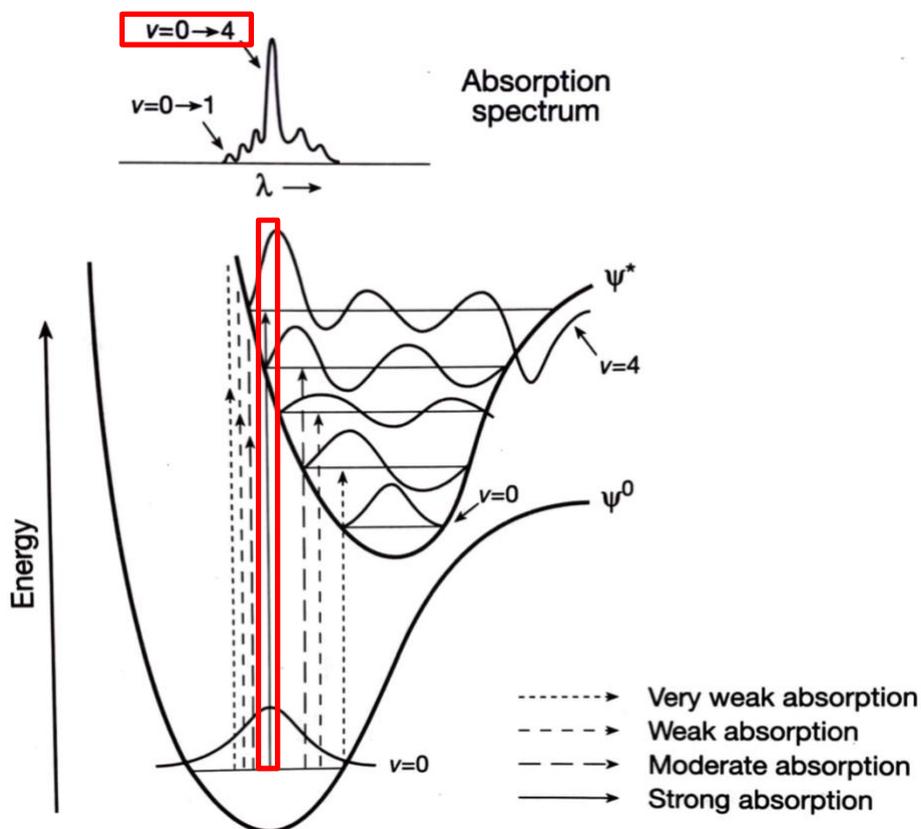
Figure 1.3.2: Representation of the Franck Condon principle in the molecular absorption of light.

Figure 1.3.2 is reproduced with permission from University Science Books.[2]

In a molecular absorption spectrum, the FC factors affect both the energy (or associated $\tilde{v}$ or λ) and the corresponding intensity (ε) of the electronic transition. As mentioned above, such transitions can be well resolved for a conformationally rigid molecule and/or in the gas phase. However, in a solvent, there is a large population of molecules that have slightly different rotational, vibrational, and electronic transition energies due to solvent interactions and due to conformational flexibility. The vibrational states in the ground state overlap with many vibrational states in excited states over range of wavelengths. All these electronic transitions occur

instantaneously upon light perturbation, so, the otherwise sharp spectral lines are spaced so closely, they are unresolved and appear as broad-bands.[2] Figure 1.3.3 schematically illustrates a typical atomic absorption spectrum (left), molecular absorption spectrum in gas phase (middle), and molecular absorption spectrum in solvent (right). Since most of UV-visible spectroscopy is in the solvent phase (including the spectra in this study), the majority of the spectra typically appear most similar to the schematic spectrum on the right of the figure.
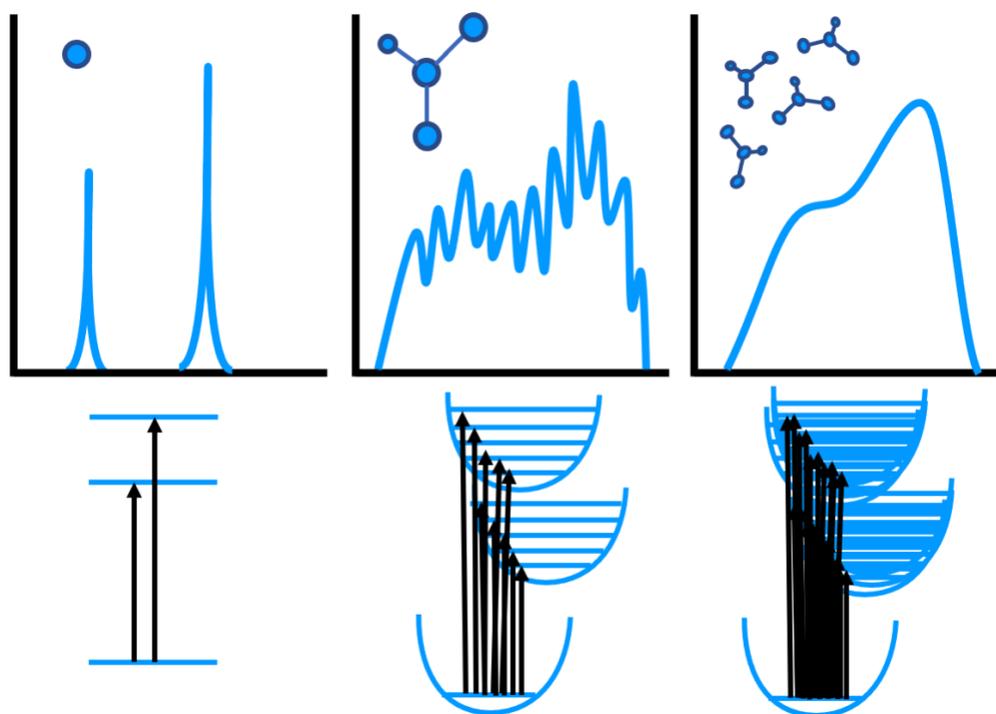
Figure1.3.3: Illustration of broadening in electronic absorption spectra.

1.4   Quantum mechanical and classical understanding of the oscillator strength

The starting point for understanding the probability of electronic transitions in quantum mechanics is with the perturbation matrix ($< \psi_1|P|\psi_2 >$). For light absorption, the perturbation P corresponds to the oscillating electromagnetic field associated with light, while ψ_1 is the initial state wave function and ψ_2 is the final state wave function. Using the approximations in equation 1.1, each wavefunction can be written as a product of an electronic wave function, nuclear wave function, and spin wave function. For spin-allowed transitions and within the Born-Oppenheimer approximation, in $< \psi_1|P|\psi_2 >$, it can be assumed that the spin and nuclear terms are not affected by the electromagnetic perturbation. The spin terms therefore simply give a factor of 1, and the nuclear terms give rise to the FC factors discussed in the previous section. What is left is the perturbation matrix now expressed only for the electronic wave functions, $< \psi_{o,1}|P|\psi_{o,2} >$. This term, arising from quantum mechanics after a series of approximations, can be related to a more classically understood quantity, the induced transition dipole moment (μ_i). μ_i defines the occurrence of an electronic transition as when an electron in a ground state wavefunction transitions to an excited state wavefunction when perturbed with light.[2] μ_i is proportional to the squared perturbation matrix, $< \psi_1|P|\psi_2 >^2$. μ_i is therefore just another parameter for describing the likelihood of electronic transition to occur, and can be linked directly to f (Equation 1.4.3).

$$f = \left(\frac{8\pi m_e \bar{v}}{3he^2}\right) \times < \psi_1|P|\psi_2 >^2.$$

Equation 1.4.3: A quantum mechanical expression for the oscillator strength.

In Equation 1.4.3, the variable, m_e, is the mass of the electron, $\bar{v}$ is the transition energy (cm^{-1}), h is Planck's constant, and e is the electric charge.

In absorption spectra, the highest ε is indicated as ε_{max}. The corresponding energy E_{max} (or $\tilde{\nu}_{max}$ or λ_{max}) indicate the energy for which the electronic transition is most likely to occur.[1, 2] However, the sum of all ε over energy for which the spectral band encompasses provides a more representative value of the *total* likelihood for an electronic transition to occur.[1] In other words, the total probability for a transition to occur is related to the area under the spectral curve. This area is named the integrated absorption coefficient (A) (Figure 1.4.4).[1]
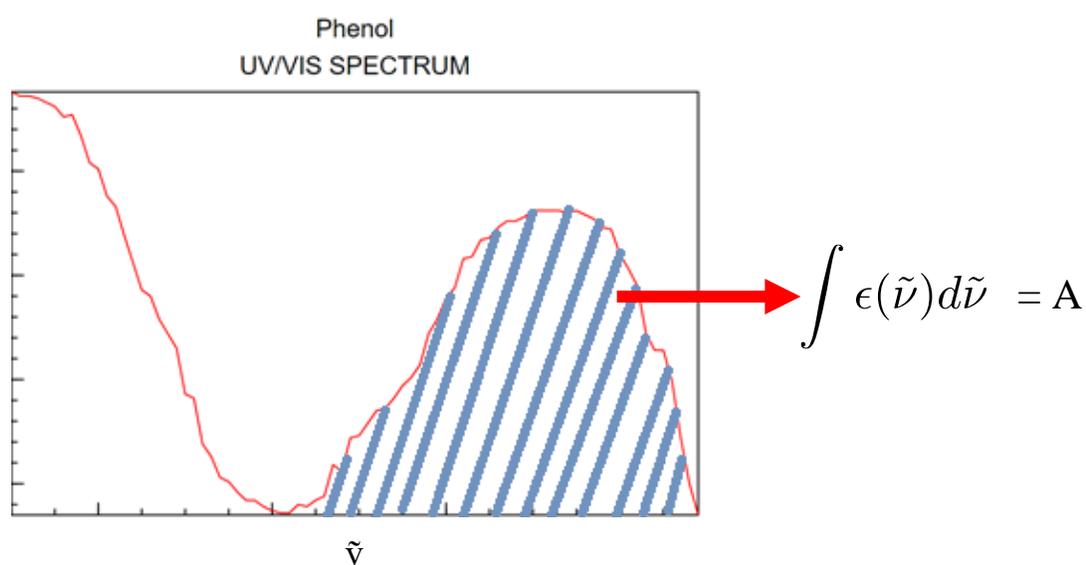
Figure 1.4.4: Representation of the integrated absorption coefficient (A).

Figure 1.4.4 is adapted from the phenol UV-visible spectrum obtained from the National Institute of Standards and Technology (NIST).[17]

The ε is often recorded in units of $dm^3\ mol^{-1}\ cm^{-1}$ or $m^2\ mol^{-1}$.[1] A more intuitive parameter exists to describe the likelihood of an electronic transition, as discussed in the previous section: *f*. *f* is unitless because it is a ratio of how much energy is absorbed or emitted by the system to complete absorption or emission of energy by a single electron in the system.[2] *f* is equal to one

when the likelihood of an electronic transition to occur is maximum and f is equal to zero when there is no likelihood of an electronic transition to occur. f is related to the ε through an integral (Equation 1.4.4).[18] This relationship is also shown schematically in Figure 1.4.4.

$$f = 2303 \frac{mc_o^2}{N_a \pi e^2 n} \int \varepsilon(\tilde{v}) d\tilde{v} = 2303 \frac{mc_o^2}{N_a \pi e^2 n} (A) = \frac{4.32 \times 10^{-9}}{n} (A)$$

Equation 1.4.4: A classical mechanical expression for the oscillator strength.

In Equation 1.4.4, f is shown as the product of the mass of the electron (m) and the square of the speed of light is (c_o). This is divided by Avogadro's number (N_a), pi (π), the square of the charge of the electron (e), and the refractive index of the medium (n). f is also expressed as an integral (with respect to energy ($\tilde{v}$)) of the product of the molar extinction coefficient (ε) and the energy ($\tilde{v}$). Since terms before the integral are all known constants, these constants can be reduced to one proportionality constant, 4.32×10^{-9}. This constant has units of cm^{-1}. In this project, the refractive index n is assumed to be 1, corresponding to the gas phase.

1.5   Significance of benchmarking

Benchmarking is the process of establishing a reliable standard for which other data can be compared. With major advancements in computing methodology, quantum chemical methods, and computational chemistry algorithms, modern quantum chemical calculations are capable of producing quantitative energies that are comparable to those obtained experimentally. A milestone study that demonstrates that computational predictions could be comparable to experimental results is the Kolos and Wolniewickz study of the H_2 molecule in the 1960s.[19] This study used an 80-term electronic wavefunction in the Schrödinger equation following the variation principle to determine the dissociation energy of H_2.[20] This calculation produced a dissociation energy of H_2

(36117.4 cm^{-1}), a more accurate value than the best spectroscopically derived dissociation energy

of H$_2$ (36113.6 ± 0.6 cm^{-1}) at the time (which later was improved and shown to better match with

the computational predictions.[19] Their work not only popularized the use of quantum

computational work to compare energy parameters to those experimentally derived but also

generated momentum in the use of electronic structures of atoms and molecules. This study

inspired others alike to invest in the development and improvement of computational methods.

1.6   Related work

This work provides a benchmark model of f_{exp} for one-hundred small organic molecules.

These experimental benchmarks are then compared with oscillator strengths computed using

B3LYP/6-31+G* computational calculations. While this study is not the first to derive oscillator

strengths from experimental UV-visible spectra and/or use them to benchmark computed oscillator

strengths, what distinguishes this work is the systematic fitting approach used and the use of a

large number of molecules collected in one reference. Previously published works have provided

benchmarked f for a significantly smaller sets of molecules or benchmark their methods against

higher -level quantum mechanical methods instead of experiments.[19, 21-25] A few of these studies

are described below.

Matsuzawa et al. compared TDDFT methods to derive the excited-state energies and f for

four molecules: formaldehyde, benzene, ethylene, and methane.[21] Their study focused on a direct

comparison between f_{comp}s computed with different electronic structure methods, with no direct

comparison to experiments. Similar to Matsuzawa et al., Caricato et al. provides comparisons of

excited-state energies and f_{comp}s using different computational methods for eleven small organic

molecules.[22] Specfically, the benchmark TDDFT methods against equation-of-motion coupled

cluster with single and double excitations (EOM-CCSD), a Post-Hartree-Fock method.[22] This

comparison yielded several instances where the computations were not internally consistent. Chrayteh et al. similarly computed oscillator strengths using couple cluster methods as a reference.[23] Several other studies have also avoided comparing f_{comp} against f_{exp}, typically citing difficulty to extract strengths from experimental data due to "line broadening and overlapping of excitation bands"[22, 23] which highlights the importance of the work presented in this thesis.

A few recent papers, however, have started to compare computational and experimental f, directly. Jacquemin et al. (2016) compared f_{comp} from Bethe-Salpeter (BSE/GW) methods with TDDFT methods.[25] While Jacquemin et al. (2016) computed oscillator strengths for approximately 200 molecules, only 30 were compared with experimentally derived f of these same molecules; this is most likely due to the scarcity of reported f_{exp} in experimental studies.[25] The experimental protocol of our project derives f_{exp} from carefully integrating the extinction coefficient from UV-visible spectra.

Therefore, current oscillator strength benchmarks are not as extensive as typical benchmarks for other properties. Consider, for example, benchmark studies of excitation energies; One example by Jacquemin et al. (2009) provided a comparison of TDDFT methods against one another and against experimentally derived excitation energies for approximately 500 bio-organic compounds and dyes.[24] This is significantly larger than the current largest benchmarks of oscillator strengths.

Methods for the calculation of the f are readily available and widely implemented in multiple quantum chemical software packages and for many methods. However, the validation of those computations against experimental absorption strengths is complicated by the fact that f is not usually reported in experimental literature.[19] When they are, experimental studies do not always describe how they fit the spectra to obtain f_{exp}, thus making it unclear what error may be

introduced by the fitting. Therefore, in this work, f_{exp} is obtained using a systematic protocol starting from UV-visible spectra reported in literature. Specifically, to be as consistent and systematic as possible, we obtain all spectra from a single source, the UV Atlas of organic compounds[26], which clearly reports experimental details including the solvent and concentration at which the spectra were taken. The UV Atlas is one of the resources used by the NIST UV-visible online database.[27] A specific protocol is then used to integrate these spectra and obtain experimental f values while implementing a series of tests to identify sources of error and quantifying these errors. This approach also addresses variations in broadening and overlapping peaks inherent to molecular absorption spectra. To our knowledge, this would be the largest collection of oscillator strengths for such molecules which are generated from a single source.

2   EXPERIMENTAL DETAILS

2.1   Overview

To derive oscillator strengths, we follow the protocol shown in Figure (2.1.5). Specifically, we digitize the spectra (**B**) and perform both numerical integration following the Midpoint rule (**C**) as well as a deconvolution using Gaussian curves to get the area under each peak (**D**). Gaussian curves are used for fitting the spectra because the shape of Gaussian curves, characterized as smooth, symmetric, and bell-like, are similar the shape of electronic molecular spectra (a consequence of the distribution of thermal motion). However, a single Gaussian fit for each peak is rarely sufficient, leading to some inaccuracy (see for instance **E1** and **E2**). Overlapping peaks also complicate the fitting. In those cases, we initially fit the overall spectrum using as few Gaussian curves as possible. We then truncate each peak and fit it while accounting for a constant spillover from the other peak (red regions in **E1** and **E2**). We keep increasing the number of Gaussian curves used to fit each peak until convergence (**F1** and **F2**), i.e., until the error in the root mean square error (RMSE) of the fit relative to the range of molar absorptivities is below a certain threshold. This entire process is then repeated from scratch to ensure the reproducibility of the fitting. As discussed later, several criteria are used to check the fitting quality: the grid fineness of the digitized spectra, the RMSE from the least-squares fitting, comparison of the Midpoint rule and Gaussian curve integration, and comparison of multiple trials to check for consistency.
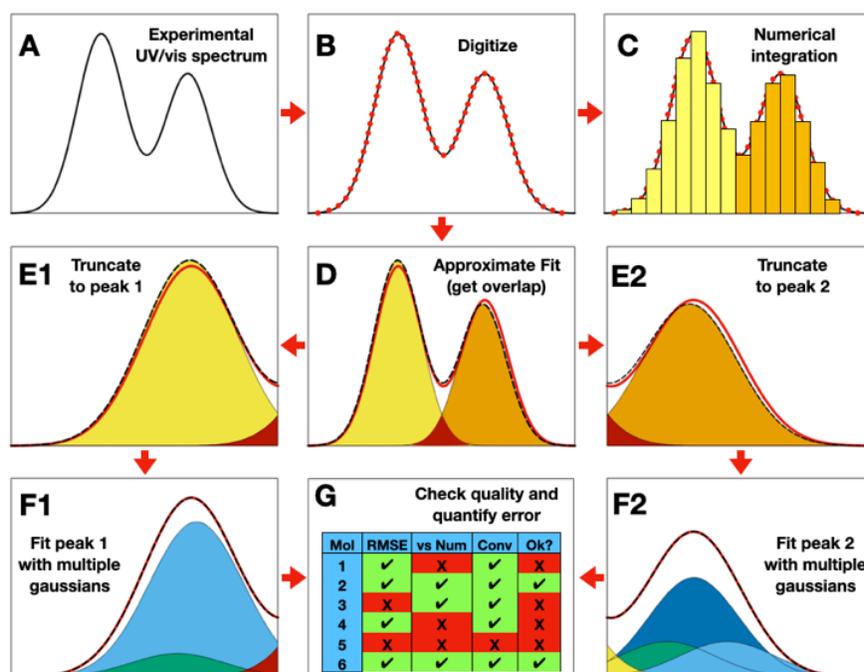
Figure 2.1.5: Overview of experimental methods.

2.2   Materials

To execute the analytical methods, all UV-visible spectra are taken from the *UV Atlas of Organic Compounds*.[26] Webplot Digitizer is the platform used to digitize the UV-visible spectral curves.[28] The Macintosh application, Numbers, is used to view the csv files containing digitized spectral data. All data, experimental and computational, is tabulated in the Microsoft application, Excel. In Excel, the plugin, Excel Solver, is used for the fitting of the digitized UV-visible spectra.

To perform the computational methods, IQmol, a free open-source molecular editor and visualization software, is the application used to digitize the molecular models of all organic compounds.[29] The Macintosh application, Terminal, acts as a communication window to a supercomputer. Both the SDSC Comet supercomputer and a local Georgia State University (GSU) supercomputer, PHOTON, were used. The quantum chemical software Gaussian is employed to

run the optimization, frequency, and excited-state calculations. Gaussian provides a variety of computational methods and basis sets that are readily available for use.[30] The computational method used is B3LYP[7, 10], a hybrid DFT method with Hartree-Fock exchange, and the basis set used is 6-31+G*. Additionally, a solvent model, the polarizable continuum model (PCM)[31], is used to account for the dielectric constant of the solvent during the optimization, frequency calculation, and excited-state calculations for each molecule. The dielectric constant used for each molecule is consistent with the solvents reported for the experiments in the UV-visible Atlas.

To visualize the digitized and fitted UV-visible spectra and f_{comp}, Gnuplot, a command-line driven graphing utility, is used.[32] All organic molecules' structures are drawn in the Perkin Elmer application, Chemdraw, a molecule editor software.

2.3 Analytical methods

Each molecule and their respective solvents are recorded from the *UV Atlas of Organic Compounds* with an ID of 000 through 108. This is tabulated in Excel (Table 2.3.1).

Table 2.3.1: Identification of molecules.

