Department of Materials
Imperial College

BEng Thesis

Predicting the colour of organic dyes using theory and simulation

Ruochen Zhao

Supervisors: Prof A. Mostofi and Dr J.C.A Prentice

April, 2019

Abstract

Density-functional theory and time-dependent density-functional theory are used to calculate excited states of neutral phenolphthalein, neutral thymol blue and phenolphthalein in the doubly negative charged state, in implicit and explicit water solvents using PBE and B3LYP functionals. Ideally, we would like to use B3LYP for all calculations, but its relative high cost dictates that we need to use PBE. The difference between the excitation energies calculated by PBE and B3LYP functionals, termed the spectral warping shift, is investigated in depth. It is found that across all three systems, the spectral warping shifts are uncorrelated to the excited energies calculated by PBE and a particularly useful way to extract the average spectral warping shift is to perform calculations on randomly sampled configurations from a molecular dynamics simulation of the system of interest. It is also found that the spectral warping shifts are uncorrelated to the individual bond rotations of the systems.

Collaboration and Supervision

This thesis presents work carried out from October 2018 to March 2019, in the Theory and Simulation of Materials group at the Department of Materials, Imperial College London, under the supervision of Professor Arash Mostofi and Dr Joseph C. A. Prentice. This thesis is entirely my own work apart from the section on molecular dynamics, where the input files were prepared by Dr Joseph C. A. Prentice.

Acknowledgements

First of all, I would like to thank my supervisors, Professor Arash Mostofi and Dr Joseph C. A. Prentice for their willingness and patience in supporting every aspect of my research project. Thanks to both of you for making it such an enjoyable experience in my final year as an undergraduate.
I would also like to thank Dr Johannes Lischner and Dr Éammon Murray for their fantastic lectures in density-functional theory, which forms the theoretical basis of my work.
I gratefully acknowledge the Research Computing Service staff at Imperial College London for teaching me how to use supercomputers.
Many thanks also to all my friends for their company during the nights when I was writing this thesis in the library and finally, to my dearest family.

Contents

1 Aims and Context

In this project, we use density-functional theory and time-dependent density-functional theory implemented in software packages which run on supercomputers to calculate excited states of organic dyes in solution. Ideally, we would use hybrid functionals to carry out all the calculations because they are known to be relatively accurate, but their computational costs are prohibitive for large systems, such as those with explicit solvents. In practice, we use less expensive functionals such as generalised gradient approximations (GGA) and they tend to produce incorrect excitation energies. Nevertheless, it was found by previous works that the less accurate results produced by GGA can be corrected by a linear transformation of the excitation energies where the hybrid functional results are used as benchmark—this method is termed 'spectral warping'. Our objective for this work is to investigate the accuracy of the spectral warping method because it has only been exploited by a small number of researchers for specific systems and has not yet been tested in depth. We will check the validity of this method when it is applied to large systems.

2 Literature review

In this section, we will begin by introducing the theoretical tools used in materials modelling and explain how we use them to calculate excited states of materials. Then, we will progress more deeply into the problem of interest: modelling absorption spectra of molecules in solution, while outlining progress in the state-of-the-art research.

2.1 Materials modelling and the many-body problem

Any material, irrespective of its complexity, is merely a collection of nuclei and electrons in the eyes of a reductionist. We may rewrite this statement into a simple equation [1]:

$$\text{materials} = \text{electrons} + \text{nuclei}$$

If we could model the electrons and nuclei, then the properties of materials can be understood at the atomic level. They are quantum particles that have kinetic energy and potential energy that arises from the Coulomb interactions between each other. In many cases, we use the Born-Oppenheimer approximation and treat the nuclei as classical particles. Thus, the nuclei only act as a source of external potential. [2] We may therefore describe a system with N interacting electrons by the many-body Schrödinger equation:

$$\hat{H}\Psi\left(\mathbf{r}_1, \ldots, \mathbf{r}_N\right) = E\Psi\left(\mathbf{r}_1, \ldots, \mathbf{r}_N\right) \tag{2.1}$$

where Ψ is the many-body wavefunction of N electrons, $\mathbf{r}_1, \ldots, \mathbf{r}_N$ are the positions of the electrons, E is the total energy of the system and $\hat{H}$, the many-body Hamiltonian, is given by (in Hartree units)

$$\hat{H} = \sum_{i=1}^{N} -\frac{\nabla_i^2}{2} + \sum_{i=1}^{N} V\left(\mathbf{r}_i\right) + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{N} 1/\left(|\mathbf{r}_i - \mathbf{r}_j|\right) \tag{2.2}$$

where V is the potential energy of the electrons. If we solve this equation, we would obtain the wavefunction that describes the probability of finding the electrons at particular locations. However, this equation is prohibitively expensive to solve once N becomes large due to the fact that it depends on $3N$ variables. On the other hand, if our primary goal is to model material properties, such as forces, structures and phase transitions, it is not necessary to find the wavefunctions. These properties can be readily obtained from the energy of the system. Rather than solving the many-body equation and finding the wavefunction, we can work out the energy in a smarter way. This is when the Hohenberg-Kohn Theorem [3] becomes important, which states that:

> The ground state energy is a unique functional of the electron density $n(\mathbf{r})$

where the electron density is given by:

$$n(\mathbf{r}) = N\int \left|\Psi\left(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N\right)\right|^2 d\mathbf{r}_2 \ldots d\mathbf{r}_N \tag{2.3}$$

The electron density is 3-dimensional, meaning that we only need to solve N 3-dimensional equations rather than one $3N$-dimensional equation, which is less computationally expensive. Although the theorem is useful, it does not provide information regarding the exact form of the energy functional. We can obtain the energy from quantum mechanics, which states that $E = \langle\Psi|\hat{H}|\Psi\rangle$ and if we substitute Ψ with $n(\mathbf{r})$ in the expression of the energy, we would find that the first and third terms in the Hamiltonian, namely the kinetic energy and Coulomb energy, cannot be expressed in terms of $n(\mathbf{r})$ only. Thus, it is not possible to calculate the energy using only $n(\mathbf{r})$ without making a new approximation. One approach to this problem is the Kohn-Sham density-functional theory, which will be discussed next.

2.2 Density-functional theory

Density-functional theory (DFT) is a quantum mechanical method used in a wide range of fields such as physics, chemistry and materials science. It was first put on a successful theoretical footing by Kohn and Hohenberg in 1964. [3] Although it was popular among solid state physicists in the 1970s, it was not the method of choice in chemistry and materials science until the 1990s, when significant improvements were made to the description of the exchange and correlation interactions in complex atomic systems. In recent years, we have seen exponential growth in the number of published papers which are related to DFT, making it the most popular computational method for electronic structure today. [4] The success of DFT is due to a number of reasons, such as its transferability and simplicity, as outlined by Giustino. [1] Here, we will briefly summarise the most important aspects of DFT because it would be much easier to introduce the time-dependent DFT knowing the basics of DFT, in the later sections.

2.2.1 The Kohn-Sham equation

Kohn and Sham's approach is to assume the interacting electrons can be described by a system of non-interacting electrons plus an additional energy that accounts for the difference between the non-interacting and interacting systems, where the electron density is the same for both. Because the electrons are not interacting, we have an analytic expression of their kinetic energy in terms of single-electron orbitals. The total energy thus has the following form [5]:

$$E[n] = -\sum_i \int d\mathbf{r}\phi_i^*(\mathbf{r})\frac{\nabla^2}{2}\phi_i(\mathbf{r}) + \int d\mathbf{r} n(\mathbf{r})V(\mathbf{r}) + \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{xc}}[n] \quad (2.4)$$

where ϕ is the Kohn-Sham orbital, the first term is the kinetic energy, the second term is the external potential from the nuclei, the third term is the Hartree energy that is associated with the electronic charge distribution and the last term is the unknown exchange and correlation energy (XC energy). The problem thus reduces to finding the electron density. It is found by Hohenberg and Kohn [3] that the electron density at the ground state is the function that minimises the energy. This is termed the 'Hohenberg-Kohn variational principle' and we find the following expression by applying it to the

energy together with the normalisation condition of the wavefunction:

$$\left[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) + V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \tag{2.5}$$

This is the Kohn-Sham equation [5], where V is the external potential, V_H is the Hartree potential and $V_{\mathrm{xc}}(\mathbf{r})$ is the unknown exchange and correlation potential. This equation has to be solved self-consistently because the potential depends on the density, which depends on the Kohn-Sham orbitals, which depends on the potential (method illustrated in Figure 2.1). The remaining question is to make approximations to V_{xc}, which will be discussed next.