Molecule ID	Name	Solvent
000	cis -1, 2-Dichloroethylene	Heptane
001	trans-1,2-Dichloroethylene	Heptane
002	Trichloroethylene	Heptane
003	Tetrachloroethylene	Heptane
004	Penta-1,3-diene	Heptane
005	2,5-dimethylhexa-2,4-diene	Heptane
006	Cycloocta-1,3-diene	Hexane
007	1-(cyclohexen-1-yl)cyclohexene	Heptane
008	cis-Hex-3-en-1-yne	Ethanol
009	trans-Hex-3-en-1-yne	Ethanol
010	1,6-Dihydroazulene	Hexane
011	p-Benzoquinone	Hexane
012	Hydroxy-p-benzoquinone	CCl4

013	Methoxy-p-benzoquinone	Hexane
014	2-Sulfinylhexahydro-1,3-benzodithiole	Ethanol
015	Thiourea	Water
016	S-ethyl thioacetate	Heptane
017	Nitromethane	Hexane
018	t-Butyl nitrite	Hexane
019	Isopentyl nitrite	Hexane
020	N-Benzylideneaniline	Ethanol
021	4- Nitro-N-benzylideneaniline	Ethanol
022	N-Benzylidene-4-dimethylaminoaniline	Ethanol
023	Tetramethylazodicarboxamide	Dioxane
024	Phenylazoformate	Ethanol
025	Dimethyl [(E)-phenyldiazenyl]phosphonate	Dioxane
026	trans-azobenzene	Hexane
027	4-(Dimethylamino)azobenzene cation	H2SO4 (aq)
028	4-Dimethylamino-4'-nitroazobenzene	Ethanol
029	1-Ethyl-2-(Phenylazo)-pyridinium cation	Acetonitrile
030	1-Phenyl-2-(phenylthio)diazene	Dioxane
031	1-Phenyl-2-(phenylsulfonyl)diazene cation	Dioxane
032	N,N,N′,N′-Tetramethylformamidinium	Ethanol
033	N-[3-(Dimethylamino)allylidene]-N-methylmethanaminium	Ethanol
034	2,4-Cyclopentadien-1-ylidenehydrazine	Hexane
035	Cyclohexanone semicarbazone	Ethanol
036	1-Isopropylidene-2-p-nitrophenyl-hydrazine	Ethanol
037	1-ethylidene-2-(2,4-dinitrophenyl)-hydrazine	CHCl3/Ethanol
038	1-ethylidene-2-(2,4-dinitrophenyl)-hydrazone anion	CHCl3/Ethanol
039	Crotonaldehyde 2,4-dinitrophenylhydrazone	CHCl3/Ethanol
040	N,N'-di(2-hydroxybenzylidene)hydrazine	Methanol
041	o-Xylene	Heptane
042	m-Xylene	Heptane
043	p-Xylene	Heptane
044	2,3,5,6,7,8-hexahydro-1H-benz(f)indene	Petroleum
045	1,2,3,4,5,6,7,8-octahydroanthracene	Petroleum
046	Phenol	Water
047	Phenolate	NaOH (aq)
048	N-diethylaniline	Heptane
049	Nitrobenzene	Petroleum
050	o-chlorotoluene	Heptane

051	p-chlorotoluene	Heptane
052	p-bromofluorobenzene	Heptane
053	o-Nitrophenol	Water
054	m-Nitrophenol	Water
055	p-Nitrophenol	Water
056	p-Nitrophenolate	0.01 M NaOH
057	o-Acetylphenol	Ethanol
058	o-Methoxyacetophenone	Ethanol
059	o-Hydroxybenzaldehyde	Ethanol
060	o-Methoxybenzaldehyde	Ethanol
061	o-aminobenzaldehyde	Ethanol
062	p-Hydroxybenzaldehyde	Ethanol
063	p-Dimethylaminobenzaldehyde	Cyclohexane
064	p-cyanobenzaldehyde	Cyclohexane
065	potassium hydrogen phthalate	Water
066	p-Nitrophenylhydrazine	Methanol
067	2,3-xylenol anion	0.025 M KOH
068	4-Methyl-2-nitroaniline	Methanol
069	Cinnamaldehyde	Methanol
070	Benzylideneacetone	Ethanol
071	Benzophenone	Ethanol
072	4-Methylbenzophenone	Cyclohexane
073	4-Aminobenzophenone	Ethanol
074	4-Dimethylaminobenzophenone	Cyclohexane
075	4-Hydroxybenzophenone	Hexane
076	Diphenylamine	Ethanol
077	2-Nitrodiphenylamine	Methanol
078	4-Nitrodiphenylamine	Methanol
079	Thiobenzophenone	Methanol
080	1-Naphthol	Methanol
081	1-Naphthol anion	0.1 M KOH
082	2-Naphthol anion	0.1 M KOH
083	1-Naphthoic acid	Methanol
084	2-Naphthaldehyde	Methanol
085	1,5-Dinitronaphthalene	Methanol
086	Thiazole	Heptane
087	Thiazoline-2-thione	Water
088	Pyridinium	H2SO4 (aq)
089	2-Methoxypyridine	Heptane

090	2-chloro-6-ethoxypyridine	Heptane
091	cis-1,2-Di-3'-pyridylethylene	Heptane
092	2-Hydroxyquinoline anion	1 M NaOH
093	4-Hydroxyquinolynium	0.1 M HCl
094	4-Hydroxyquinolyne anion	0.01 M KOH
095	5-Hydroxyquinoline	Methanol
096	5-Hydroxyquinolynium	0.1 M HCl
097	5-Hydroxyquinolyne anion	0.1 M NaOH
098	3-Hydroxyphthalamide	Ethanol
099	3-Amino-1-methylphthalamide	Ethanol

2.3.1   Digitization of spectral curves

All UV-visible spectra in the UV Atlas show the x-axis in units of $\tilde{v}$ and the y-axis in units

of ε.[26] Webplot Digitizer is used to digitize the spectral curves of molecules 000 through 108.

Digitization is the procedure of translating human-readable data (e.g., an image) into a language

that can be processed by a computer. In this case, digitization involves automated detection of

points on the spectral curve in which each point is characterized by a specific x-value, $\tilde{v}$, and a y-

value, ε. The first step in digitization with Webplot Digitizer is to calibrate the x and y axes. For

example, the upper and lower bounds are chosen on the x and y axis for *Cis* -1, 2Dichloroethylene

(Figure 2.3.1.6).[28] To simplify the following discussion, *Cis* -1, 2-Dichloroethylene is referred to

by its ID, 000.
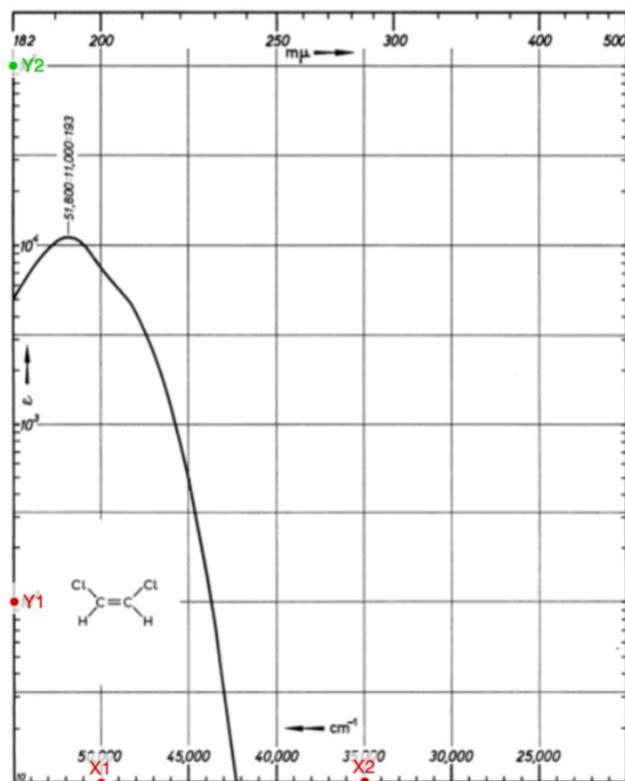
Figure 2.3.1.6: Choosing two bounds on the x and y axes of the UV-visible spectrum of

000 in Webplot Digitizer to calibrate the axes.

To complete calibration of the x and y axes, the values of the bounds are entered in the "X and Y Axes Calibration" window (Figure 2.3.1.7).[28] Note that the extinction coefficients are plotted on a logoritmic scale.
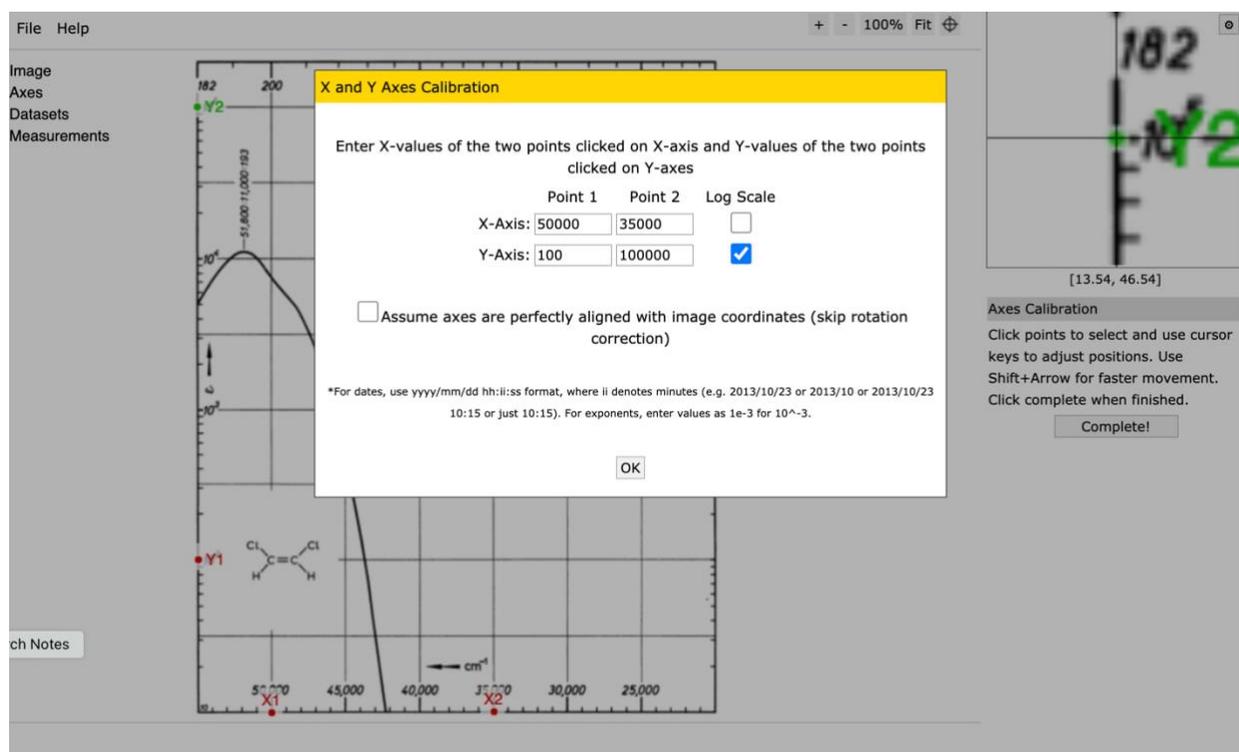
Figure 2.3.1.7: "X and Y Axes Calibration" window in Webplot Digitizer.

Next, the spectral curve is captured using the tools under "Automatic Extraction" (highlighted in pink) (Figure 2.3.1.8).[28] Black is chosen as the "Foreground Color" because our spectral curve is black against a white background. Delta X and Delta Y indicate the number of pixels between two points along the x and y axes respectively. Delta X and Delta Y are set to 5 Px. Lastly, the pen tool is used to trace the spectral curve. The width on the pen can be adjusted.
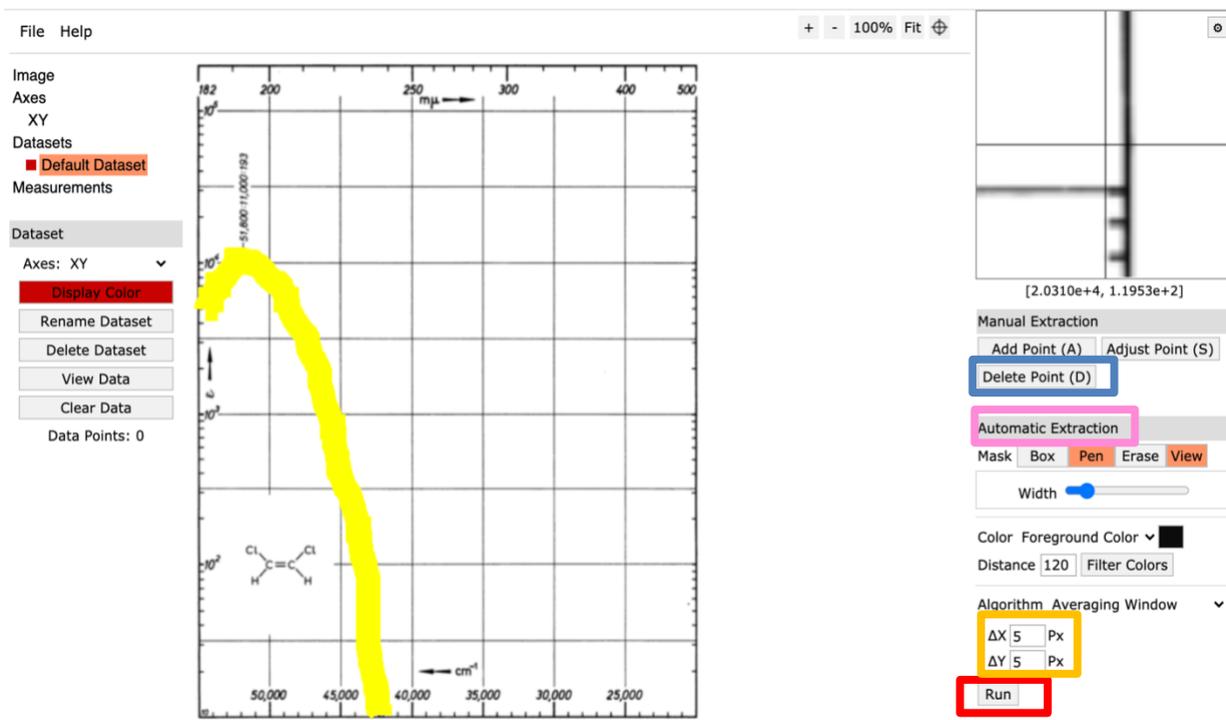
Figure 2.3.1.8: "Automatic Extraction" tools in Webplot Digitizer.

The detected points on the spectral curve are viewed once the "Run" option (highlighted in red) is chosen (Figure 2.3.1.8).[28] Using the "Delete Point (D)" option (highlighted in blue), points that do not lie directly on the spectral curve are deleted. The x and y values of each detected point are listed with the "View Data" option (highlighted in green). (Figure 2.3.1.9).[28]
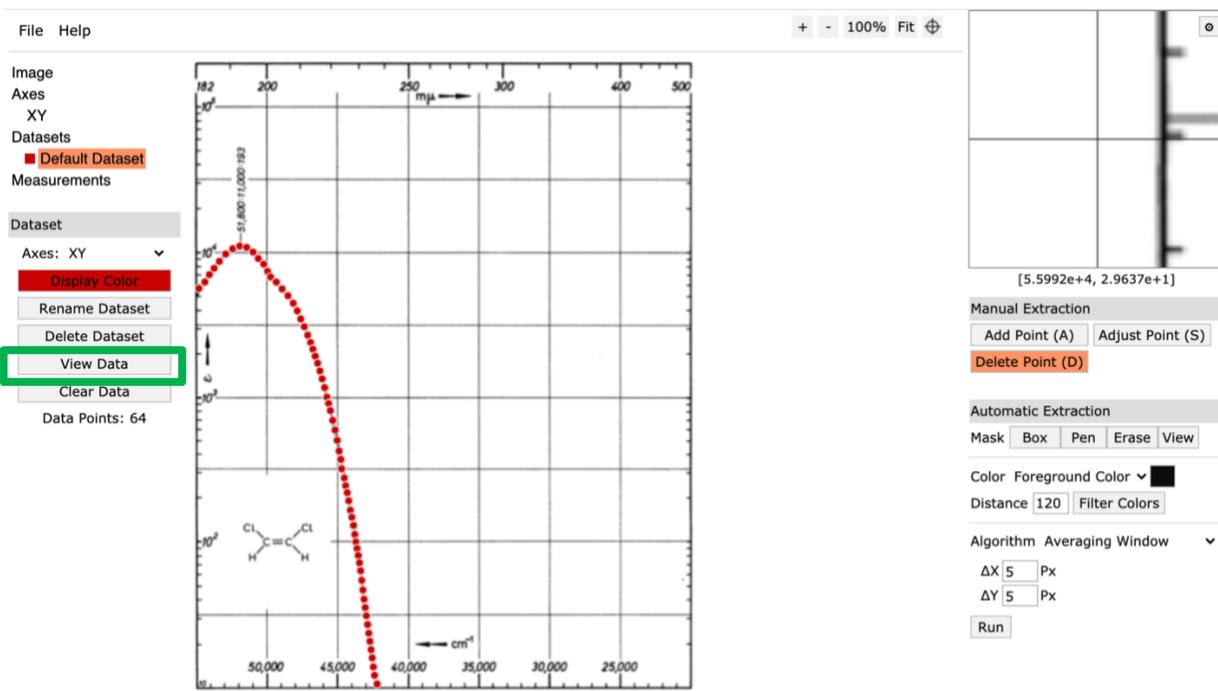
Figure 2.3.1.10: Points detected on the spectral curve in the UV-visible spectrum of 000 in Webplot Digitizer.
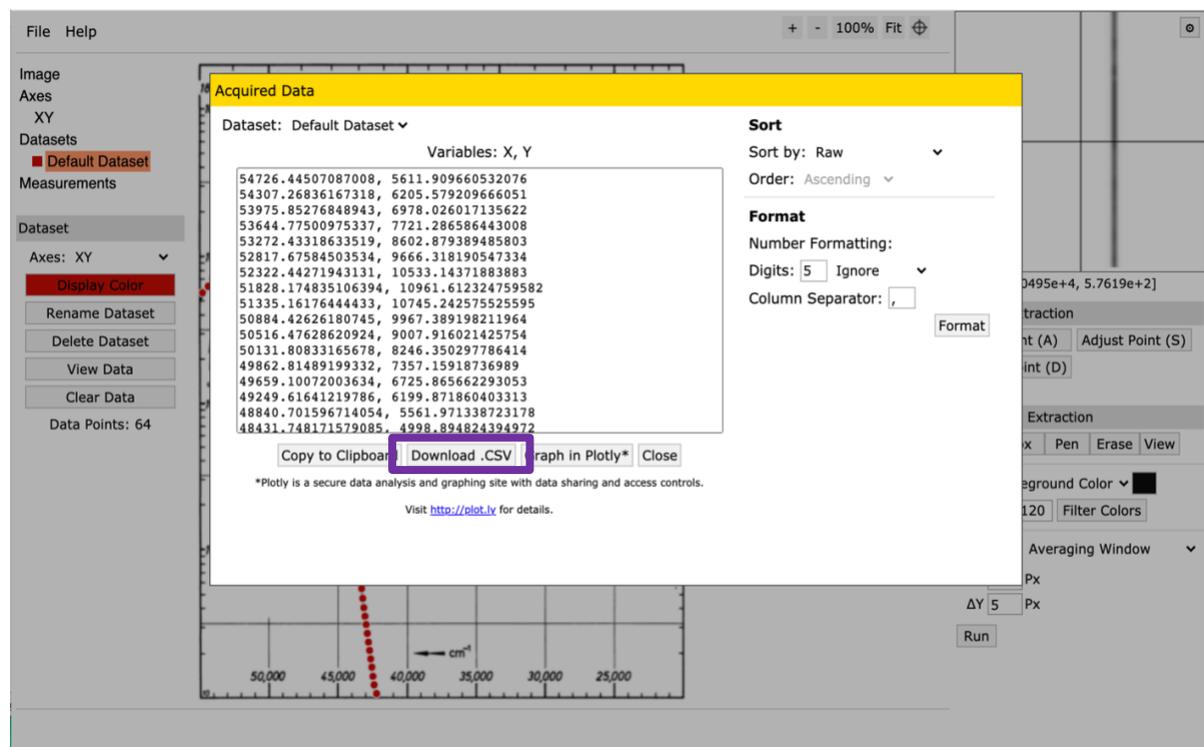
Figure 2.3.1.9: "Acquired Data" window in Webplot Digitizer.

This data is downloaded by choosing the "Download .CSV" option (highlighted in purple) in the "Acquired Data" window (Figure 2.3.1.10).[28] The csv file is opened in Numbers and the file content is copied and pasted into Excel, where the spectral curve of 000 is plotted (Figure 2.3.2.12). Digitization and plotting of the UV-visible spectrum of 000 is repeated following the procedure above exactly, except Delta X and Delta Y (highlighted in orange) are set to 3 Px after the spectral curve is captured to get a higher resolution for the digitization (Figure 2.3.1.8). So far, two csv files are generated for each molecule's spectral curve. The first csv file contains points separated by five pixels in the x and y direction and the second csv file contains points separated by three pixels in the x and y direction, thus these points are composed in a five-point grid and a three-point grid, respectively (Figure 2.3.1.11). All Excel spectra generated from the five-point grid are grouped and separated from all Excel spectra generated from the three-point grid.