2.2.2 Local density approximations

Local density approximations (LDA) are a family of approximations made to the XC energy that depend only on the electron density at each point in space. They are considered the simplest approximation and a starting point for the construction of more advanced approximations. There are numerous ways of approximating the XC energy using the local density and the most successful one is based on the homogeneous electron gas (HEG) [5], which has the general form:

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}(n(\mathbf{r}))\mathrm{d}\mathbf{r} \tag{2.6}$$

where $\epsilon_{xc}(n)$ is the exchange-correlation energy per particle of a homogeneous electron gas of density n. It is common in practice to decompose the XC energy into the exchange and correlation energy. For an HEG, the exchange energy has an analytic form, whereas the derivation of the analytic forms of the correlation energy requires parameters interpolated from quantum Monte Carlo calculations. Different analytic forms of the correlation energy yield different LDAs, e.g. VWN [6], PZ81 [7] , CP [8]. In spite of the simplicity of LDAs, they show good agreement with experiments for the ionisation energy and bond length calculations for some systems. [9]

2.2.3 Generalised gradient approximations

Generalised gradient approximations (GGA) are modifications of the LDA, that take into account both the density and the gradient of density (also referred to as semi-local functionals). It has the following form:

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}[n(\mathbf{r})]F_{\mathrm{xc}}[n(\mathbf{r}), \nabla n(\mathbf{r})]d\mathbf{r} \tag{2.7}$$

where F_{XC} is the enhancement factor which can contain experimental parameters or be derived from first-principles for different GGAs. The most widely used GGA is the PBE functional, [10] which we will use for our study. While GGA may lead to some improvements over LDAs, for instances when modelling hydrogen bonds, [11] it is found that they both generally underestimate bandgap in solids. [12]

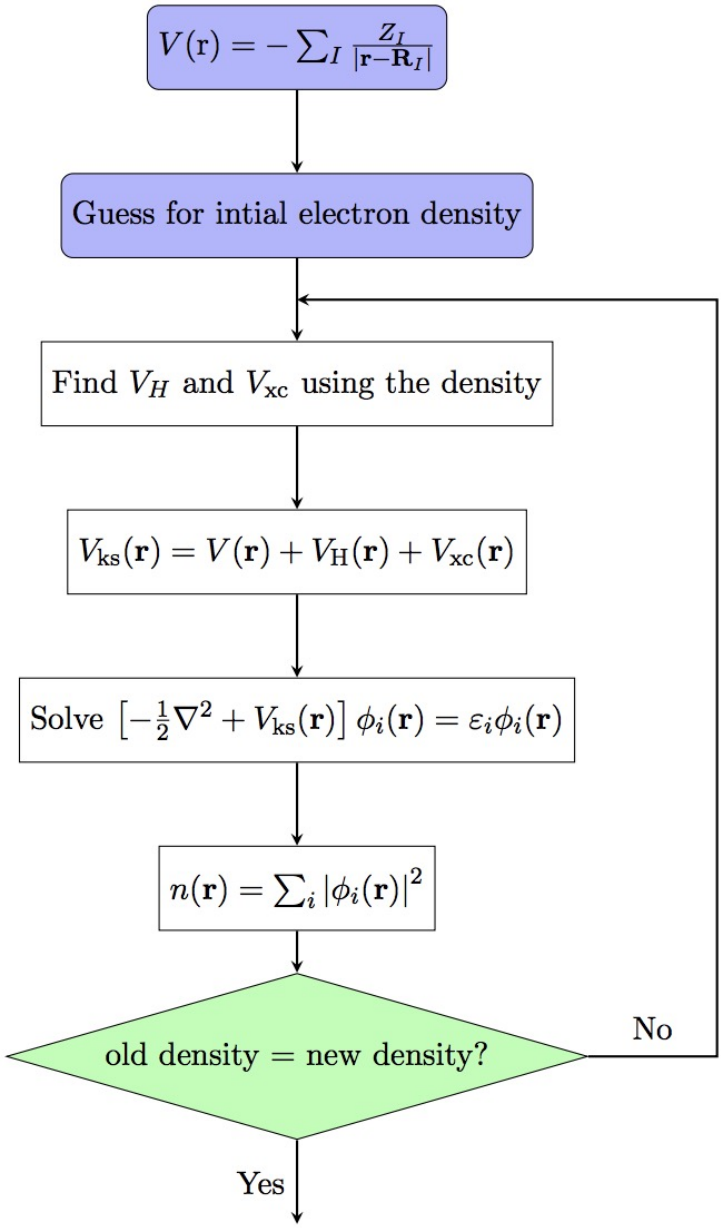


Figure 2.1: Schematic flowchart of self-consistent field method in DFT calculations

2.2.4 Hybrid functionals

Some of the deficiencies of LDA and GGA may be improved by using hybrid functionals. These functionals take in a portion of the exact exchange energy from the Hartree-Fock method [1] and the rest of XC energy from other methods, such as LDA and GGA. One of the most popular hybrid functionals and also the one that we will use, is the B3LYP functional, [13] defined as the following:

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = E_{\mathrm{x}}^{\mathrm{LDA}} + a_0 \left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LDA}}\right) + a_{\mathrm{x}} \left(E_{\mathrm{x}}^{\mathrm{GGA}} - E_{\mathrm{x}}^{\mathrm{LDA}}\right) + E_{\mathrm{c}}^{\mathrm{LDA}} + a_{\mathrm{c}} \left(E_{\mathrm{c}}^{\mathrm{GGA}} - E_{\mathrm{c}}^{\mathrm{LDA}}\right) \tag{2.8}$$

where $a_0 = 0.20, a_{\mathrm{x}} = 0.72,$ and $a_{\mathrm{c}} = 0.81$. $E_{\mathrm{x}}^{\mathrm{HF}}$ is the Hartree-Fock exact exchange functional [1], $E_{\mathrm{x}}^{\mathrm{GGA}}$ is the Becke 88 exchange functional [14], $E_{\mathrm{c}}^{\mathrm{GGA}}$ is the correlation functional of Lee, Yang and Parl [15] and $E_{\mathrm{c}}^{\mathrm{LDA}}$ is the VWN correlation functional. [6] In general, hybrid functionals are considered more accurate but computationally expensive compared to other functionals.

2.3 Time-dependent density-functional theory

Time-dependent density-functional theory (TDDFT) is an extension of DFT to model the dynamics of the physical systems under time-varying potentials, whereas DFT is only concerned with ground state properties. Many familiar concepts follow from DFT in TDDFT and the most popular applications of TDDFT are calculations of excitation energies of molecular systems. These calculations are usually based on linear response theory, which tells us how the electron density changes with respect to small perturbations to the system.

2.3.1 The time-dependent Kohn-Sham equation

Analogously to the Hohenberg-Kohn theorem for the ground state, we also have the Runge-Gross theorem [16] for the excited states, which states:

The time-dependent potential $v(\mathbf{r}, t)$ is a unique functional of the electron density $n(\mathbf{r})$, assuming the potential can be expanded in a Taylor series about a given time

Hence, we can use the Kohn-Sham approach as we did before, yielding the time-dependent Kohn-Sham equation [16]:

$$\left[-\frac{1}{2}\nabla^2 + V(\mathbf{r}, t) + V_{\mathrm{H}}(\mathbf{r}, t) + V_{\mathrm{xc}}(\mathbf{r}, t)\right] \phi_i(\mathbf{r}, t) = i\frac{\partial}{\partial t}\phi_i(\mathbf{r}, t) \tag{2.9}$$

One of the remaining problems is to find an approximation for the time-dependent XC potential, which is similar to the DFT case.

2.3.2 Adiabatic approximations

One approach to approximate the time-dependent XC potential is the adiabatic approximation:

$$V_{\mathrm{xc}}^{\mathrm{A}}(\mathbf{r}, t) = V_{\mathrm{xc}}^{\mathrm{gs}}\left[n_0(\mathbf{r})\right]\big|_{n_0(\mathbf{r})=n(\mathbf{r},t)} \tag{2.10}$$

where $V_{\text{xc}}^{\text{gs}}$ is the ground-state XC potential evaluated at the instantaneous time-dependent density. This equation becomes exact when the system varies infinitely slowly and is at its ground state for any time. This is often not the case in practice, but numerical studies show that this approach works reasonably well in many scenarios. [17, 18, 19] The majority of TDDFT calculations today use the adiabatic LDA (ALDA) or adiabatic GGA (AGGA), by substituting the ground state density with the instantaneous time-dependent density. [2]