Five-point Grid	**Three-point Grid**
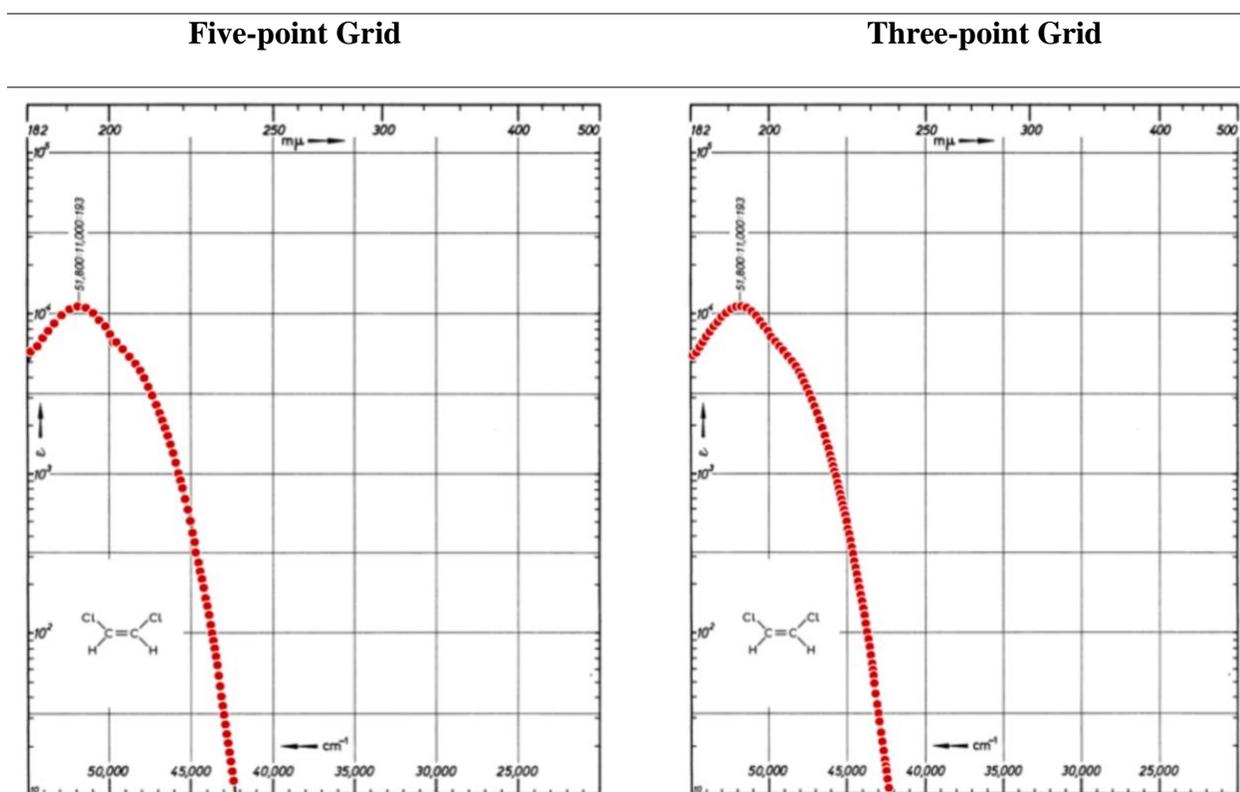
Figure 2.3.1.11: Digitized UV-visible Spectra of 000 with the five-point grid (left) and the three-point grid (right) generated in Webplot Digitizer.

A third set of csv files of digitized spectra in variable grids is generated. For the variable-grids, the images of the spectra used in digitization are of higher resolution than those images used to generate the five-point grid and three-point grid sets above. The procedures for the digitization, fitting, and error analysis used for the variable-grids are identical to those discussed for the five-point grid and three-point grid sets.

The fineness of the grid is quantified by determining the product of the range divided by the count (range/count) for each generated spectrum. The count is the number of points generated by Webplot Digitizer to digitize the spectral curve. The range is the span of $\tilde{v}$ in which a spectral

curve exists. The range/count is tabulated for all molecules' spectra for the five-point grid, the three-point grid, and the variable-grid in Excel (see in Appendix B.2). The variable-grid set typically has the highest fineness (range/count) out of the three sets, since the number point points used in this set was varied in such a way to ensure that the range/count remains under 100 for all molecules (i.e., to ensure the fineness passes some minimum threshold).

2.3.2 The Gaussian curves
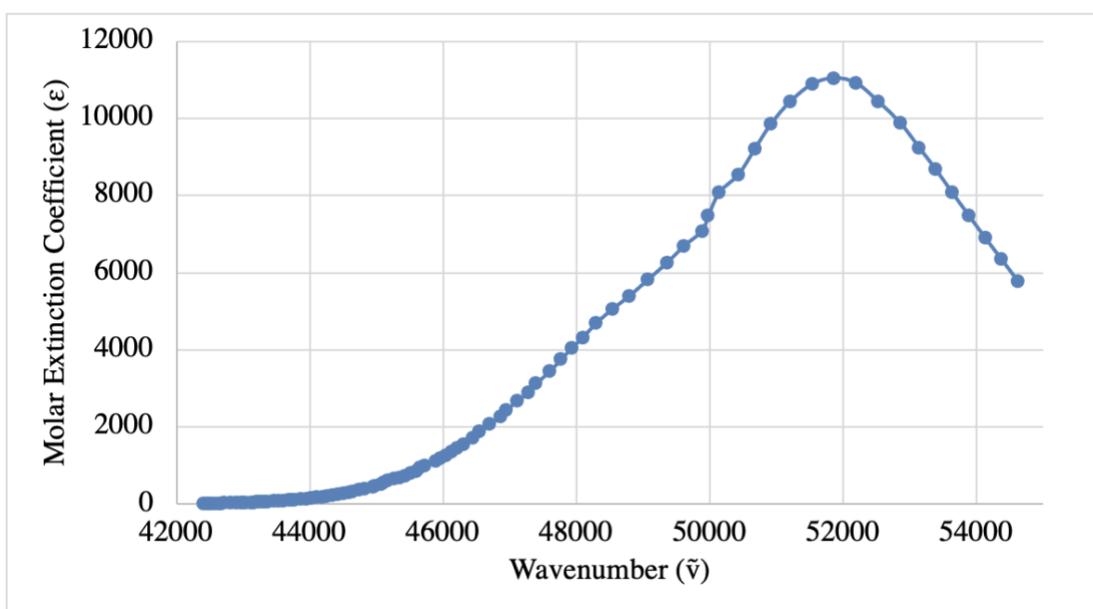
Figure 2.3.2.12: Spectral curve of 000 plotted in Excel with points in the three-point grid.

In Figure 2.3.2.12, note that this spectrum is no longer plotted along a logarithmic scale, and therefore, the spacing between the points are not evening distributed.

With Excel Solver, one Gaussian curve is approximately fitted to the spectral curve by minimizing the root-mean square error (RMSE) (Figure 2.3.2.13).
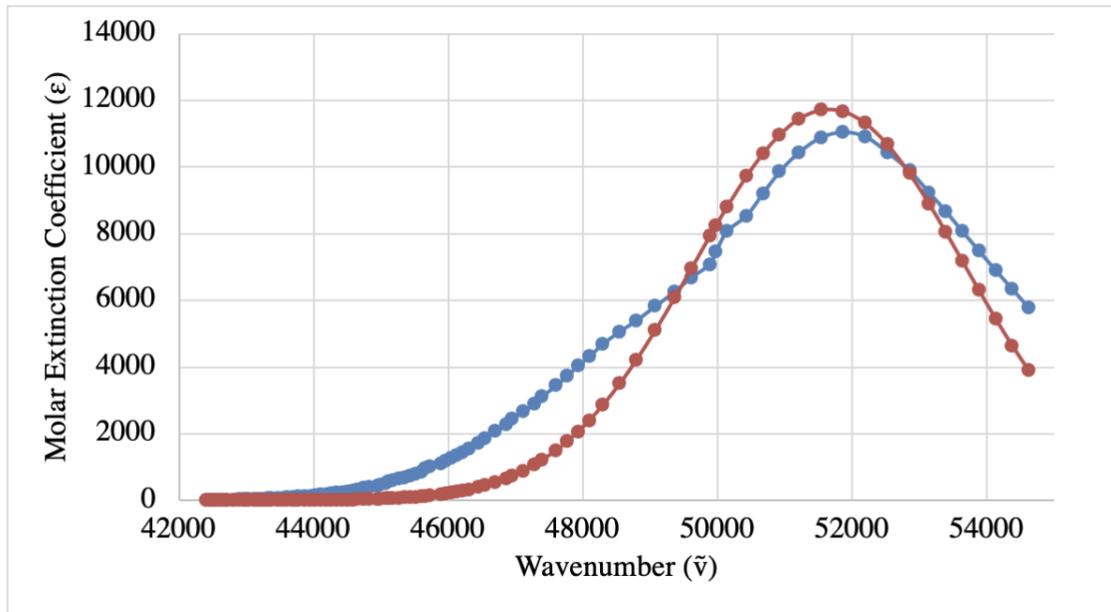
Figure 2.3.2.13: Spectral curve of 000 (blue) fitted with a Gaussian curve (red).

The error is a consequence of the difference between ε corresponding to a point on the spectral curve ($\hat{y}_j$) and the respective ε contributed by the Gaussian curve (y_j) (Equation 2.3.2.5).[33] The RMSE attributes significantly more weight to large error compared to small error due to the squaring function performed on the error before an average is taken.[34]

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Equation 2.3.2.5: Mathematical expression for root-mean square error.

In Equation 2.3.2.5, the number of non-missing data points is n, and the variable is j. A sum, with respect to n and j, of the squared difference of is the observed value (y_j) and the

expected value ($\hat{y}_j$) is taken. RMSE is found when the square-root is applied to the quotient of this sum and n.

The RMSE is preferred over other widely used statistical measures because the former allows large errors to be easily recognized. A normalized RMSE (NRMSE) is defined by dividing the RMSE by the range of ε (Equation 2.3.2.6).[35] The range of ε is the difference between the maximum ε and the minimum ε of the respective molecule's spectrum and normalizes RMSE by confining it to the space in which the error exists. The NRMSE is expressed as a percentage and is a more intuitive parameter for understanding the error relative to the intensity of the absorption of the peak. This accounts for the fact that for more intense peaks, the RMSE is naturally going to be larger than for less intense ones.

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}$$

Equation 2.3.2.6: Mathematical expression for the normalized root-mean square error.

In Equation 2.3.2.6, the maximum ε is y_{max} and the minimum ε is y_{min} of the molecule's spectrum, respectively.

Clearly, a single Gaussian curve does not fit the spectrum 000 well, as shown in Figure 2.3.2.13. To minimize the NRMSE, additional Gaussian curves are fitted to the spectral curve of 000 (Figure 2.3.2.14 and Table 2.3.2.2). With multiple Gaussian curves, the sum of the relative areas of each Gaussian curve is used to find the oscillator strengths (f_{exp}) (Equation 1.4.4). The relative area under the Gaussian curve is then determined by finding the total area of all Gaussian curves used in the fitting, which is equivalent to integrating the molar absorptivity with respect to change in ṽ ($\int \varepsilon(\bar{v})d\bar{v}$). The corresponding most optimal oscillator strength associated with the

smallest NRMSE is recorded in Excel for the five-point grid and the three-point grid (see in

Appendix B.2).
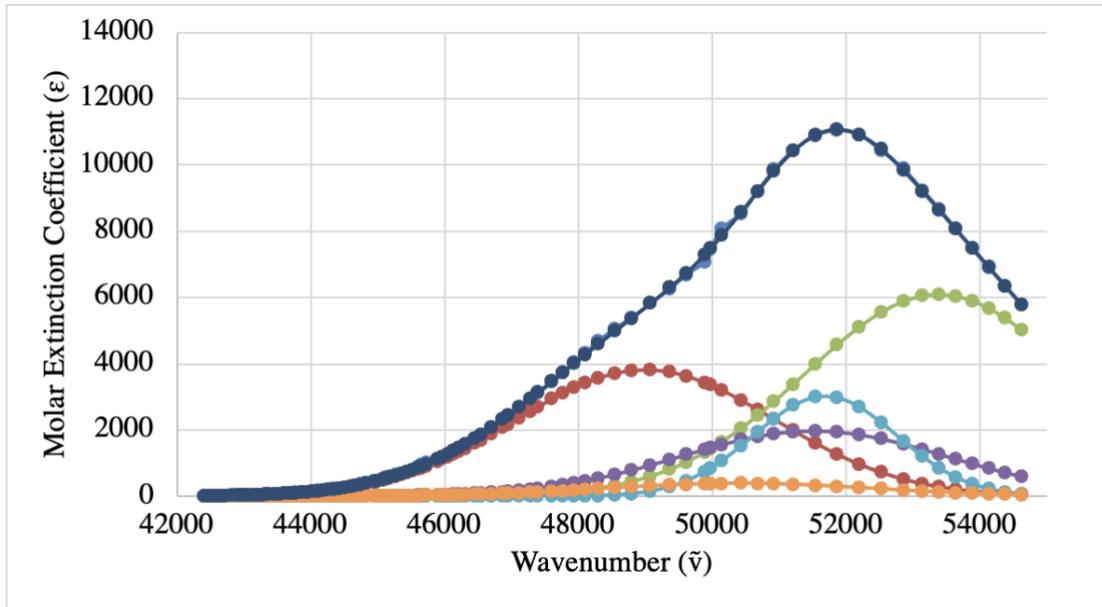
Figure 2.3.2.14: Spectral curve of 000 fitted with the optimal number of gaussian

curves.

In Figure 2.3.2.14, the spectral curve of 000 is fitted with the several gaussian curves until

convergence is low. This is determined by NRMSE and by the change in total oscillator strength

as additional Gaussian curves are added. Here, five Gaussian peaks are used. The Gaussian curves

are shown in light blue, green, yellow, orange, and gray (Figure 2.3.2.14). The total fit to the

spectral curve is navy blue. The original digitized spectral curve is in regular blue (but is mostly

hidden under the navy blue curve due to the very good fit).

Table 2.3.2.2: Analysis of Gaussian curve fitting to the spectral curve of 000.

Number of Gaussian Curves (n)	Oscillator Strength (f_{exp})	RMSE	NRMSE (%)
1	0.254	907.6	8.23
2	0.287	130.8	1.19
3	0.291	107.3	0.97
4	0.300	39.8	0.36
5	0.298	39.1	0.35

For compounds showing spectral curves with multiple peaks (e.g., see Figure 2.3.2.15), each well-defined peak provides a distinct f_{exp} value for a different state. Thus, in order to resolve the values of the f_{exp}, each well-defined peak is isolated from the rest of the spectral curve. This is done firstly by fitting the spectral curve with the least number of Gaussian curves to later account for any overlap. Again, using Excel Solver, these Gaussian curves are fitted to minimize the NRMSE. The example shown in Figure 2.3.2.15 is Methoxy-p-benzoquinone. This molecule has a spectral curve with two well-defined peaks. To simplify the proceeding discussion, Methoxy-p-benzoquinone is hereon referred to by its ID, 013.
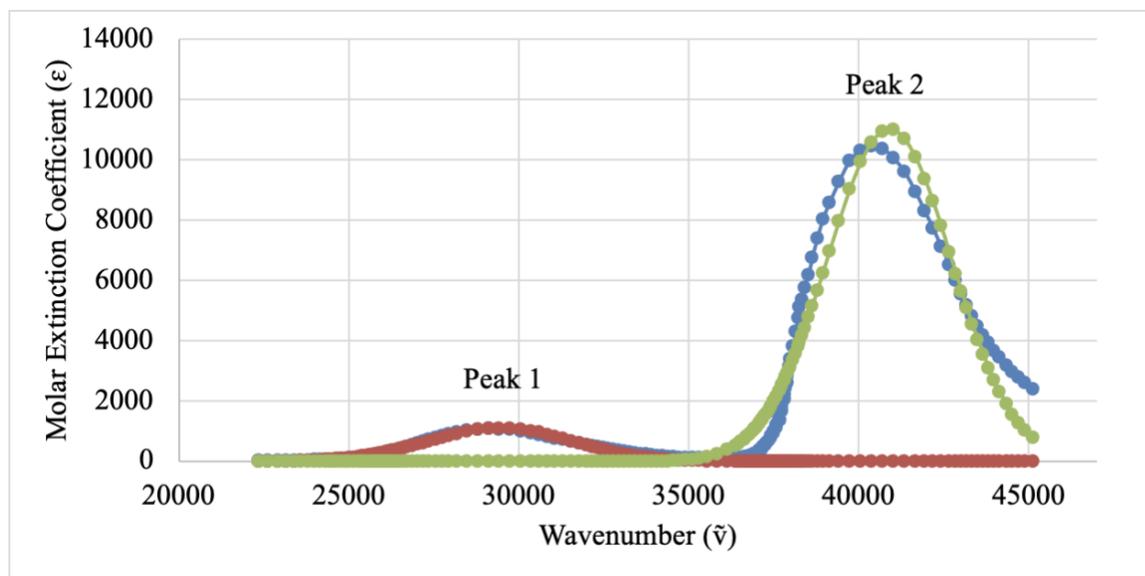
Figure 2.3.2.15: Spectral curve of 013 with a Gaussian curve fitted to each well-defined peak.

In Figure 2.3.2.15, the spectral curve is shown in blue. The Gaussian curve mainly fitting peak 1 is red and the Gaussian curve mainly fitting peak 2 is green.

Then, Peak 1 is isolated from the spectral curve (Figure 2.3.2.16). While the relative area under the green Gaussian curve is not factored into the calculation for the f_{exp} associated with peak 1, it is essential to account for how peak 2's intensity has spilled into peak 1 (Figure 2.3.2.16). The relative area under the orange Gaussian curve is used to calculate the f_{exp} of peak 1 (Equation 1.4.4).
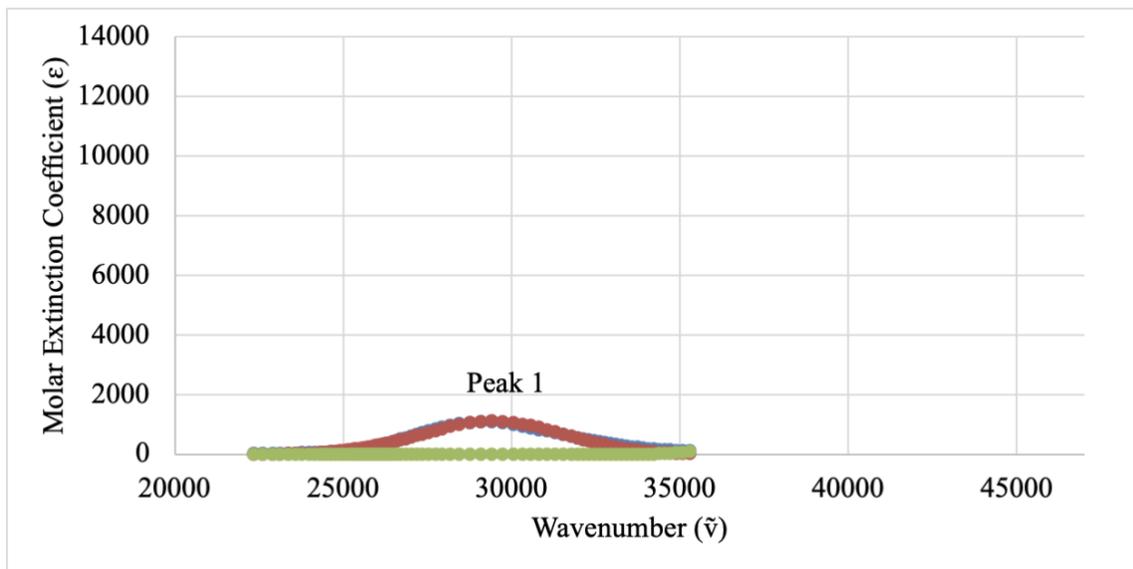
Figure 2.3.2.16: Peak 1 of the spectral curve of 013.

In Figure 2.3.2.16, the spectral curve is blue. The Gaussian curve mainly fitting peak 1 is red. The Gaussian curve mainly fitting peak 2 is green.

Additional Gaussian curves are then fitted to peak 1 as usual (Figure 2.3.2.17). However, the main difference is that this is now done in the presence of the peak 2 tail, which is kept frozen during the fitting. Gaussian curve fitting becomes more optimal as the NRMSE decreases (Table 2.3.2.3). The f_{exp} associated with the smallest NRMSE is recorded in Excel.
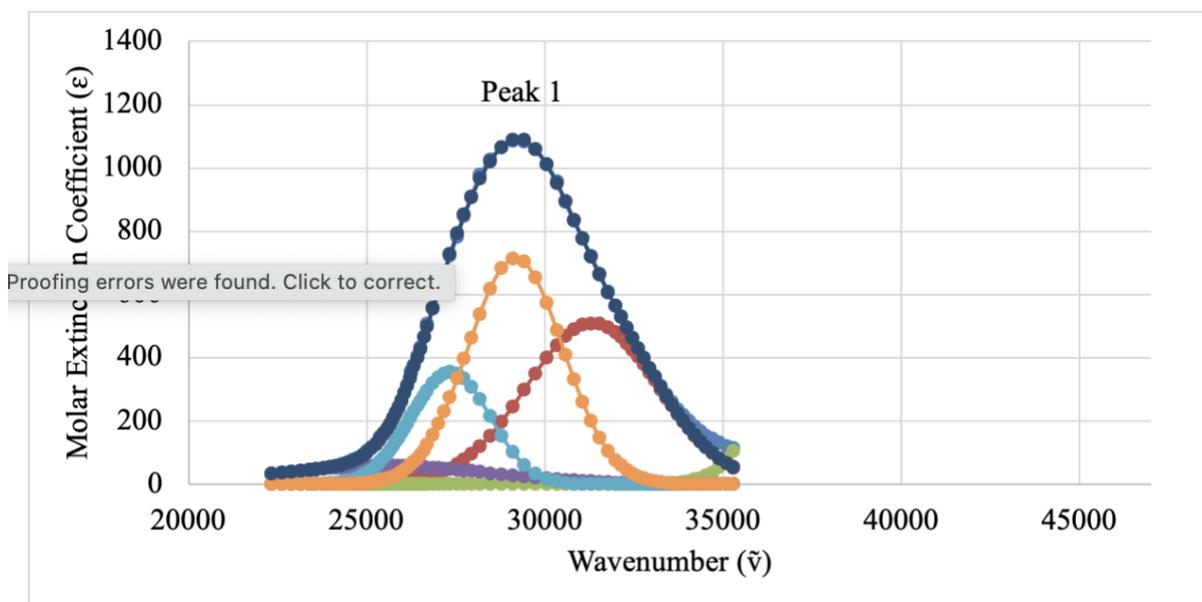
Figure 2.3.2.17: Peak 1 of the spectral curve of 013 fitted with the optimal number of

Gaussian curves.

In Figure 2.3.2.17, the spectral line is blue. The Gaussian curves mainly fitting peak 1 are orange, purple, light blue, and red. The tail of the Gaussian curve mainly fitting peak 2 is green, and is not included in the calculation of f_{exp}. The navy-blue curve is the total fit to spectral curve.

Table 2.3.2.3: Analysis of Gaussian curve fitting peak 1 of the spectral curve of 013.

Number of Gaussian Curves (n)	Oscillator Strength (f_{exp})	RMSE	NRMSE (%)
1	0.025	50.0	4.75
2	0.027	25.5	2.42
3	0.027	11.4	1.08
4	0.027	7.8	0.74

Next, Peak 2 is isolated from the spectral curve (Figure 2.3.2.18). Again, while the relative area of the red Gaussian curve is not factored into the calculation for the f_{exp} of peak 2, it is there

to account for the overlap of the peaks (Figure 2.3.2.18). The relative area under the green Gaussian

curve is used to calculate the f_{exp} of peak 2 (Equation 1.4.4).
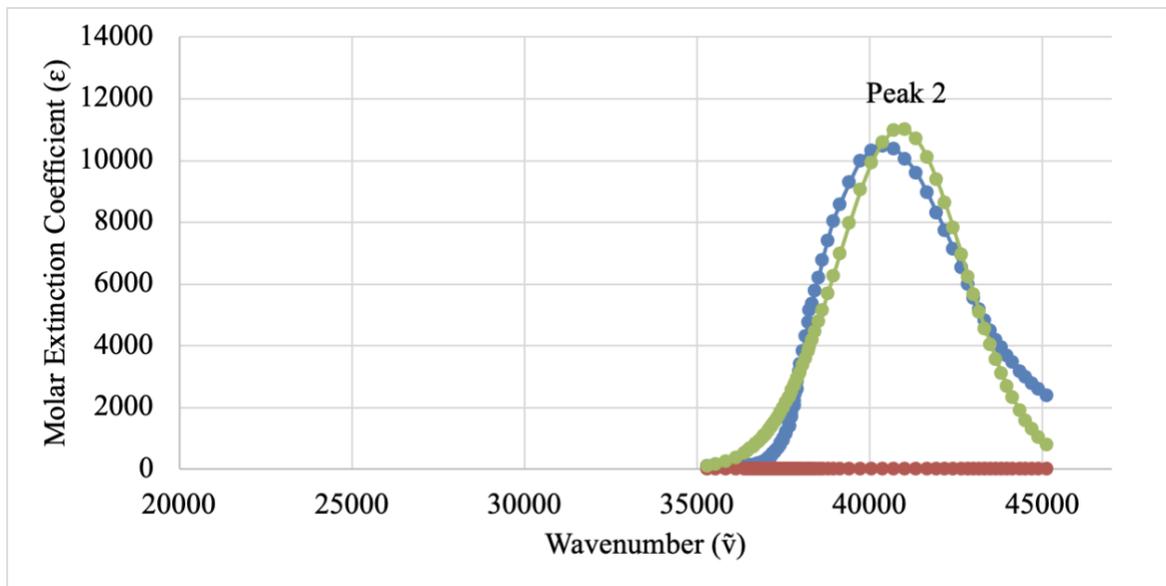
Figure 2.3.2.18: Peak 2 of the spectral curve of 013.

In Figure 2.3.2.18, the spectral curve is blue. The Gaussian curve mainly fitting peak 1 is

red. The Gaussian curve mainly fitting peak 2 is green.

Additional Gaussian curves are fitted to peak 2 (Figure 2.3.2.19). In this case, the

contribution of the red curve from peak 1 to peak 2 is minimal, but it is still accounted for and

excluded from the calculation of f_{exp} for peak 2.
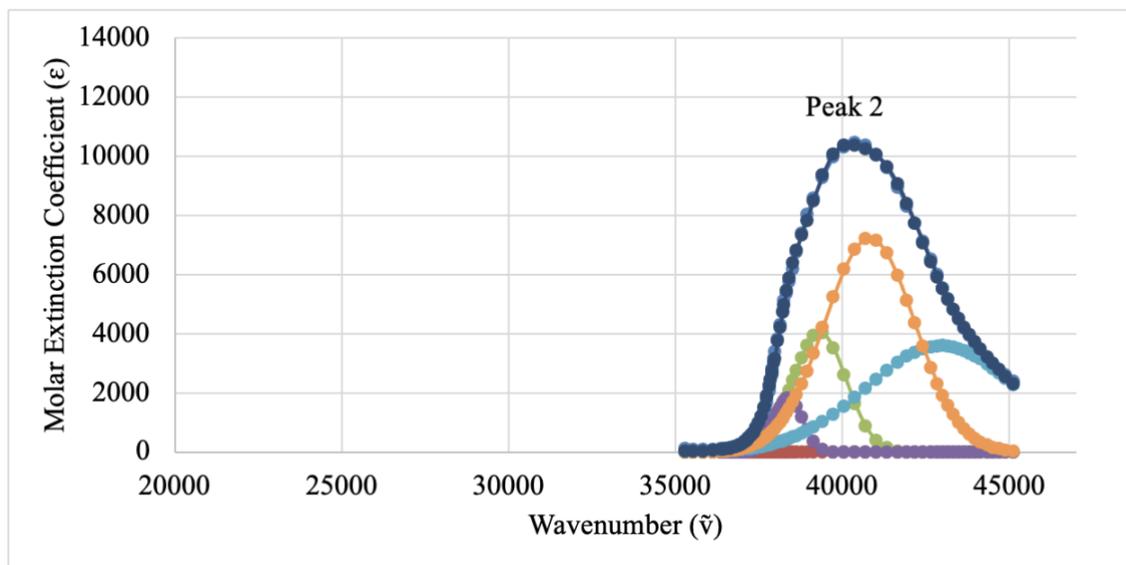
Figure 2.3.2.19: Peak 2 of the spectral curve of 013 fitted with the optimal number of

Gaussian curves.

In Figure 2.3.2.19, the spectral line is blue. The Gaussian curves mainly fitting peak 2 are purple, green, light blue, and orange. The Gaussian curve mainly fitting peak 2 is red. The navy-blue curve is the total fit to spectral curve. Gaussian curve fitting becomes more optimal as the NRMSE decreases (Table 2.3.2.4). The f_{exp} associated with the smallest NRMSE is recorded in Excel.

Table 2.3.2.4: Analysis of Gaussian curve fitting to the peak 2 of the spectral curve of

013.

Number of Gaussian Curves (n)	Oscillator Strength (f_{exp})	RMSE	NRMSE (%)
1	0.219	925.1	8.91
2	0.226	414.6	4.00
3	0.231	165.9	1.60
4	0.236	80.7	0.78

2.3.3   Numerical integration

The use of the Gaussian curves is hypothesized to be the most compatible fitting and integration method of the molecule's spectra because the molecule's spectra are Gaussian-like. To get an idea about how well the fitted peaks account for overlaps with other peaks, we also perform numerical integration of each peak. If a peak is well isolated from other peaks and appears fully in the spectrum, the numerical and Gaussian fit integrations should be very similar. If two peaks overlap but those two peaks are almost identical in shape and intensity, then their relative contributions in the overlap region are similar and, again, numerical and Gaussian integration should give similar results. However, when the numerical and Gaussian integrations are different, this is an indication that part of the peak is missing (not included in the spectrum) or of an unequal contribution of two peaks to an overlap region. In conclusion, the comparison of analytical (Gaussian fit) and numerical integrations gives an important indication about the quality and trustworthiness of the fits.

The numerical integration was performed using the midpoint rule. The midpoint rule uses a finite number of rectangles to approximately fit the total area of the spectral curve. Each rectangle has a width equal to the distance between two digitized points on the spectral curve along the x-

axis and a height equal to the average value of two digitized points on the spectral curve along the y-axis (Figure 2.3.3.20).[36]
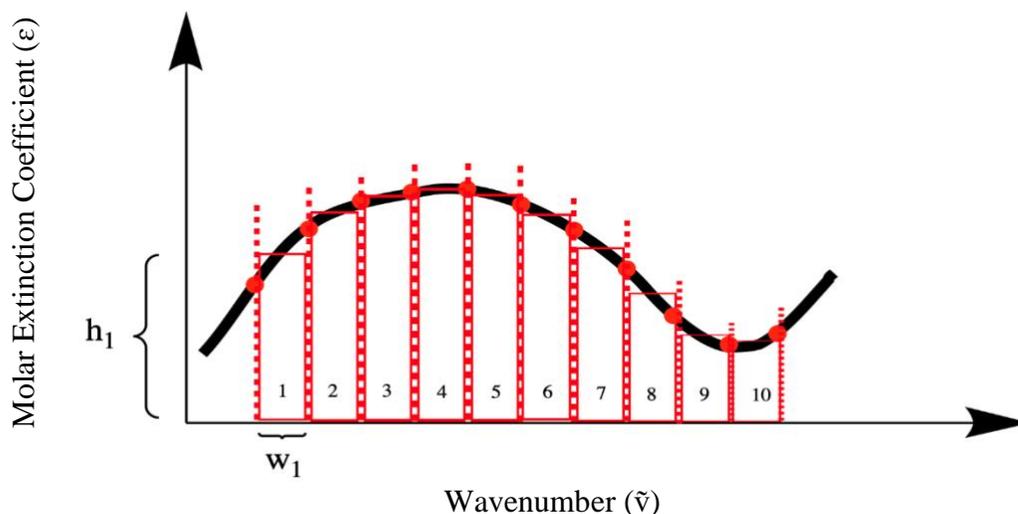
Figure 2.3.3.20: Illustration of the Midpoint rule.

The relative area of one rectangle is the product of the width and the height. The sum of the relative areas of all rectangles fitted under the spectral curve equals the total relative area under the spectral curve. By knowing the relative area under the spectral curve, the oscillator strength is computed using the midpoint rule fit ($f_{m,exp}$) and recorded in Excel (Equation 1.4.4).

The percent error of the f_{exp} against the $f_{m,exp}$ is determined (Equation 2.3.3.7).

$$Percent\ error\ (\%) = \left| \frac{f_{m,exp} - f_{exp}}{f_{exp}} \right| \bullet (100)$$

Equation 2.3.3.7: Mathematical expression for percent error of f_{exp} versus $f_{m,exp}$.

This percent error is tabulated in Excel for all molecules' spectra of the five-point grid and the three-point (see in Appendix B.2).

2.4   Computational chemistry methods

A molecular model of the molecule of interest, such as 000, is produced in IQmol (Figure 2.4.21). Then, this model of 000 is digitized. In this case, digitization refers to taking a chemical structure known to chemists and translating it into x,y,z coordinates that are known to computers. These parameters are obtained in IQmol under the "Q-Chem Setup" option in the menu bar (Figure 2.4.22).
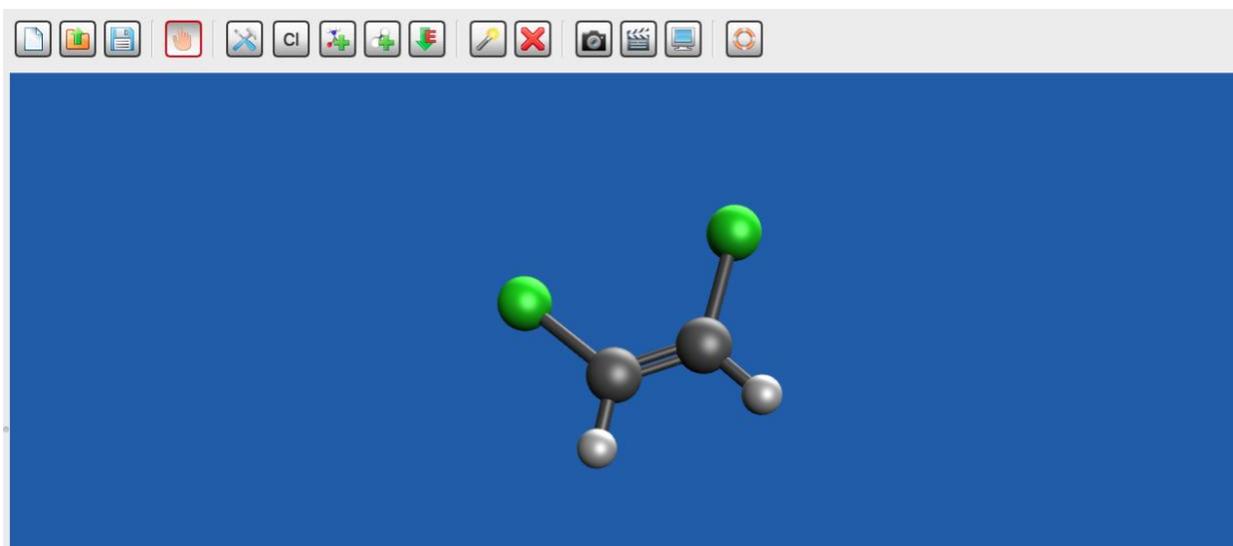
Figure 2.4.21: Molecular model of 000.

In Figure 2.4.21, Green spheres are chlorine atoms, dark gray spheres are carbon atoms, and light gray spheres are hydrogen atoms.

Figure 2.4.22: x,y,z-coordinates of 000.

2.4.1 The input script

Most of the calculations were performed on the Photon cluster housed at GSU. After connecting to GSU's virtual private network (VPN), Photon is accessed using ssh through the terminal application. A directory is created to organize all computations for this project using the command **mkdir [Project Name]**. To change the current directory to [Project Name], the user types the command **cd [Project Name]**. Then, a directory is made and labeled with the molecule ID using the command **mkdir [molecule ID]**. This directory is entered using the command **cd [molecule ID]**.

While in the directory for the molecule of interest, two scripts are required to run calculations on these molecules: an input file and a submission script. The input file contains the x,y,z molecular coordinates and instructions to Gaussian for which jobs to run. A template of an input file is shown below.

```
 1 %chk=opt[molecule ID].chk
 2 %nprocshared=4
 3 #p B3LYP/6-31+G* opt freq scrf=(pcm, solvent=[experimental solvent])
 4
 5 Optimization of [molecule ID]
 6
 7 0 1
 8 coordinates here...
 9
10 --link1--
11
12 %chk=opt[molecule ID].chk
13 %nprocshared=8
14 #P B3LYP/6-31+G* TD(NStates=[15 or 30], Root=1, Singlets) scrf=(pcm, solvent
   =[experimental solvent]) geom=checkpoint guess=read
15
16 TD-DFT calculation
17
18 0 1
19
20 █
```

Figure 2.4.1.23:Contents of a Gaussian input file, template.in.

Line 1 of template.in indicates a check point file. A check point file saves intermediate data on the job for the molecule, so if the job crashes while running, not all information is lost. The ID of the molecule of interest is written in place of [molecule ID] at lines 1, 5, and 12. Line 2 indicates the number of processors used to run these calculations. Since most molecules in this benchmark set are small, 4 processors are used instead of 8. Line 3 provides keywords on the calculations to be run. For instance, on line 3, the quantum computational calculation, 6-31+G*, includes diffuse functions which may be important, particularly for anionic systems. The command **scrf=(pcm,**

solvent=[experimental solvent]) instructs Gaussian to run the energy calculation in presence of a solvent using pcm. The solvent used for the pcm should reflect the solvent used to perform UV-visible spectroscopy on the molecule of interest. The solvent used for UV-visible spectroscopy varies for the molecules used in this project. So, the user must ensure the solvent used for the UV-visible spectroscopy of a molecule matches the solvent written in the input script for such molecule. A list of solvents available for Gaussian energy calculations is found in the reference, *https://gaussian.com/scrf/.*[30] Line 4 and line 6 must be left blank. Line 7 indicates the charge, the first digit, and multiplicity, the second digit, of the molecule. All moleules in this benchmark are either neutral (the charge is 0), monoanionic (the charge is -1), or monocationic (the charge is 1). All small molecules in this project have a multiplicity of 1 because all molecules considered have singlet ground states. Line 8 indicates the placement of x,y,z-coordinates of molecules with "coordinates here…." The x,y,z-coordinates of the molecule are copied from the "Q-Chem Setup" window in IQmol and pasted in place of "coordinates here…" (Figure 2.4.22).

So far, the input script contains instructions for Gaussian to run an optimization calculation and a frequency calculation. These calculations are indicated by **opt** and **freq** in the input script (Figure 2.4.1.23). The optimization calculation modifies the geometry of the molecule to minimize its quantum mechanical energy. The frequency calculation checks that the geometry has been optimized to a true minimum by performing a second derivative calculation of the energy with respect to coordinates for the optimization calculations. If geometry has been optimized correctly to an equilibrium structure, the frequency calculations are all positive. Positive values to a second derivative calculation indicate the molecule is optimized to a minimum rather than a or saddle point. Line 9 must be left blank. On line 10, **—link1—** combines the optimization and frequency calculations to the proceeding excited state energy calculations. Just as a check point file is

generated for the optimization and frequency calculations, a check point file is also generated for the excited state energy calculations on line 12. The number of processors used for the excited state energy calculations are instructed in line 13. On line 14, the number of excited states to solve for is indicated by the command **NStates=[15 or 30]**. Fifteen is a satisfactory number of states for calculation of small molecules. For relatively more complex molecules, we request a larger number of states (e.g., 30) since there may be many low-lying excited-states and more states are needed to simulate the full range of wavelengths in the experimental spectrum. The modifications of the input script described above for each of molecules in this project.

An example of the input script for 000, 000.in, commands Gaussian on the optimization, frequency, and excited state energy calculations (Figure 2.4.1.24).

```
 1 %chk=opt000.chk
 2 %nprocshared=4
 3 #p B3LYP/6-31+G* opt freq scrf=(pcm, solvent=heptane)
 4
 5 Optimization of 000
 6
 7 0 1
 8   C    -0.0000000   -1.0678244   -0.6683184
 9   C     0.0000000   -1.0678244    0.6683184
10  Cl     0.0000000    0.4190179   -1.6100959
11  Cl    -0.0000000    0.4190179    1.6100959
12   H     0.0000000   -2.0136027   -1.2005203
13   H    -0.0000000   -2.0136027    1.2005203
14
15 --link1--
16
17 %chk=opt000.chk
18 %nprocshared=8
19 #P B3LYP/6-31+G* TD(NStates=15, Root=1, Singlets) scrf=(pcm, solvent=heptane
   ) geom=checkpoint guess=read
20
21 TD-DFT calculation
22
23 0 1
24
25 ▌
```

Figure 2.4.1.24: Input script for molecule 000.

2.4.2 The submission script

Before executing the quantum chemical calculations, a submission script is needed. The submission script instructs the supercomputer on how to locate Gaussian and to run the job. A template of the submission script is shown in Figure 2.4.2.25.

```
1 #SBATCH -J opt[molecule ID]
2 #SBATCH -N 1
3 #SBATCH -n  8
4 #SBATCH -p  qPHO
5 #SBATCH -t  48:00:00
6 #SBATCH -e %J.err
7 #SBATCH -o %J.out
8 #SBATCH --mem-per-cpu=3000
9
10 mkdir -p /runjobs/RS10237/$USER/$SLURM_JOBID
11 module load ComputationalChemistry/g16
12 export GAUSS_SCRDIR=/runjobs/RS10237/$USER/$SLURM_JOBID/
13 echo $SLURM_JOB_NODELIST > nodename
14 cd $GAUSS_SCRDIR
15 cp $SLURM_SUBMIT_DIR/$SLURM_JOB_NAME.in $GAUSS_SCRDIR/
16 numactl -l /apps/Computation_Chemistry/Gaussian/Legacy/g16/g16 $GAUSS_SCRDIR
   /$SLURM_JOB_NAME.in
17 rsync -a --ignore-existing $GAUSS_SCRDIR/ $SLURM_SUBMIT_DIR
18 rm -r $GAUSS_SCRDIR/
19 █
```

Figure 2.4.2.25: Submission script for Gaussian in Photon.

All commands are performed in the node where the job is run. The label "#SBATCH" means that the commands are meant for the queueing system Slurm. The command **-J template** instructs the name of this job where "[molecule ID]" is replaced with the molecule's ID of interest. The commands **-N 1** and **-n 8** instructs the job to be run on one node with 8 CPUs on the one node, respectively. **-p qPHO** tells the computer the type node to run the job on a Photon node. **-t 48:00:00** instructs the job to stop running at 48 hours. **-e %J.err** and **-o %J.out** are commands for the computer to generate an error file and an output file, respectively, and named with the given job ID. **--mem-per-cpu=3000** instructs the computer on how much memory (in megabytes) to

allocate for this job per cpu. The command on line 10 **mkdir -p /runjobs/RS10237/$USER/$SLURM_JOBID** generates a temporary scratch folder, in which all Gaussian calculations are executed. "$USER" is the bash user variable, corresponding to the user name. "$SLURM_JOBID" is also a bash variable and defines the job with a specific ID. By assigning each job with a specific ID, the user avoids overriding a job when running multiple jobs at a time. In bash, variables are indicated with "$". Line 11 instructs the computer to load all necessary Gaussian (g16) files onto the node. The command on line 12 **export GAUSS_SCRDIR=/runjobs/RS10237/$USER/$SLURM_JOBID/** assigns an environmental variable, GAUSS_SCRDIR, to the scratch file and this variable is read by g16. Line 13 defines a nodename in the scratch file which can be deleted manually by the user once the job is finished. The nodename file is a way to document where the job was executed, in case of error or an incomplete job. Line 14 instructs the computer to access the bash variable composing the scratch file. Line 15 copies the content of the folder, $SLURM_SUBMIT_DIR, containing the input file, $SLURM_JOB_NAME.in, to the working scratch file, $GAUSS_SCRDIR. The command on line 16-17 **numactl -l /apps/Computation_Chemistry/Gaussian/Legacy/g16/g16 $GAUSS_SCRDIR/$SLURM_JOB_NAME.in** calls on Gaussian 16 and informs it of the input file, **$SLURM_JOB_NAME.in**. "numactl" is a memory management utility. The command on line 18 synchronizes the necessary work performed in the scratch file, $GAUSS_SCRDIR, to the submission folder, $SLURM_SUBMIT_DIR. When this command is executed, only the final calculations performed in $GAUSS_SCRDIR are synced to $SLURM_SUBMIT_DIR, avoiding all lengthy scratchwork. Lastly, line 19 instructs the computer to remove $GAUSS_SCRDIR.

To submit the job, the command **sbatch [molecule ID].sub** is entered in the command line. To check the job was submitted successfully, the command **squeue -u [username]** is entered. The

user ensures status of the job is "R" for running. The modifications of submission script described above are done for all molecules in this project.

2.4.3   Extracting the data

To ensure the geometry of a single molecule has been optimized, first, the directory of the molecule is opened with command **cd [molecule ID]** and then, the log file of the molecule is viewed with command **vi [molecule ID].log**. Finally, "Stationary point found" is searched for with command **/Stationary point found**. "Stationary point found" is indicated in the log file once the geometry optimization calculation is completed successfully. To find "Stationary point found" for all lines beginning with "opt" in all log files, the user can use the command **grep "Stationary point found" opt*/*.log**. The grep command is used to scan files for a string of characters that all match a single pattern. To ensure all frequency calculations are positive for a single molecule search the log file of the molecule with command **/Frequencies**. To find "negative Signs" for all lines beginning with "frequ" in lines beginning with "opt" within all log files, use the command **grep "frequ" opt*/*.log | grep "negative Signs"**. Ideally, no lines will exhibit "negative Signs" and this indicates all molecules reached optimal geometries with the optimization calculations. If "negative Signs" is found in a molecule's log file, the molecular model of such molecule is redrawn in IQmol and the geometries of certain substituents (alkyls) or atoms (hydrogens) are altered. Then, the x,y,z-coordinates are obtained once more and pasted into the molecules input script to rerun the job.

Once the user confirms the optimization and frequency calculations have completed successfully, the excited state energies and f_{comp} are extracted. To extract the molecule IDs pertaining to each excited state resolved in the calculation (either 15 or 30 states), the command **grep "Excited State" opt*/*.log|awk '[print \$1]' > ID** is typed. The awk command is commonly

used for extracting large amounts of data from a file. The ">" generates a file named "ID" which contains the extracted data resolved with grep and awk. Then, the command **cut -c 4-6 ID > IDcut** is used to remove columns four through six in the file. Also in this command, ID is rename the IDcut. Finally, the command **cat IDcut** is used to display the molecule IDs; this data is copied and pasted into Excel. To abstract the excited state numbers (1-15 for small molecules and 1-30 for large molecules), the command **grep "Excited State" opt*/*.log|awk '[print \$4]' > State#** is used (where # corresponds to the actual state number). Then, the command **sed 's/://g' State#** is typed which scans all lines of the log files for colons ":" and deletes them. To extract the excited state energies for all molecules, the command **grep "Excited State" opt*/*.log|awk '[print \$6]'** is used and these energy values (units of eV) are copied and pasted into Excel. To exstract the f_{comp} values, the command **grep "Excited State" opt*/*.log|awk '[print \$10]' > OS** is used. Following this command, **sed 's/f=//g' > OScut** is typed. Then, the command **cat OScut** is used to copy and paste these values into Excel. Following this procedure, the user has generated in Excel, a table comprised of the molecule IDs, the state number corresponding to each excited-state energy resolved in the calculation, the excited-state energies, and the computed oscillator strengths. For example, the data of 000 is shown (Table 2.4.3.5).

For this study, the f_{comp} related to one excited energy state (one UV-visible spectral peak) is found as the sum of all computed fs approximately lying within the range of wavelengths for which the molecules' spectral curve exists. The fs composing f_{comp} are chosen approximately to account for discrepancies in the values of computed wavelengths versus experimentally reported wavelengths.

Table 2.4.3.5: Computationally derived excited state energies and f comp of 000.

Molecule ID	State number	Excited state energy (eV)	Oscillator strength
000	1	5.713	0.000
000	2	6.321	0.385
000	3	6.509	0.003
000	4	6.843	0.000
000	5	6.917	0.000
000	6	7.103	0.000
000	7	7.372	0.000
000	8	7.390	0.002
000	9	7.422	0.025
000	10	7.623	0.005
000	11	7.887	0.083
000	12	8.114	0.004
000	13	8.183	0.000
000	14	8.217	0.075
000	15	8.407	0.008

In Table 2.4.3.5, the sum of the computed oscillator strengths highlighted in yellow is the f_{comp} of 000. See the Appendix B.4 for the state number, excited state energies, and computed fs for all small organic molecules in this project by corresponding IDs.

2.5   Visualization methods

The f_{comp} and the respective molecules' UV-visible spectra are visualized in Gnuplot.[32] The latest version of Gnuplot was downloaded to the Applications folder on the user's computer. In Terminal, Gnuplot is opened with command **cd /Applications/gnuplot-5.4.1**. The command **mkdir bin** is used. The command **./configure --with-cairo --prefix=/Applications/gnuplot-5.4.1/bin** is performed. This command ensures all dependencies are available in the computer's

system settings before the installation. Cairo is a visualization package for Gnuplot to generate the output file. Then the command **make** is used and following, the command **make install** is used.

2.5.1   The gnu file

Before generating a plot in Gnuplot, a data file containing the UV-visible spectral data, the fs composing f_{comp}, and f_{exp} of the molecule is required. Firstly, this data is copied from the working Excel sheet and orderly pasted in a new Excel sheet (Figure 2.5.1.26).

A	B	C	D	E	F	G
$\tilde{v}$ from Webplot Digitizer	ε from Weblplot Digitizer	ε of First Fitted Gaussian Curve	...	ε of nth Fitted Gaussian Curve	Computed $\tilde{v}$ of Vertical Excitation Energies	Computed f

Figure 2.5.1.26: Template Excel sheet of spectral data and oscillator strengths.

This Excel file is saved as a csv file, such as 000.csv, on the user's computer. The location of 000.csv should be easily identified and accessible to the reader. In terminal, the user should access the folder containing 000.csv file using a command such as **cd ~/Desktop/gnuplot/000**. This command indicates the 000.csv is located in a folder named 000 which is in a folder named gnuplot on the user's Desktop. Then, 000.csv is converted to a data file using the command **mv 000.csv 000.data**. The data file is read by Gnuplot. The 000.data file is viewed with the command **vi 000.data** and all commas in this file are replaced with tabs with the command **:%s/,/^I/g**. It is important to note that **^I** indicates the tab key.

The template Gnuplot input script, plot.gnu, from the downloaded and uncompressed .zip file is opened and viewed with the commands, **cd ~/Desktop/gnuplot** and **vi plot.gnu**, respectively (Figure 2.5.1.27). This input script is a set of instructions for the Gnuplot program on how to generate the plot for the molecule of interest.

```
1 set term svg size 1920,1080
2 set output 'molecule ID.svg'
3 set title '[molecule ID (molecule name)]'
4 unset key
5 set xrange[35000:55000]
6 set y2tics 0,0.1
7 set ytics nomirror
8 set yrange[0:20000]
9 set y2range[0:1]
10 set ylabel "Molar Extinction Coefficient, cm^{-1}mol^{-1}"
11 set y2label "Oscillator Strength"
12 set xlabel "Wavenumber (cm^{-1})"
13 set xtics font ", 24"
14 set ytics font ", 24"
15 set y2tics font ", 24"
16 set xlabel font ", 32"
17 set ylabel font ", 32"
18 set y2label font ", 32"
19 set lmargin 24
20 set rmargin 24
21 set ylabel offset -8,0,0
22 set y2label offset 8,0,0
23 set xlabel offset 0,-3,0
24 set title font ", 36"
25 plot "[path to file].data" u [column A]:[column B] w l lw 6 lc 6 axis x1y1, "" u [column A]
   :[column C] w l lw 4 lc 7 dt 2 axis x1y1, "" u [column A]:[column D] w l lw 4 lc 5 dt 2 axi
   s x1y1, "" u [column A]:[column E] w l lw 4 lc 4 dt 2 axis x1y1, "" u [column A]:[column F]
    w l lw 4 lc 3 dt 2 axis x1y1, "" u [column A]:[column G] w l lw 4 lc 2 dt 2 axis x1y1, ""
   u [column A]:($[column C]+$[column D]+$[column E]+$[column F]+$[column G]) w l lw 4 lc 0 ax
   is x1y1, "" u [column H]:[column I] w impulses lw 8 lc 8 axis x1y2
```

Figure 2.5.1.27: Template of the gnu script use to generate Gnuplot plots

Lines 1-24 instruct Gnuplot on how to define global parameters of the generated plot. These global parameters are independent of the data composed in the molecule's data file. Line 1 of this script indicates the type of file and pixel size of the plot to be generated. For all molecules' Gnuplot input scripts, the type of output file is "svg" and the size is "1920, 1080". Line 2 names the output file with the molecule's ID . Line 3 labels the title of the plot. For all molecule's the generated plot will be titled with the molecule ID and respective molecule name in parenthesis. Line 4 commands Gnuplot to not display a legend on the plot. Lines 5-24 are commands for formatting the x- and y-axes. The "x", "y", "y1", and "y2" indicate the x-axis, y-axes (both 1 and 2), y1-axis, and y2-axis, respectively. Two y-axes are used since one is used to plot the extinction coefficient for the experimental data, ε, and the other is used to plot f_{comp}. The commands including "range" instruct

Gnuplot on formatting the upper and lower limits of the axis for tic placement. The commands including "tics" instruct Gnuplot on the specific points on the axis. For instance, on Line 6, the command **set y2tics 0, 0.1** instructs Gnuplot to begin labeling tics at 0 with a spacing of 0.1. On Line 7, the command **set ytics nomirror** instructs Gnuplot to not add unlabeled tics on the other y-axis.[37] The commands with "label" instruct Gnuplot on the titles of the x- and y-axis. The commands with "font" instruct Gnuplot on the font style and font size. The default font style is "the ascii portion of Hershey simplex Roman provided in the file canvastext.js" and the default font size is 600 by 400 pixels.[38] This command follows the form **set [parameter] font "[style], [size]"**. When nothing is indicated for style, the default is set. The commands with "lmargin" and "rmargin" instructs Gnuplot on the the size of the margin on the left and the right, respectively.[38] The commands with "offset" follow the format **set [parameter] offset [left], [right], [top], [bottom]** where 0 is the default.[38] Offsets instruct Gnuplot on providing a boundary surrounding the data and labels on the generated plot.[38]

Lines 25 instructs Gnuplot on what data to plot, and in which form. These are dependent on the data composed in molecule's respective data. Gnuplot offers a variety of variables, expressions, and functions to be expressed consecutively or as a "string" with the plotting command.[38] For this input script, the command **plot "[path to file].data"** instructs Gnuplot to locate the data file to plot. Strung to this command is **u [column A]:[column B]** and this tells Gnuplot to plot using the data in column A and column B (Figure 2.5.1.26 and 2.5.1.27). Strung to this command is a command on how Gnuplot should display this data on the plot **w l lw 6 lc 6 axis x1y1**. "w l" is the abbreviated command for with lines, "lw 6" indicates a line width of six, and "lc 6" indicates a line color of six.[38] Each number following "lw" and "lc" corresponds to a specific width and color. For instance, six corresponds to blue for the line color command. "axis

x1y1" instructs Gnuplot to plot this data along the x1- and y1-axes. Altogether, the string of commands **plot "[path to file].data" u [column A]:[column B] w l lw 6 lc 6 axis x1y1** generates the UV-visible spectral curve of the molecule (Figure 2.5.1.26 and 2.5.1.27). A comma follows this string of command to introduce a new string of commands **" " u [column A]:[column C] w l lw 4 lc 7 dt 2 axis x1y1** (Figure 2.5.1.26 and 2.5.1.27). While this string of commands follows a similar format as the first string of commands, there are some notable differences. The set of parenthesis is an abbreviated notation commanding Gnuplot to operate on the same data file, defined earlier. "dt 2" is the abbreviated command for dashtype 2, which requests dashed instead of solid lines. This format of string of commands is repeated for plotting each fitted Gaussian curve until the nth one. Then the command **" " u [column A]:($[column C]+ $[column D]+ $[column E]+ $[column F]+ $[column G]) w 1 lw 4 lc 0 axis x1y1** instruct Gnuplot on plotting the sum of the fitted Gaussian curves (Figure 2.5.1.26 and 2.5.1.27). Lastly, the command **" " u [column H]:[column I] w impulses lw 8 lc 8 axis x1y2** tells Gnuplot to plot the fs of f_{comp} using impulse lines along using the y2- axis scale install of y1 (Figure 2.5.1.26 and 2.5.1.27).

For 000, 000.data and 000.gnu are generated (Figure 2.5.1.28 and 2.5.1.29).

54626.92976	5773.692098	1804.055086	3677.629409	131.436175	25.2105911	92.11745674	51587.09677	0.313
54379.47364	6342.32223	2009.033492	3954.900755	181.6356102	52.96181693	144.7397134	51619.35484	0.004
54132.09658	6901.818243	2179.708296	4188.3828	247.1706672	104.0059036	219.6646247		
53884.74587	7487.190966	2304.083264	4368.273811	331.2344228	190.9631075	322.0381898		
53637.44787	8071.507541	2372.954476	4486.702619	437.125596	327.8107498	456.0568285		
53390.17622	8674.222955	2381.106782	4538.403314	568.1006337	526.1603719	623.9045836		
53142.98363	9234.79026	2327.949063	4521.061817	727.0638575	789.6036478	824.4903858		
52854.634	9885.603366	2194.148943	4414.849857	950.9911583	1164.766388	1092.632735		
52525.29204	10434.22325	1963.840222	4188.683408	1259.834149	1623.951393	1422.989327		
52196.02914	10910.32277	1678.492842	3867.850246	1624.234138	2009.907817	1743.049212		
51867.04295	11039.18304	1370.186015	3476.419156	2037.675834	2208.392898	2008.114269		
51538.27417	10884.73179	1068.354644	3041.467057	2487.808717	2154.751894	2176.250321		
51209.74256	10434.22325	795.7472659	2590.302198	2956.061288	1867.439128	2218.828871		
50922.44047	9862.415937	592.30006	2201.718573	3362.061639	1495.259092	2146.774963		
50676.35472	9205.920358	447.2762444	1884.397738	3692.784569	1150.120695	2010.991884		
50430.32167	8539.480808	329.1909792	1588.632455	3995.035429	827.8102707	1820.509633		
50143.01299	8077.828204	222.7544628	1276.774643	4296.360192	518.3576562	1552.154394		
49979.22154	7475.478536	175.5091167	1116.852029	4437.037105	381.1674251	1388.005905		
49883.75724	7076.88438	151.9403496	1029.859238	4507.178989	314.3391049	1291.407146		
49610.15879	6689.063062	98.36856681	806.0495286	4655.276645	171.1570165	1020.693563		
49364.04668	6263.383618	64.74248638	636.3428589	4716.631607	92.34668896	796.7221519		
49076.87635	5828.181771	38.46487538	473.7801532	4698.43732	41.32706271	571.4660871		
48789.75873	5389.364023	22.06705634	345.5674286	4584.837761	16.89677054	391.2740036		

Figure 2.5.1.28: The contents of 000.data which is used by Gnuplot to generate the plot.

```
 1 set term svg size 1920,1080
 2 set output '000.svg'
 3 set title '000 (cis-1,2-Dichloroethylene)'
 4 unset key
 5 set xrange[42000:55000]
 6 set y2tics 0,0.1
 7 set ytics nomirror
 8 set yrange[0:14000]
 9 set y2range[0:1]
10 set ylabel "Molar Extinction Coefficient, cm^{-1}mol^{-1}"
11 set y2label "Oscillator Strength"
12 set xlabel "Wavenumber (cm^{-1})"
13 set xtics font ", 24"
14 set ytics font ", 24"
15 set y2tics font ", 24"
16 set xlabel font ", 32"
17 set ylabel font ", 32"
18 set y2label font ", 32"
19 set lmargin 24
20 set rmargin 24
21 set ylabel offset -8,0,0
22 set y2label offset 8,0,0
23 set xlabel offset 0,-3,0
24 set title font ", 36"
25 plot "/Users/astridtarleton/Desktop/gnuplot/000/000.data" u 1:2 w l lw 6 lc 6 axis x1y1, ""
   u 1:3 w l lw 4 lc 7 dt 2 axis x1y1, "" u 1:4 w l lw 4 lc 5 dt 2 axis x1y1, "" u 1:5 w l lw 4
    lc 4 dt 2 axis x1y1, "" u 1:6 w l lw 4 lc 3 dt 2 axis x1y1, "" u 1:7 w l lw 4 lc 2 dt 2 axi
   s x1y1, "" u 1:($3+$4+$5+$6+$7) w l lw 4 lc 0 axis x1y1, "" u 8:9 w impulses lw 8 lc 8 axis
   x1y2
```

Figure 2.5.1.29: The modified gnu script for generating a Gnuplot plot for 000.

The labels, column A, column B, column C etc. in plot.gnu, are replaced with 1, 2, 3 etc., respectively and this is done for all molecule's gnu files (Figure 2.5.1.27 and 2.5.1.29). With 000.data and 000.gnu, the plot of 000 is generated by Gnuplot (Figure 2.5.1.30).
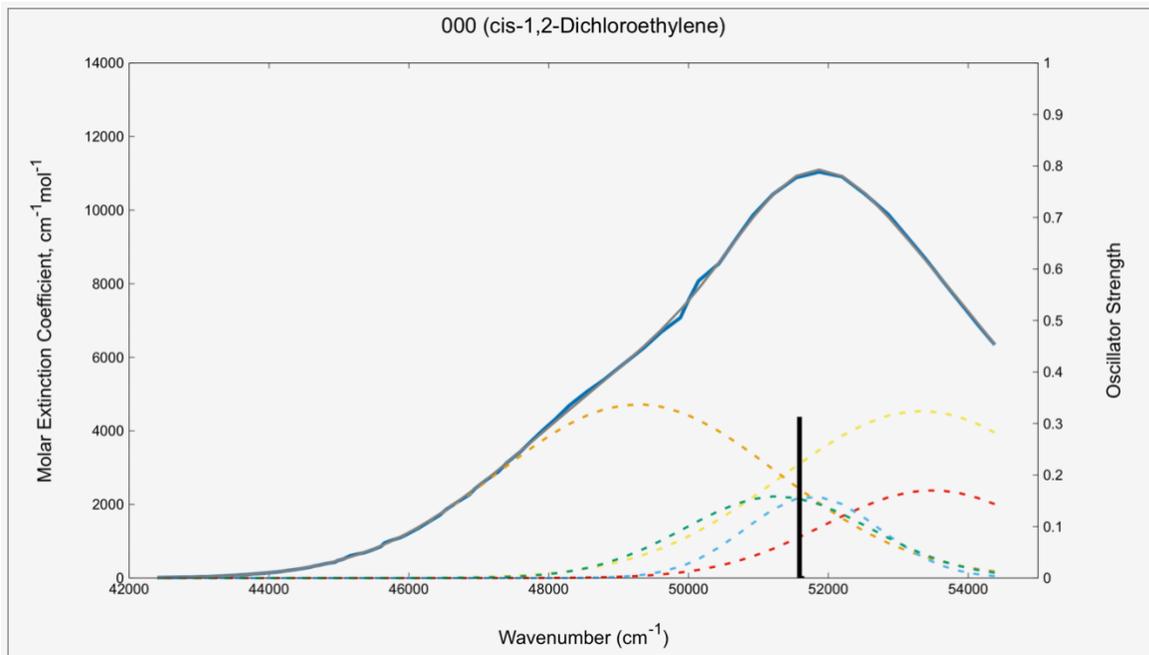
Figure 2.5.1.30: Plot of 000.

3   RESULTS

We have used several diagnostic tools to quantify the accuracy of the numerical and analytical fitting procedures. As mentioned in the Experimental Details section, the quality of the digitization is determined by the number of points digitized for each molecules' spectra (the count) and the density of points along the range of wavenumbers ($\tilde{v}$) for which the molecules' spectra exist (the range / count). To ensure reproducibility of the digitization and fitting, these processes were repeated three times (using a three-point grid, a five-point grid, and a third set of data referred to as a "variable-point grid" set). The range/count for each molecule in each set is recorded in Appendix B.2. Note that the range/count of all molecules' spectra in the variable-count grid is kept under 100 cm^{-1}. On average the range/count is 195 cm^{-1}, 133 cm^{-1} , and 68 cm^{-1} for the five-, three-, and variable-, point grids, respectively.

The NRMSE, normalized root-mean-square error, for the fitting the molecules' spectra with Gaussian curves in the five-, three-, and variable-point grids are on average 1.41%, 1.30%, and 0.45%, respectively. In large part, this is because additional Gaussian curves were used in the fitting in the case of the variable-point grid. In almost all excited states of molecules, the NRMSE was kept under 0.50%, meaning that the relative error introduced by the fitting relative to the intensity of the peak is kept small. A recording of the NRMSE for each of the molecules' spectra can be found in the Appendix B.2. To interpret the correlation between the experimentally derived oscillator strengths with Gaussian curves (f_{exp}) and the oscillator strengths found with numerical integration following the Midpoint rule ($f_{m,exp}$), f_{exp}s are plotted against $f_{m,exp}$s (Figure 3.31).
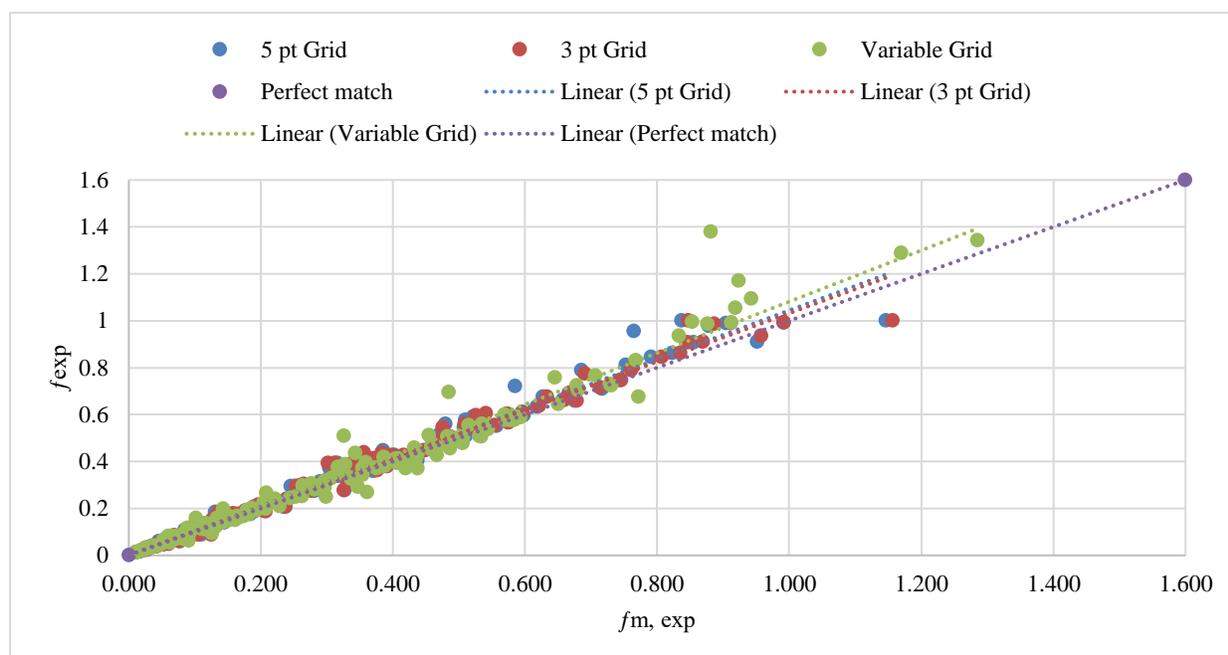
Figure 3.31: ƒexp vs ƒm,exp.

See <u>Appendix B.1</u> for exact values of f shown on the plot (Figure 3.31).

The slope of the linear trendline of the three-, five-, and variable-point grid are 1.0317, 1.0443, and 1.0691, respectively. The R^2 value of the linear trendline of the three-, five-, and variable-point grids are 0.9951, 0.9935, and 0.9847, respectively. The percent deviations of the f_{exp} versus $f_{m,exp}$ are found in the <u>Appendix B.2</u>. The average percent deviations of f_{exp} versus $f_{m,exp}$ in the five-, three-, and variable-point grids are 6.4%, 6.2%, and 7.7%, respectively. This indicates that, on average, the difference between numerical and analytical fitting with Gaussian curves is around 7%. There are several factors that contribute to this difference, but the mains ones are:

- Truncated peaks where the numerical integration is missing part of the peak, however, this missing part is accounted for in the Gaussian fitting. An example is molecule 000 in Figure 2.3.2.14; the percent deviations of f_{exp} versus $f_{m,exp}$ is around 12%.

- Overlapping peaks where a nearby large peak cuts into a smaller peak, and therefore, the Gaussian fitting gives a smaller f_{exp} relative to $f_{m,exp}$. An example is peak 1 of 013, shown in Figure 2.3.2.17.

Therefore, the f_{exp} versus $f_{m,exp}$ percent difference should not be interpreted as an error. Rather, it may be viewed as an indicator in some molecules that there is an uncertainty in fitting. Specifically, when the f_{exp} versus $f_{m,exp}$ percent difference is large, this indicates that there is a large truncation of the peak or an uncertainty in how to deal overlaps with nearby peaks.

To show the correlation between f_{exp} and the quantum chemically computed oscillator strengths (f_{comp}), f_{exp}s are plotted against f_{comp}s (Figure 3.32).

As stated in the Experimental Details section, the sum of all computed fs corresponding to wavenumbers for which the molecules' spectral curve exists is reported as the f_{comp}. See the Appendix B.1 for exact values of f shown on the plot (Figure 3.32).
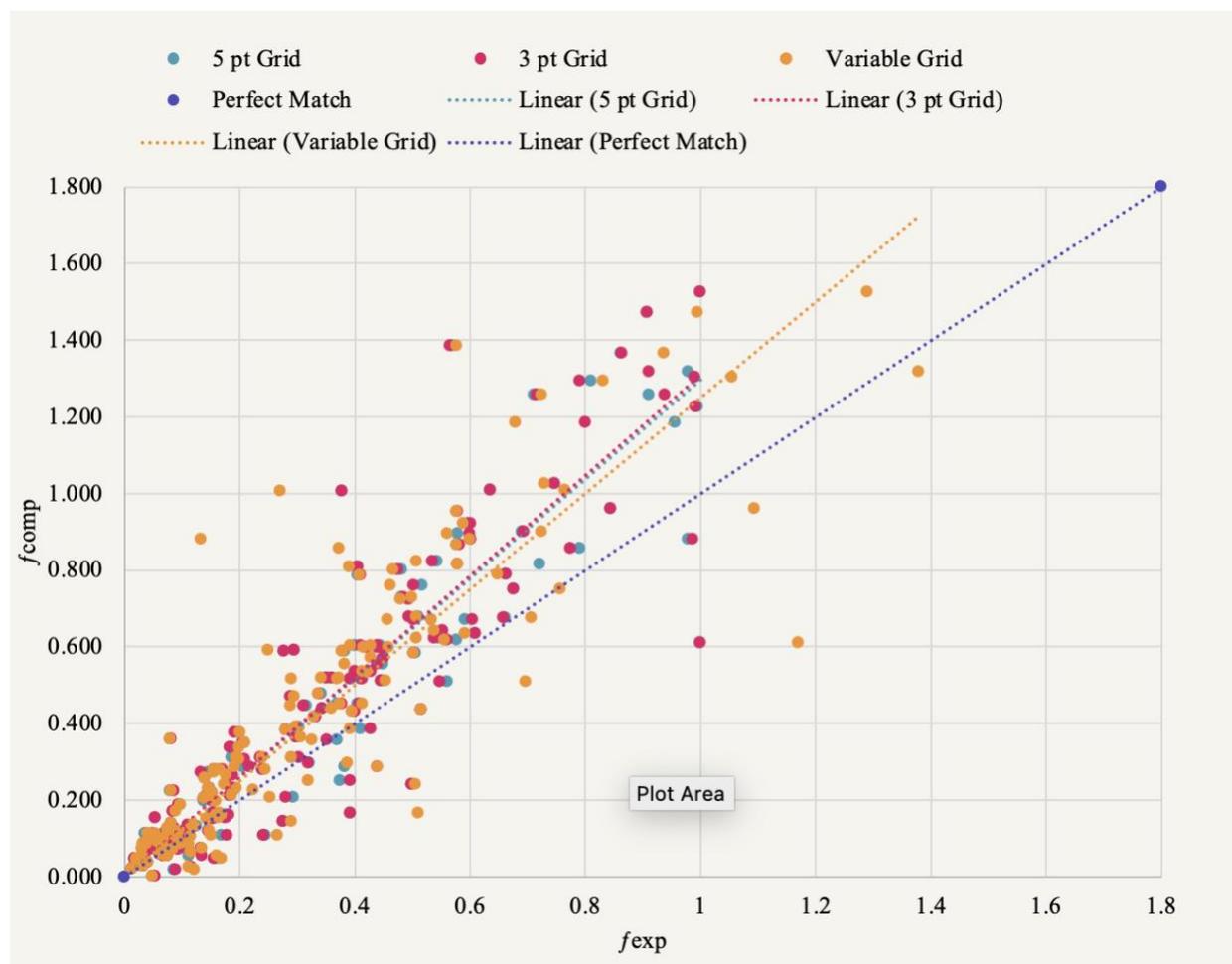
Figure 3.32: fexp vs. fcomp.

The slope of the linear trendline of the three-, five-, and variable-grids are 1.2967, 1.3080, and 1.2485, respectively (Figure 3.32). The R^2 value of the linear trendline of the three-, five-, and variable-grids are 0.9338, 0.9325,and 0.8959, respectively.

Finally, it is important to note that there is no unique way to fit these spectra, due to uncertainties with peak truncation and overlaps. This is why the results of the five-point grid, three-point grid, and variable-point grid were not identical. We may use these three trials to also get a sense of the errors and uncertainties associated with digitizing the spectra with Webplot Digitizer and fitting them. The percent deviations are found given the expression in Equation 3.8 and 3.9.

$$Percent\ error\ (\%) = \left| \frac{f_{exp,max\ in\ the\ three\ grid} - f_{exp,min\ in\ the\ three\ grid}}{f_{exp,variable-point\ grid}} \right| \bullet (100)$$

Equation 3.8: Mathematical expression for percent error of f_{exp}.

$$Percent\ error\ (\%) = \left| \frac{f_{m,exp,max\ in\ the\ three\ grid} - f_{m,exp,min\ in\ the\ three\ grid}}{f_{m,exp,variable-point\ grid}} \right| \bullet (100)$$

Equation 3.9: Mathematical expression for percent error of $f_{m,exp}$.

The percent deviations of f_{exp} between the three grids capture differences in how the fitting was done. A large error in f_{exp} is a clear indication of an uncertainty in how to fit the peaks (i.e., due to truncation or overlaps). On the other hand, difference in $f_{m,exp}$ are attributable directly to errors introduced by digitizing the spectra. The percent deviations of f_{exp} and separately, $f_{m,exp}$, are reported in Appendix B.3.

4   DISCUSSION

We have obtained experimental derived oscillator strengths, f_{exp} and $f_{m,exp}$, and computationally derived oscillator strengths, f_{comp}, for nearly one-hundred small to medium sized organic compounds. To classify the confidence in the digitization, truncation, and fitting of the molecules' UV-visible spectra, a ranking system is employed and is found in the <u>Appendix B</u> for each of the diagnostic tools. The range/count, consequential to the digitization and a measure of the fineness of the grid, is ranked with high confidence if the value is less than or equal to 100, average confidence if the value is between 100 and 200, inclusively, and poor confidence if the value is equal to or greater than 200. Similarly, the NRMSE is ranked with high confidence if the value falls below 0.5%, average confidence if the value falls between 0.5% and 1.0%, inclusively, and poor confidence if the value is greater than 1.0%. The percent deviation between f_{exp} versus $f_{m,exp}$ is also ranked with high confidence if the value falls below 5%, average confidence if the value falls between 5% and 15%, inclusively, and poor confidence if the value is greater than 15%. Lastly, the percent deviations of f_{exp} in the three-, five-, and variable-point grids and separately, the percent deviations of $f_{m,exp}$ in the three-, five-, and variable-point grids are ranked. The percent deviations of f_{exp} are classified with high confidence if the value lies below 5%, average confidence if the value is between 5% and 10%, respectively, and poor confidence if the value is above 10%. The percent deviations of $f_{m,exp}$ are classified with high confidence if the value lies below 3%, average confidence if the value is between 3% and 10%, respectively, and poor confidence if the value is above 10%.

Generally, high confidence classification of the range/count is reported more often in the three-, and variable-point grids than the five-point grid due to the greater fineness of the former two. This is expected for the three-point grids, given more points (greater count) are generally

digitized by Webplot Digitizer compared to the five-point grid. This is also to be expected for the variable-point grids because they are generated from high resolution images. In other words, these images contain a greater total pixel count per image compared to the images used in the three-, and five- point grids.

The digitized spectra are smoother when plotted with points of a finer grid. Thus, the Gaussian curves are more easily fitted in a three- and variable- point grid set than in a five-point grid set given Gaussian curves are smooth in character. The NRMSE is an indication of the effect of the fineness of the grid on Gaussian curve fitting to molecules' spectra. The NRMSE is classified more often with high confidence in the finer grids sets than the five-point grid set. It is also worth pointing that in the case of the variable-point grid, a larger number of Gaussian curves were used to fit the data as well, contributing to the better NRMSE.

Another consequence of high grid fineness is low percent deviations of f_{exp} verses $f_{m,exp}$. In accordance with the Midpoint rule, a finer grid (high count) permits a greater number of rectangles to fit the spectral curve. Even still, the f_{exp} are on average recorded slightly greater in value than the $f_{m,exp}$. The characteristics of a Gaussian curve (smooth, symmetric, and bell-shaped) are similar to such characteristics of a UV-visible spectrum showing complete broad-band peaks. So, the Gaussian curves can more closely fit to the area under the curve compared to rectangles (from the Midpoint rule), especially at the tail ends of a spectrum. The average percent deviation of f_{exp} verses $f_{m,exp}$ for the variable-point grid set is slightly higher than such averages for the three- and five-point grid sets. This is due to the fact that Gaussian curves were fit to molecular spectra with constraints more often in the three- and five-point grids than in the variable-point grids.

Indeed, the ranking system is imperative to categorizing the reliability and accuracy of the fitting. It is also an important tool for identifying outlier data, especially pertaining the method of truncation and fitting. For example, an outlying f_{exp} or $f_{m,exp}$ reported for a particular spectrum may be attributed to a difference in the method of truncation. The discrepancy in truncation mostly occurs when two peaks overlap significantly in a molecule's spectrum and the point of inflection, for which truncation is performed, is unclear. The difference in method of truncation may also occur when a molecule's spectrum shows a peak with a trailing tail. In this case, the point of inflection on the tail of the peak is unclear. An example of discrepancies in the user's method of fitting is shown in a high percent deviation of f_{exp} of the three-point grid compared to f_{exp} in the variable-point grid of the same spectrum. This can be attributed to a difference in the use or disuse of the Excel Solver constraints. This difference is especially prevalent in spectra of unresolved or incomplete peaks. Such spectra are non-Gaussian-like and to compensate for the challenge of fitting such spectra, constraints are used to conduct a realistic fit.

On average, the f_{exp} are smaller than the f_{comp}. This is an indication that computations with the B3LYP method and 6-31+G* method overestimate the experimental oscillator strengths. This can also be seen in some molecules where computations predict an f_{comp} larger than 1. Experimentally, while an f describing a single electronic transition is reported as a value between 0 and 1, when electronic transitions occur at similar energy values, they may be reported as greater than 1. This is observed in the spectra with overlapping peaks.

In figures 3.31 and 3.32, R^2 value is an indication of how well the plotted points are fitted to the linear trendline Ideally, $R^2=1$. The average R^2 value between the f_{exp} and the f_{comp} is 0.9324, indicating on average the f_{comp} values are greater than the f_{exp} values (Figure 3.32). By allowing f_{exp} of the molecules' spectra to exceed 1 and by fine-tunning or eliminating oscillator strength

outliers, it is expected that the R^2 value to improve. f outliers are typically a consequence of fitting unresolved or incomplete peaks in the molecules' spectra. With an improved R^2-value, a f_{exp} may be determined with the f_{comp} alone, using a scaling factor determined by the slope of the linear trendline.

We acknowledge the B3LYP/6-31+G* method is just an approximative approach for solving the Schrödinger equation and is not exact. While this method is widely used in quantum chemical computations, it is still an approximation that fails in several cases. Future endeavors for this project include computing and comparing different quantum chemical methods and basis sets using the same set of experimental data generated in this work.

Our research goal is developing a data benchmark set to understand correlations between computed and oscillator strengths derived from integration of spectra for nearly 100 small to medium sized organic molecules. This benchmark set demonstrates the extent of the correlation between the computational and analytical methods. By validating the computed oscillator strengths with experimentally derived oscillator strengths, quantum chemical computations may be used solely to compare and even interpret electronic structures of molecules.

5   SUMMARY AND FUTURE DIRECTION

This project has found computed and experimentally derived oscillator strengths (f) for small organic compounds. Computed oscillator strengths (f_{comp}) are calculated with the most widely used quantum chemical method, B3LYP, a hybrid of Hartree-Fock and DFT methods, and the 6-31+G* basis set. f_{exp} were derived from digitizing UV-visible spectra, fitting them with Gaussian curves, and determining the total area under these spectral curves by analytical integration. We also employed numerical integration with the Midpoint rule to derive $f_{m,exp}$ and compared $f_{m,exp}$ against f_{exp}. The f_{comp} and f_{exp} correlate very well, with few outliers. With such high correlation, we find that f_{exp} can be compared or even predicted by using B3LYP/6-31+G* calculations and a simple scaling factor. Thus, quantum chemical calculations may be solely used to define electronic structures of molecules.

Future endeavors for this study involve deriving a statistical measure (besides NRMSE) which more so reflects the error due to fitting Gaussian curves to the molecules' spectra. Another major step would be to then use those analytical tools to better estimate the errors associated with fitting. Ultimately, the goal would be to score the experimental spectra using one single metric that would help indicate when we have very high confidence (<2% error), high confidence (<5% error), medium confidence (<10% error), low confidence (<20% error), and no confidence (>20% error). Such a single metric, which is a combination of metrics reported so far or a new statistical tool, would help better quantify the errors associated with the experimental fitting and therefore also quantify the errors in computational predictions of the oscillator strength relative to the experiments.

Another important direction would be to understand the effect of solvation on the oscillator strengths of these molecules. In the computations reported in this work, the solvent is included

using a polarizable continuum model (PCM) with a dielectric constant matching the experimental

solvent. However, it is not clear whether this adequately captures the effect of solvation, nor is it

clear how big a change the solvent introduces in the oscillator strength of the molecule. On the

experimental side, there is also uncertainty in whether the refractive index term should be

accounted for in expression for f (Equation 1.4.4). This would decrease f_{exp} and $f_{m,exp}$, which

would only reduce the agreement between experiments and computations further. Therefore,

further testing and analysis maybe done, both on the experimental and computed strengths, to

understand the effect of the solvent dielectric and strength on the oscillator strength.

Finally, the main goal for this work is to prepare a set of experimental data that can be used

to test a wide range of computational methods, not just a single method. A natural future direction

of this work is the benchmarking of the experimental data against a range of DFT functionals,

basis sets, and potentially post-Hartree-Fock methods. Such testing would provide important

guidelines into which methods and basis sets are best suited for the calculation of quantitative

oscillator strengths, and how sensitive the calculations are to these methods and basis sets.

REFERENCES

1.   Atkins, P.; Paula, J. d.; Keeler, J., *Atkins' Physical Chemistry*. 11 ed.; Oxford University Press, 2018.

2.   Turro, N. J.; Ramamurthy, V.; Scaiano, J. C., *Principles of Molecular Photochemistry: An Introduction*. 1 ed.; University Science Books: 20 Edgehill Rd, Mill Valley, California 94941, 2015.

3.   Cederbaum, L. S., The Exact Molecular Wavefunction as a Product of an Electronic and a Nuclear Wavefunction. *The Journal of Chemical Physics* **2013,** *138*.

4.   Jensen, F., Atomic Orbital Basis Sets. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *3* (3), 273-295.

5.   Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry*. General Publishing Company: 30 Lesmill Road, Don Mills, Toronto, Ontario, 1989.

6.   Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*. 2 ed.; John Wiley & Sons, Ltd: 2004.

7.   Becke, A. D., A New Mixing of Hatree-Fock and Local Density-functional Theories. *Journal of Chemical Physics* **1993,** *98* (2), 1372-1377.

8.   Baseden, K. A.; Tye, J. W., Introduction to Density Functional Theory: Calculations by Hand on the Helium Atom. *Journal of Chemical Education* **2014,** *91* (12), 2116-2123.

9.   Manby, F., Density Functional Theory. In *Theoretically Speaking*, 2018.

10.   Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** *37* (2), 785-789.

11.   *Fundamentals of Time-dependent Density Functional Theory*. Springer Science & Business Media: Chicago, 2012; Vol. 837.

12.   Marques, M. A. L.; Gross, E. K. U., TIME-DEPENDENT DENSITY FUNCTIONAL THEORY. *Annual Review of Physical Chemistry* **2004,** *55* (1), 427-455.

13.   McNaught, A. D.; Wilkinson, A., IUPAC: The "Gold Book". Blackwell Scientfic Publications: 1997.

14.   Condon, E. U., Nuclear Motions Associated with Electron Transitions in Diatomic Molecules. *Physical Review* **1928,** *32*.

15.   Wenzel, T. Introduction.

16.   Coon, J. B.; DeWames, R. E.; Loyd, C. M., The Franck-Condon Principle and the Structures of Excited Electronic States of Molecules. *Journal of Molecular Spectroscopy* **1962,** *8*, 285-299.

17.   Martynoff, M., Note de laboratoire: Spectres d'absorption de quelques p-quinones. *Bulletin de la Société Chimique de France 16*, 258-261.

18.   Valeur, B.; Berberan-Santos, M. N., Absorption of Uv-Visible Light. *Molecular Fluorescence, Weinheim* **2002,** *27*.

19.   Mata, R. A.; Suhm, M. A., Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction? *Angewandte Chemie International Edition English* **2017,** *56* (37), 11011-11018.

20.   Beckel, C. L.; Satler, J. P., Spectroscopic Constants of the Kolos-Wolniewicz Potential of H2. *Journal of Molecular Spectroscopy* **1966,** *20*, 153-158.

21.   Matsuzawa, N. N.; Ishitani, A.; Dixon, D. A.; Uda, T., Time-Dependent Density Functional Theory Calculations of Photoabsorption Spectra in the Vacuum Ultraviolet Region. *The Journal of Physical Chemistry A* **2001,** *105* (20), 4953-4962.

22.   Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B., Oscillator Strength: How Does TDDFT Compare to EOM-CCSD. *Journal of Chemical Theory and Computation* **2011,** *7*, 456-466.

23.   Chrayteh, A.; Blondel, A.; Loos, P.-F.; Jacquemin, D., Mountaineering Strategy to Excited States: Highly Accurate Oscillator Strengths and Dipole Moments of Small Molecules. *Journal of Chemical Theory and Computation* **2020**.

24.   Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C., Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *Journal of Chemical Theory and Computation* **2009,** *5* (9), 2420-2435.

25.   Jacquemin, D.; Duchemin, I.; Blondel, A.; Blase, X., Assessment of the Accuracy of Bethe-Salpeter (BSE/GW) Oscillator Strengths. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3969-3981.

26.   UV Atlas of Organic Compounds. Springer Science+Business Media, LLC: Buttersworths, London, 1967.

27.   V., T.; A.N., Y.; A.N., L.; A.A., U.; A.A., G.; N.A., M.; N.V., U.; M.V., E.; Aristova, E. V., "UV/Visible Spectra". In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, Linstrom, P. J.; Mallard, W. G., Eds. National Institute of Standards and Technology (NIST): Gaithersburg MD, 20899.

28.   Rohatgi, A. Webplot Digitizer-Web Based Plot Digitizer. (accessed May 5, 2020).

29.   Gilbert, A. *IQmol*, 2.11 (2018-03-29); 2018.

30.   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

31.   Mennucci, B.; Tomasi, J.; Cammi, R.; Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Gabriel, S.; Stephens, P. J., Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotations of Chiral Molecules. *The Journal of Physical Chemistry A* **2002,** *106* (25), 6102-6113.

32.   Williams, T.; Kelley, C. *Gnuplot-5.4.1: An Interactive Plotting Program*, 2020.

33.   Saskia How to Normalize the RMSE.

34.   JJ. MAE and RMSE—Which Metric is Better? *Human in a Machine World* [Online], 2016.

35.   Statistics—Normalization. 5 June 2014 ed.; Costal Inlets Research Program (CIRP): 2014.

36.   Numerical Integration—Midpoint, Trapezoid, Simpson's Rule.

37.   Shimoyama, H. Gnuplot Manual: Frequently Used Methods.

38.   Williams, T.; Kelley, C., *gnuplot 5.2: An Interactive Plotting Program*. 2019.

APPENDICES

Appendix A

Atomic orbital: A single electron function describing the shape and motion of that electron around an atom.

Basis set: A set of functions that can resemble those of an atomic orbital (quantum chemistry definition)

Born-Oppenheimer approximation[1,2]: An ansatz that the motion of the nuclei and the motion of electrons can be treated separately. The approximation assumes that since nuclei are much heavier than electrons, the speed of the nuclei is much slower than the speed of an electron, and therefore the nuclei can be regarded as stationary relative to the electron. The benefit of this approximation is that the Schrödinger equation can be solved for electrons independently from the nuclei.

Electron correlation: A measures how one electron is affected by all other electrons in the system.

Electron density: a measure of the probability of an electron to exist in a certain region of space.

Electron exchange energy[2,9]: electron-electron repulsive energy occurring in molecular orbitals (following the Pauli exclusion principle). In essence, this is a term arising from quantum mechanics and it cannot be observed; rather, it serves as a correction for the classical coulombic electron-electron repulsion required to obtain the correct answer for an energy state.

Electronic spin (S): The natural angular momentum of an electron[2]

Electronic wavefunction (ψ_o): A function that describes the shape and motion of electrons around a fixed field of nuclei in a molecule.[2]

Molecular orbital: A single electron function describing the shape and motion of that electron around an atom. Often expressed as a linear combination of functions composed in the basis set for particular molecule.[2]

Operator: An expression containing mathematical operations to be performed on a function.

Orbital Approximation: An ansatz that the total wavefunction for a multi-electron atom is the product of one-electron orbitals. $\Psi(r1, r2, \ldots)=\Psi(r1)\ \Psi(r2)\ldots$ [2]

Vibrational wavefunction (χ): Describes the shape and motion of nuclei in a molecule.[2]

Variational principle: For a wavefunction used to calculate an energy parameter of a system, the eigenvalue determined for the energy in never smaller than the true energy of such system

Wavefunction (ψ): holds all structural information of a molecule in a particular state.[2]

Appendix B

Appendix B.1

Molecule ID	Peak#	f_{comp}	f_{exp}			$f_{m,exp}$		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
000	1	0.389	0.303	0.298	0.297	0.265	0.260	0.264
001	1	0.363	0.299	0.299	0.306	0.282	0.283	0.277
002	1	0.416	0.332	0.330	0.328	0.304	0.307	0.307
003	1	0.536	0.400	0.400	0.419	0.378	0.381	0.387
004	1	0.899	0.691	0.692	0.723	0.666	0.669	0.679
005	1	1.010	0.634	0.635	0.765	0.618	0.621	0.706
006	1	0.263	0.187	0.187	0.175	0.184	0.187	0.183
007	1	0.724	0.492	0.493	0.479	0.490	0.487	0.505
008	1	0.438	0.343	0.343	0.360	0.324	0.325	0.325
009	1	0.432	0.394	0.399	0.395	0.359	0.361	0.360
010	1	0.134	0.108	0.109	0.117	0.085	0.088	0.089
011	1	0.437	0.515	0.514	0.513	0.477	0.480	0.455
012	1	0.030	0.033	0.033	0.033	0.033	0.032	0.033
	2	0.335	0.185	0.184	0.199	0.131	0.136	0.143

Molecule ID	Peak#	f_{comp}	f_{exp} (continued)			$f_{m,exp}$ (continued)		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
013	1	0.032	0.026	0.026	0.027	0.026	0.026	0.027
	2	0.312	0.237	0.236	0.240	0.221	0.221	0.221
014	1	0.308	0.191	0.208	0.198	0.204	0.203	0.208
	2	0.086	0.112	0.109	0.112	0.094	0.091	0.097
015	1	0.349	0.208	0.207	0.209	0.235	0.238	0.229
	2	0.165	0.392	0.392	0.509	0.316	0.312	0.326
016	1	0.360	0.078	0.080	0.079	0.076	0.076	0.077
017	2	0.203	0.137	0.137	0.136	0.116	0.117	0.117
018	2	0.085	0.078	0.081	0.097	0.084	0.084	0.085
019	2	0.114	0.071	0.071	0.081	0.060	0.061	0.060
020	1+2	0.922	0.597	0.601	0.587	0.598	0.596	0.593
021	1+2	0.674	0.659	0.658	0.705	0.676	0.679	0.680
022	1+2	0.800	0.481	0.473	0.467	0.467	0.460	0.459
023	2	0.171	0.084	0.084	0.088	0.084	0.083	0.083
024	2	0.296	0.320	0.319	0.387	0.330	0.333	0.328
025	2	0.587	0.277	0.278	0.378	0.327	0.326	0.323
026	2	0.952	0.577	0.578	0.576	0.578	0.577	0.570
	3	0.208	0.293	0.280	0.253	0.246	0.256	0.262
027	1	1.257	0.710	0.716	0.724	0.717	0.714	0.731
	2+3	0.243	0.499	0.499	0.505	0.481	0.477	0.483
028	1	1.387	0.570	0.566	0.576	0.575	0.575	0.584
029	1				0.014			0.012
	2	0.881	0.601	0.601	0.598	0.572	0.574	0.569
030	1	0.143	0.276	0.276	0.290	0.282	0.277	0.280
	2				0.248			0.253
031	2	0.784	0.405	0.411	0.407	0.437	0.428	0.428
032	1	0.478	0.341	0.336	0.336	0.311	0.311	0.313
033	1	1.026	0.746	0.746	0.729	0.744	0.746	0.728
	2	0.053	0.031	0.029	0.030	0.027	0.027	0.027
034	2+3	0.518	0.350	0.352	0.342	0.351	0.352	0.354
035	1	0.591	0.294	0.295	0.248	0.252	0.254	0.299
036	1	0.727	0.481	0.485	0.500	0.491	0.491	0.505
	2	0.356	0.369	0.352	0.326	0.333	0.326	0.337
037	1+2	0.639	0.552	0.554	0.537	0.557	0.554	0.544
038	1+2	0.822	0.543	0.536	0.506	0.534	0.531	0.534

Molecule ID	Peak#	ƒcomp	ƒexp (continued)			ƒm,exp (continued)		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
039	1	0.790	0.662	0.663	0.646	0.657	0.659	0.650
040	1	0.809	0.405	0.405	0.391	0.423	0.426	0.422
	2	0.617	0.577	0.561	0.555	0.510	0.509	0.515
041	2	0.018	0.087	0.088	0.123	0.110	0.105	0.132
	3	1.473	0.907	0.907	0.995	0.855	0.846	0.853
042	2	0.025	0.115	0.114	0.112	0.115	0.112	0.114
	3	1.368	0.862	0.863	0.937	0.823	0.835	0.833
043	2	0.074	0.133	0.133	0.135	0.129	0.124	0.127
	3	1.319	0.979	0.911	1.378	0.878	0.869	0.881
044	1	0.087	0.033	0.034	0.034	0.033	0.033	0.033
	2				0.068			0.084
	3				0.994			0.912
	2+3	1.226	0.993	0.992		0.992	0.992	
045	1	0.048	0.018	0.018	0.019	0.018	0.018	0.019
	2				0.062			0.091
	3				0.988			0.876
	2+3	1.255	0.911	0.937		0.952	0.958	
046	1	0.045	0.019	0.019	0.020	0.020	0.020	0.020
	2	0.057	0.112	0.134	0.160	0.103	0.101	0.102
047	1	0.085	0.039	0.039	0.039	0.038	0.039	0.039
	2	0.375	0.198	0.191	0.201	0.191	0.186	0.186
048	1	0.051	0.034	0.034	0.034	0.034	0.033	0.033
	2+3	0.469	0.289	0.287	0.294	0.281	0.282	0.286
	4	0.759	0.516	0.501	0.461	0.471	0.476	0.468
049	1+2	0.281	0.240	0.239	0.245	0.240	0.239	0.244
050	2	0.064	0.111	0.114	0.114	0.115	0.108	0.110
	3	1.294	0.811	0.790	0.830	0.753	0.761	0.768
051	2	0.166	0.153	0.156	0.166	0.154	0.152	0.150
	3	1.183	0.955	0.800	0.677	0.765	0.763	0.772
052	1	0.022	0.011	0.011	0.012	0.011	0.011	0.011
	2	0.197	0.138	0.152	0.158	0.138	0.127	0.134
	3	0.751	0.676	0.676	0.756	0.627	0.633	0.646
053	1	0.064	0.070	0.070	0.068	0.074	0.072	0.070
	2	0.279	0.166	0.169	0.166	0.167	0.168	0.164
054	1	0.037	0.041	0.041	0.039	0.036	0.037	0.038

Molecule ID	Peak#	ƒcomp	ƒexp (continued)			ƒm,exp (continued)		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
	2	0.218	0.151	0.152	0.152	0.147	0.145	0.142
	3	0.094	0.102	0.089	0.093	0.122	0.125	0.126
	4	0.310	0.291	0.303	0.291	0.273	0.266	0.265
055	1	0.383	0.280	0.279	0.281	0.285	0.282	0.282
	2	0.047	0.161	0.155	0.168	0.146	0.142	0.150
056	1	0.587	0.381	0.380	0.378	0.386	0.385	0.384
057	1	0.101	0.066	0.067	0.067	0.067	0.068	0.068
	2	0.256	0.140	0.140	0.140	0.144	0.142	0.142
058	1	0.056	0.069	0.069	0.071	0.070	0.070	0.071
	2	0.230	0.147	0.145	0.144	0.145	0.142	0.143
059	1	0.080	0.068	0.068	0.068	0.067	0.068	0.068
	2	0.279	0.155	0.156	0.156	0.160	0.159	0.160
060	1	0.128	0.080	0.079	0.079	0.081	0.080	0.080
	2	0.243	0.176	0.176	0.174	0.178	0.175	0.175
061	1	0.116	0.092	0.092	0.092	0.092	0.092	0.092
	2	0.060	0.061	0.061	0.064	0.078	0.077	0.067
	3	0.386	0.410	0.428	0.392	0.410	0.416	0.428
062	1	0.449	0.369	0.376	0.374	0.368	0.378	0.375
	2	0.227	0.190	0.187	0.223	0.204	0.207	0.204
063	1+2	0.679	0.509	0.495	0.506	0.511	0.499	0.506
	3	0.155	0.172	0.175	0.163	0.169	0.165	0.167
064	1	0.045	0.023	0.023	0.025	0.027	0.027	0.026
	2+3	0.599	0.438	0.440	0.459	0.432	0.437	0.433
065	1	0.029	0.018	0.018	0.020	0.018	0.018	0.019
	2	0.879	0.979	0.986	0.132	0.878	0.886	0.114
066	1	0.518	0.360	0.361	0.372	0.370	0.376	0.375
	2	0.110	0.243	0.241	0.265	0.211	0.209	0.209
067	1	0.113	0.036	0.039	0.040	0.042	0.043	0.042
	2	0.272	0.144	0.133	0.156	0.135	0.135	0.133
	3	0.508	0.561	0.547	0.697	0.480	0.476	0.485
068	1	0.115	0.095	0.094	0.095	0.093	0.093	0.092
	2	0.223	0.079	0.086	0.082	0.060	0.069	0.068
	3+4	0.598	0.419	0.419	0.415	0.433	0.436	0.436
069	1	0.864	0.576	0.580	0.577	0.573	0.581	0.577
070	1	0.817	0.722	0.579	0.578	0.585	0.579	0.579

Molecule ID	Peak#	ƒcomp	ƒexp (continued)			ƒm,exp (continued)		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
	2	0.163	0.175	0.181	0.167	0.184	0.177	0.174
071	2	0.534	0.427	0.427	0.412	0.402	0.399	0.407
072	2	0.601	0.444	0.438	0.429	0.443	0.435	0.437
073	1	0.452	0.404	0.408	0.412	0.400	0.402	0.404
	2	0.447	0.315	0.311	0.288	0.290	0.295	0.297
074	1	0.511	0.446	0.446	0.454	0.447	0.449	0.458
	2	0.554	0.447	0.437	0.383	0.385	0.383	0.389
075	2+3	0.585	0.504	0.503	0.501	0.487	0.489	0.490
076	1	0.672	0.503	0.501	0.456	0.482	0.489	0.486
077	1	0.210	0.142	0.144	0.148	0.142	0.144	0.145
	2+3	0.859	0.789	0.774	0.371	0.686	0.691	0.437
078	1+2	0.633	0.610	0.608	0.590	0.596	0.596	0.586
	3+4	0.248	0.374	0.392	0.317	0.304	0.302	0.297
079	1	0.573	0.450	0.452	0.428	0.460	0.466	0.467
	2				0.376			0.317
080	1	0.130	0.124	0.123	0.122	0.123	0.123	0.122
	2	1.005	0.376	0.378	0.269	0.381	0.391	0.361
	3	0.612	1.000	1.000	1.171	0.837	0.847	0.924
	2+3				1.343			1.285
081	1	0.161	0.154	0.161	0.158	0.146	0.148	0.147
	2	0.515	0.411	0.412	0.370	0.355	0.370	0.354
082	1	0.088	0.041	0.041	0.041	0.043	0.043	0.043
	2	0.105	0.111	0.108	0.113	0.091	0.090	0.097
	3	1.301	0.990	0.990	1.055	0.905	0.912	0.919
083	1	0.150	0.154	0.153	0.143	0.141	0.140	0.142
	2	1.525	1.000	1.000	1.289	1.147	1.157	1.170
084	1	0.076	0.030	0.033	0.031	0.025	0.027	0.025
	2	0.110	0.168	0.179	0.151	0.164	0.158	0.162
085	1	0.288	0.207	0.215	0.192	0.189	0.196	0.192
	2	0.288	0.382	0.439	0.437	0.341	0.356	0.343
086	1	0.113	0.067	0.050	0.054	0.061	0.061	0.063
	2	0.072	0.060	0.046	0.067	0.046	0.053	0.053
087	1	0.311	0.186	0.195	0.194	0.190	0.195	0.194
088	1	0.123	0.074	0.073	0.074	0.074	0.074	0.074
	2	0.001	0.047	0.053	0.047	0.051	0.047	0.049

Molecule ID	Peak#	ƒcomp	ƒexp (continued)			ƒm,exp (continued)		
			Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
089	1	0.090	0.057	0.057	0.057	0.057	0.056	0.057
	2	0.212	0.149	0.150	0.152	0.131	0.137	0.140
090	1	0.139	0.082	0.082	0.081	0.081	0.082	0.082
	2	0.229	0.189	0.189	0.193	0.176	0.177	0.179
091	1	0.604	0.400	0.410	0.393	0.410	0.413	0.407
092	1	0.122	0.089	0.088	0.087	0.087	0.086	0.086
093	1	0.120	0.146	0.146	0.148	0.147	0.146	0.149
	2	0.961	0.844	0.844	1.093	0.791	0.807	0.942
094	1	0.212	0.182	0.182	0.186	0.178	0.180	0.182
	2				0.371			0.419
095	1	0.071	0.091	0.093	0.083	0.093	0.093	0.079
	2	0.896	0.578	0.598	0.561	0.528	0.526	0.535
096	1+2	0.054	0.074	0.073	0.074	0.073	0.072	0.074
097	1	0.189	0.096	0.094	0.096	0.092	0.091	0.092
	2	0.620	0.545	0.538	0.508	0.508	0.524	0.530
098	1+2	0.154	0.053	0.054		0.053	0.054	
	3+4+5	0.515	0.391	0.391	0.290	0.337	0.338	0.348
099	1	0.127	0.085	0.085	0.078	0.081	0.081	0.080
	2+3+4	0.673	0.592	0.604	0.532	0.522	0.541	0.536

Appendix B.2

Molecule ID	Peak #	Range / Count (cm-1)			NRMSE (normalized root mean square error)			Percent Error (ƒexp vs. ƒm,exp)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
000	1	204	131	59	0.50%	0.39%	0.20%	12.5%	12.8%	11.3%
001	1	168	130	58	0.60%	0.59%	0.25%	5.7%	5.4%	9.6%
002	1	247	173	65	0.24%	0.46%	0.28%	8.4%	7.0%	6.3%
003	1	245	176	71	0.64%	0.72%	0.45%	5.5%	4.8%	7.8%
004	1	196	131	60	0.36%	0.43%	0.43%	3.6%	3.3%	6.1%
005	1	161	120	61	0.91%	0.93%	0.37%	2.5%	2.2%	7.7%
006	1	213	135	62	0.35%	0.37%	0.29%	1.6%	0.0%	4.3%
007	1	201	124	62	2.49%	2.00%	0.49%	0.4%	1.2%	5.6%
008	1	176	126	57	0.68%	1.10%	0.25%	5.5%	5.2%	9.8%

Molecule ID	Peak #	Range / Count (cm-1) (continued)			NRMSE (normalized root mean square error) (continued)			Percent Error (fexp vs. fm,exp) (continued)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
009	1	204	130	56	0.25%	0.23%	0.24%	8.9%	9.5%	9.1%
010	1	219	154	70	0.96%	1.03%	0.36%	21.3%	19.3%	24.0%
011	1	233	141	62	1.29%	0.62%	0.36%	7.4%	6.6%	11.4%
012	1	252	161	74	0.68%	0.39%	0.38%	0.0%	3.0%	1.5%
	2	144	99	46	0.55%	1.14%	0.37%	29.2%	26.1%	28.2%
013	1	250	166	77	0.80%	1.10%	0.35%	0.0%	0.0%	1.1%
	2	189	123	55	0.76%	0.74%	0.36%	6.8%	6.4%	7.8%
014	1	154	141	58	0.30%	0.38%	0.30%	6.8%	2.4%	4.7%
	2	332	222	91	0.24%	0.27%	0.27%	16.1%	16.5%	13.1%
015	1	196	123	55	0.44%	0.65%	0.45%	13.0%	15.0%	9.6%
	2	420	289	127	2.99%	2.42%	0.48%	19.4%	20.4%	36.0%
016	1	209	148	65	0.41%	0.26%	0.29%	2.6%	5.0%	2.6%
017	2	178	121	67	0.16%	0.21%	0.23%	15.3%	14.6%	13.8%
018	2	248	196	97	0.29%	0.27%	0.21%	7.7%	3.7%	12.6%
019	2	317	213	100	0.37%	0.39%	0.18%	15.5%	14.1%	25.5%
020	1+2	189	133	66	0.24%	0.23%	0.27%	0.2%	0.8%	1.0%
021	1+2	218	138	70	0.37%	0.39%	0.47%	2.6%	3.2%	3.5%
022	1+2	165	110	52	0.54%	0.40%	0.36%	2.9%	2.7%	1.8%
023	2	279	179	95	0.37%	0.34%	0.30%	0.0%	1.2%	5.3%
024	2	219	158	83	0.43%	0.35%	0.17%	3.1%	4.4%	15.3%
025	2	224	144	73	0.40%	0.34%	0.27%	18.1%	17.3%	14.6%
026	2	186	109	64	0.70%	0.66%	0.81%	0.2%	0.2%	1.2%
	3	242	157	95	1.48%	4.11%	0.58%	16.0%	8.6%	3.9%
027	1	105	69	40	0.78%	0.86%	0.40%	1.0%	0.3%	1.0%
	2+3	241	161	91	0.98%	1.16%	0.47%	3.6%	4.4%	4.4%
028	1	150	97	47	0.37%	0.57%	0.36%	0.9%	1.6%	1.4%
029	1			75			0.86%			14.0%
	2	311	152	84	0.52%	0.63%	0.38%	4.8%	4.5%	4.9%
030	1	188	122	68	0.22%	0.28%	0.25%	2.2%	0.4%	3.3%
	2			121			0.59%			1.7%
031	2	212	95	96	0.31%	0.25%	0.25%	7.9%	4.1%	5.0%
032	1	232	159	73	0.30%	0.29%	0.22%	8.8%	7.4%	6.9%
033	1	104	66	51	0.40%	0.34%	0.45%	0.3%	0.0%	0.1%
	2	113	76	53	0.71%	1.28%	0.55%	12.9%	6.9%	9.7%
034	2+3	306	207	89	0.61%	0.71%	0.40%	0.3%	0.0%	3.5%
035	1	193	160	67	2.32%	2.35%	0.19%	14.3%	13.9%	20.3%

Molecule ID	Peak #	Range / Count (cm-1) (continued)			NRMSE (normalized root mean square error) (continued)			Percent Error (ƒexp vs. ƒm,exp) (continued)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
036	1	179	128	75	0.54%	0.61%	0.41%	2.1%	1.2%	1.0%
	2	311	208	122	3.13%	1.89%	0.52%	9.8%	7.4%	3.4%
037	1+2	272	177	78	0.39%	0.41%	0.37%	0.9%	0.0%	1.2%
038	1+2	197	136	93	0.69%	0.51%	0.34%	1.7%	0.9%	5.4%
039	1	228	161	71	0.39%	0.37%	0.23%	0.8%	0.6%	0.7%
040	1	116	84	42	3.04%	3.18%	0.36%	4.4%	5.2%	8.1%
	2	233	160	90	1.56%	1.95%	0.42%	11.6%	9.3%	7.1%
041	2	81	39	34	#####	#####	0.36%	26.4%	19.3%	7.5%
	3	130	87	68	3.74%	5.34%	0.44%	5.7%	6.7%	14.3%
042	2	77	50	34	0.91%	1.09%	0.42%	0.0%	1.8%	1.8%
	3	140	102	67	2.22%	2.39%	0.63%	4.5%	3.2%	11.1%
043	2	120	80	38	1.85%	1.62%	0.71%	3.0%	6.8%	5.9%
	3	238	154	75	1.26%	2.32%	0.56%	10.3%	4.6%	36.1%
044	1	48	33	29	2.91%	2.80%	1.40%	0.0%	2.9%	0.3%
	2			37			0.43%			23.5%
	3			96			0.37%			8.2%
	2+3	94	61		2.75%	2.72%		0.1%	0.0%	
045	1	76	47	27	4.57%	4.07%	2.15%	0.0%	0.0%	0.5%
	2			34			0.35%			47.1%
	3			99			0.48%			11.3%
	2+3	161	101		4.97%	3.70%		4.5%	2.2%	
046	1	97	64	38	1.64%	1.06%	0.51%	5.3%	5.3%	0.5%
	2	139	103	59	1.16%	0.46%	0.32%	8.0%	24.6%	36.3%
047	1	110	69	38	0.92%	0.62%	0.42%	2.6%	0.0%	0.5%
	2	190	124	74	1.03%	1.89%	0.34%	3.5%	2.6%	7.2%
048	1	113	75	50	2.19%	3.28%	1.10%	0.0%	2.9%	2.4%
	2+3	158	105	74	0.92%	1.09%	0.64%	2.8%	1.7%	2.9%
	4	222	154	99	0.55%	0.51%	0.43%	8.7%	5.0%	1.5%
049	1+2	185	124	83	0.35%	0.35%	0.32%	0.0%	0.0%	0.1%
050	2	128	56	41	2.13%	0.69%	0.45%	3.6%	5.3%	3.2%
	3	159	110	77	2.17%	1.98%	0.55%	7.2%	3.7%	7.5%
051	2	80	50	40	2.74%	0.98%	0.55%	0.7%	2.6%	9.6%
	3	138	86	65	2.02%	1.84%	0.56%	19.9%	4.6%	14.0%
052	1	75	44	30	3.62%	3.51%	2.10%	0.0%	0.0%	8.1%
	2	158	105	69	4.61%	1.93%	0.43%	0.0%	16.4%	15.3%
	3	159	112	74	3.01%	2.47%	0.53%	7.2%	6.4%	14.7%

Molecule ID	Peak #	Range / Count (cm-1) (continued)			NRMSE (normalized root mean square error) (continued)			Percent Error (ƒexp vs. ƒm,exp) (continued)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
053	1	212	143	77	0.31%	0.54%	0.29%	5.7%	2.9%	3.1%
	2	357	213	112	0.46%	0.65%	0.41%	0.6%	0.6%	1.2%
054	1	173	120	58	0.56%	0.40%	0.28%	12.2%	9.8%	4.6%
	2	276	185	92	0.88%	0.91%	0.46%	2.6%	4.6%	6.4%
	3	247	174	90	1.75%	1.52%	0.45%	19.6%	40.4%	35.5%
	4	306	207	101	1.85%	7.54%	0.94%	6.2%	12.2%	8.8%
055	1	276	190	75	1.30%	1.24%	0.40%	1.8%	1.1%	0.5%
	2	320	215	85	1.40%	0.70%	0.40%	9.3%	8.4%	10.7%
056	1	154	100	68	0.43%	0.42%	0.36%	1.3%	1.3%	1.4%
057	1	160	101	43	0.32%	0.39%	0.27%	1.5%	1.5%	0.1%
	2	166	108	47	0.59%	0.69%	0.44%	2.9%	1.4%	1.0%
058	1	145	111	41	0.44%	0.50%	0.30%	1.4%	1.4%	0.3%
	2	235	148	55	0.73%	0.93%	0.50%	1.4%	2.1%	1.1%
059	1	157	108	52	0.39%	0.49%	0.45%	1.5%	0.0%	0.6%
	2	167	113	55	0.77%	0.70%	0.45%	3.2%	1.9%	2.6%
060	1	83	120	55	0.71%	0.61%	0.43%	1.3%	1.3%	1.4%
	2	99	138	63	0.86%	0.60%	0.38%	1.1%	0.6%	0.9%
061	1	108	97	52	0.58%	0.41%	0.37%	0.0%	0.0%	0.1%
	2	115	76	40	1.06%	1.05%	0.59%	27.9%	26.2%	4.5%
	3	260	168	96	1.91%	1.21%	0.75%	0.0%	2.8%	9.2%
062	1	210	136	60	0.40%	0.55%	0.42%	0.3%	0.5%	0.3%
	2	253	160	74	0.99%	1.14%	0.48%	7.4%	10.7%	8.3%
063	1+2	123	84	38	3.13%	1.43%	0.56%	0.4%	0.8%	0.1%
	3	220	152	73	0.97%	0.70%	0.63%	1.7%	5.7%	2.8%
064	1	170	91	39	2.44%	1.64%	0.69%	17.4%	17.4%	4.0%
	2+3	144	106	63	3.74%	2.75%	0.96%	1.4%	0.7%	5.6%
065	1	109	71	41	0.44%	0.54%	0.35%	0.0%	0.0%	6.5%
	2	342	242	95	1.20%	0.78%	0.45%	10.3%	10.1%	13.7%
066	1	277	112	70	0.47%	0.49%	0.27%	2.8%	4.2%	0.8%
	2	338	222	131	1.73%	1.42%	0.35%	13.2%	13.3%	21.2%
067	1	99	132	40	0.46%	0.44%	0.44%	16.7%	10.3%	5.3%
	2	198	131	80	0.60%	1.36%	0.63%	6.2%	1.5%	14.7%
	3	202	147	80	1.22%	2.17%	0.44%	14.4%	13.0%	30.5%
068	1	185	128	79	0.29%	0.40%	0.30%	2.1%	1.1%	2.7%
	2	188	130	81	0.74%	0.98%	0.36%	24.1%	19.8%	16.8%
	3+4	290	196	123	2.25%	1.53%	0.41%	3.3%	4.1%	4.9%

Molecule ID	Peak #	Range / Count (cm-1) (continued)			NRMSE (normalized root mean square error) (continued)			Percent Error (ƒexp vs. ƒm,exp) (continued)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
069	1	217	144	61	0.36%	0.38%	0.31%	0.5%	0.2%	0.1%
070	1	152	106	60	0.38%	0.43%	0.29%	19.0%	0.0%	0.1%
	2	173	123	70	1.61%	1.42%	0.35%	5.1%	2.2%	4.7%
071	2	198	159	57	0.35%	0.28%	0.24%	5.9%	6.6%	1.2%
072	2	199	119	60	0.49%	0.36%	0.40%	0.2%	0.7%	2.1%
073	1	257	178	67	0.78%	0.50%	0.49%	1.0%	1.5%	1.9%
	2	313	224	83	1.25%	1.33%	0.46%	7.9%	5.1%	2.9%
074	1	144	103	48	0.44%	0.53%	0.35%	0.2%	0.7%	0.9%
	2	270	180	83	0.73%	0.97%	0.38%	13.9%	12.4%	1.5%
075	2+3	209	153	73	1.21%	0.79%	0.50%	3.4%	2.8%	2.2%
076	1	231	126	62	0.52%	0.55%	0.27%	4.2%	2.4%	6.6%
077	1	211	151	78	0.57%	0.62%	0.28%	0.0%	0.0%	2.0%
	2+3	280	183	87	2.45%	3.87%	0.33%	13.1%	10.7%	18.0%
078	1+2	218	157	69	0.26%	0.29%	0.22%	2.3%	2.0%	0.7%
	3+4	411	277	123	4.26%	1.59%	0.30%	18.7%	23.0%	6.3%
079	1	128	145	57	0.22%	0.35%	0.18%	2.2%	3.1%	9.0%
	2			144			0.33%			15.8%
080	1	178	105	56	3.76%	3.83%	0.94%	0.8%	0.0%	0.5%
	2	122	80	42	0.76%	0.61%	0.66%	1.3%	3.4%	34.2%
	3	335	209	113	4.65%	5.08%	0.68%	16.3%	15.3%	21.1%
	2+3			71			0.61%			4.3%
081	1	195	123	59	0.30%	0.29%	0.30%	5.2%	8.1%	7.2%
	2	196	132	60	1.02%	0.70%	0.47%	13.6%	10.2%	4.3%
082	1	137	90	40	0.66%	0.86%	0.31%	4.9%	4.9%	4.4%
	2	237	146	66	2.09%	1.31%	0.52%	18.0%	16.7%	14.2%
	3	252	182	74	0.80%	0.82%	0.52%	8.6%	7.9%	12.9%
083	1	172	115	61	0.89%	1.04%	0.37%	8.4%	8.5%	0.8%
	2	202	131	73	6.69%	6.04%	0.49%	14.7%	15.7%	9.2%
084	1	131	93	35	1.08%	1.12%	0.41%	16.7%	18.2%	17.8%
	2	211	147	61	3.32%	0.66%	0.66%	2.4%	11.7%	7.1%
085	1	151	148	54	0.53%	0.29%	0.22%	8.7%	8.8%	0.0%
	2	213	220	75	4.71%	1.36%	0.27%	10.7%	18.9%	21.5%
086	1	149	89	48	0.44%	0.46%	0.30%	9.0%	22.0%	16.9%
	2	413	289	143	0.32%	1.99%	0.83%	23.3%	15.2%	20.3%
087	1	163	127	58	0.55%	0.32%	0.33%	2.2%	0.0%	0.4%
088	1	114	76	34	1.07%	0.88%	0.56%	0.0%	1.4%	0.0%

Molecule ID	Peak #	Range / Count (cm-1) (continued)			NRMSE (normalized root mean square error) (continued)			Percent Error (fexp vs. fm,exp) (continued)		
		Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid	Five-point grid	Three-point grid	Variable-point grid
	2	205	142	62	0.59%	0.62%	0.24%	8.5%	11.3%	5.6%
089	1	122	72	36	1.16%	1.14%	0.75%	0.0%	1.8%	0.9%
	2	203	162	77	0.73%	0.96%	0.67%	12.1%	8.7%	7.6%
090	1	136	90	39	1.67%	0.72%	0.48%	1.2%	0.0%	1.3%
	2	226	155	68	0.88%	1.05%	0.39%	6.9%	6.3%	7.2%
091	1	239	150	71	0.52%	0.48%	0.32%	2.5%	0.7%	3.5%
092	1	142	91	44	0.50%	0.57%	0.42%	2.2%	2.3%	0.8%
093	1	127	83	46	1.06%	1.07%	0.43%	0.7%	0.0%	1.2%
	2	157	107	63	1.11%	0.44%	0.41%	6.3%	4.4%	13.8%
094	1	146	95	49	0.60%	0.69%	0.49%	2.2%	1.1%	2.2%
	2			62			0.37%			12.9%
095	1	199	131	78	9.08%	7.79%	0.35%	2.2%	0.0%	5.2%
	2	135	85	57	1.03%	0.87%	0.59%	8.7%	12.0%	4.7%
096	1+2	208	161	72	1.59%	1.90%	0.34%	1.4%	1.4%	0.1%
097	1	189	123	63	0.70%	0.73%	0.21%	4.2%	3.2%	4.0%
	2	176	123	65	0.68%	0.74%	0.45%	6.8%	2.6%	4.5%
098	1+2	216	148		1.05%	0.96%		0.0%	0.0%	
	3+4+5	256	171	102	5.70%	5.76%	0.48%	13.8%	13.6%	19.7%
099	1	114	80	46	0.49%	0.47%	0.35%	4.7%	4.7%	3.0%
	2+3+4	178	119	67	1.65%	1.83%	0.48%	11.8%	10.4%	0.8%
Averages:		195	133	68	1.41%	1.30%	0.45%	6.4%	6.2%	7.7%
Ranking System:		<100			<0.5%			<5%		
		100-200			0.5-1%			5-15%		
		>200			>1%			>15%		

Appendix B.3

Molecule ID	Peak#	Percent Deviation of fexp in the three grids	Percent Deviation of fm,exp in the three grids
000	1	2%	2%
001	1	2%	2%
002	1	1%	1%
003	1	5%	2%

Molecule ID	Peak#	Percent Deviation of ƒexp in the three grids (continued)	Percent Deviation of ƒm,exp in the three grids (continued)
004	1	4%	2%
005	1	17%	12%
006	1	7%	2%
007	1	3%	4%
008	1	5%	0%
009	1	1%	1%
010	1	8%	4%
011	1	0%	5%
012	1	0%	3%
	2	8%	9%
013	1	3%	2%
	2	2%	0%
014	1	9%	2%
	2	3%	7%
015	1	1%	4%
	2	23%	4%
016	1	3%	2%
017	2	1%	1%
018	2	19%	1%
019	2	12%	2%
020	1+2	2%	1%
021	1+2	7%	1%
022	1+2	3%	2%
023	2	5%	1%
024	2	18%	1%
025	2	27%	1%
026	2	0%	1%
	3	16%	6%
027	1	2%	2%
	2+3	1%	1%
028	1	2%	2%
029	1	0%	0%
	2	0%	1%
030	1	5%	2%
	2		
031	2	1%	2%
032	1	2%	1%
033	1	2%	2%

Molecule ID	Peak#	Percent Deviation of ƒexp in the three grids (continued)	Percent Deviation of ƒm,exp in the three grids (continued)
	2	7%	0%
034	2+3	3%	1%
035	1	19%	16%
036	1	4%	3%
	2	13%	3%
037	1+2	3%	2%
038	1+2	7%	1%
039	1	3%	1%
040	1	4%	1%
	2	4%	1%
041	2	29%	20%
	3	9%	1%
042	2	3%	3%
	3	8%	1%
043	2	2%	4%
	3	34%	1%
044	1	3%	1%
	2	0%	0%
	3	0%	0%
	2+3		
045	1	3%	4%
	2	0%	0%
	3	0%	0%
	2+3		
046	1	3%	2%
	2	30%	2%
047	1	1%	3%
	2	5%	3%
048	1	1%	4%
	2+3	2%	2%
	4	12%	2%
049	1+2	2%	2%
050	2	3%	6%
	3	5%	2%
051	2	8%	3%
	3	41%	1%
052	1	11%	3%
	2	13%	8%

Molecule ID	Peak#	Percent Deviation of f_{exp} in the three grids (continued)	Percent Deviation of $f_{m,exp}$ in the three grids (continued)
	3	11%	3%
053	1	3%	6%
	2	2%	2%
054	1	4%	4%
	2	1%	3%
	3	14%	3%
	4	4%	3%
055	1	1%	1%
	2	8%	5%
056	1	1%	1%
057	1	2%	1%
	2	0%	2%
058	1	3%	1%
	2	2%	2%
059	1	0%	2%
	2	1%	1%
060	1	2%	1%
	2	1%	2%
061	1	0%	0%
	2	5%	16%
	3	9%	4%
062	1	2%	3%
	2	16%	1%
063	1+2	3%	2%
	3	7%	2%
064	1	8%	4%
	2+3	5%	1%
065	1	10%	3%
	2	645%	676%
066	1	3%	2%
	2	9%	1%
067	1	10%	2%
	2	15%	1%
	3	22%	2%
068	1	1%	1%
	2	9%	13%
	3+4	1%	1%
069	1	1%	1%

Molecule ID	Peak#	Percent Deviation of ƒexp in the three grids (continued)	Percent Deviation of ƒm,exp in the three grids (continued)
070	1	25%	1%
	2	9%	6%
071	2	4%	2%
072	2	4%	2%
073	1	2%	1%
	2	9%	2%
074	1	2%	2%
	2	17%	1%
075	2+3	1%	1%
076	1	10%	1%
077	1	4%	2%
	2+3	113%	58%
078	1+2	3%	2%
	3+4	24%	2%
079	1	6%	1%
	2		
080	1	1%	1%
	2	41%	8%
	3	15%	9%
	2+3		
081	1	4%	1%
	2	11%	4%
082	1	1%	1%
	2	4%	7%
	3	6%	1%
083	1	8%	1%
	2	22%	2%
084	1	10%	8%
	2	19%	4%
085	1	12%	4%
	2	13%	4%
086	1	31%	3%
	2	31%	13%
087	1	5%	3%
088	1	1%	0%
	2	14%	8%
089	1	0%	2%
	2	2%	7%

Molecule ID	Peak#	Percent Deviation of ƒexp in the three grids (continued)	Percent Deviation of ƒm,exp in the three grids (continued)
090	1	1%	1%
	2	2%	2%
091	1	4%	2%
092	1	3%	1%
093	1	1%	2%
	2	23%	16%
094	1	2%	2%
	2	0%	0%
095	1	12%	18%
	2	7%	2%
096	1+2	2%	3%
097	1	2%	2%
	2	7%	4%
098	1+2		
	3+4+5		
099	1	9%	1%
	2+3+4	14%	4%
Averages:		11.64%	7.40%
Ranking System:		<5%	<3%
		5-10%	3-10%
		>10%	>10%