2.3.3 Linear response theory

In the practical usage of TDDFT, if the systems of interest are subjected to small perturbations, then we can simplify our calculations by using the linear response theory, instead of solving the full Kohn-Sham equations. This is usually the case of spectroscopy calculations. [2] We consider a ground state system which is subjected to a small perturbation v_1 after a given time:

$$n(\mathbf{r},t) = n_0(\mathbf{r}) + n_1(\mathbf{r},t) + n_2(\mathbf{r},t) + \dots \tag{2.11}$$

where n_0 is the ground state density, n_1 and n_2 are the first- and second-order density response due to v_1. The assumption here is that n_1 dominates all the higher-order terms. Thus, we want to find an expression of n_1, because spectroscopic observables can be calculated from the density response. The time-density response function can be constructed from physical arguments [2], however, we are mainly interested in the frequency domain so we can see which photon frequencies are absorbed by the system, so we can Fourier transform it into the frequency domain, yielding the frequency-density response function:

$$n_1(\mathbf{r},\omega) = \int d^3r' \chi_{\text{ks}}\left(\mathbf{r},\mathbf{r}',\omega\right) v_{1\text{ks}}\left(\mathbf{r}',\omega\right) \tag{2.12}$$

where $v_{1\text{ks}}$, the perturbation to the Kohn-Sham potential is a sum between v_1, the perturbation to the external potential and linearised Hartree and XC potentials:

$$v_{1\text{ks}}(\mathbf{r},\omega) = v_1(\mathbf{r},\omega) + \int d^3r' \left\{ \frac{1}{|\mathbf{r}-\mathbf{r}'|} + f_{\text{xc}}\left(\mathbf{r},\mathbf{r}',\omega\right) \right\} n_1\left(\mathbf{r}',\omega\right), \tag{2.13}$$

$$f_{\text{xc}}\left(\mathbf{r},\mathbf{r}',\omega\right) = \left. \frac{\delta V_{\text{xc}}[n(\mathbf{r},\omega)]}{\delta n\left(\mathbf{r}',\omega\right)} \right|_{n_0(\mathbf{r})} \tag{2.14}$$

and χ_{ks}, the Kohn-Sham response function is given by:

$$\chi_{\text{ks}}\left(\mathbf{r},\mathbf{r}',\omega\right) = \lim_{\eta\to 0^+} \sum_{j,k=1}^{\infty} \left(f_k - f_j\right) \frac{\phi_j(\mathbf{r})\phi_k^*(\mathbf{r})\phi_j^*\left(\mathbf{r}'\right)\phi_k\left(\mathbf{r}'\right)}{\omega - \left(\epsilon_j - \epsilon_k\right) + i\eta} \tag{2.15}$$

where f_j and f_k are occupation numbers of the Kohn-Sham orbitals (one means occupied and zero means empty), ϵ_j and ϵ_k are Kohn-Sham eigenvalues and η comes from the Lehmann representation of the response function [20]. From there, by rewriting the

density response into the matrix form and arguing mathematically, we can derive the Casida equation (full derivation given in [21]):

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} \mathbf{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{2.16}$$

where the elements of $\mathbf{A}$ and $\mathbf{B}$ are given by:

$$A_{ij\sigma,kl\tau}(\omega) = \delta_{ik}\delta_{jl}\delta_{\sigma\tau}\omega_{ji\sigma} + B_{ij\sigma,kl\tau}(\omega) \tag{2.17}$$

$$B_{ij\sigma,kl\tau}(\omega) = \int d^3r \int d^3r' \phi^*_{i\sigma}(\mathbf{r})\phi_{j\sigma}(\mathbf{r}) f_{\mathrm{Hxc}\sigma\tau}\left(\mathbf{r},\mathbf{r}',\omega\right) \phi_{k\tau}\left(\mathbf{r}'\right) \phi^*_{l\tau}\left(\mathbf{r}'\right) \tag{2.18}$$

where we define $f_{\mathrm{Hxc}}\left(\mathbf{r},\mathbf{r}',\omega\right) = \left|\mathbf{r}-\mathbf{r}'\right|^{-1} + f_{\mathrm{xc}}\left(\mathbf{r},\mathbf{r}',\omega\right)$. The occupied and unoccupied states are summed over by i,j and k,l respectively, Ω is the excitation frequency, $\mathbf{X}$ and $\mathbf{Y}$ are excitation and de-excitation, respectively. If the excitation frequencies are not close to zero (often the case for molecules and semiconductors), we may set the off-diagonal $\mathbf{B}$ matrix to zero, reducing the equation to:

$$\mathbf{AX} = \Omega\mathbf{X} \tag{2.19}$$

which is termed the Tamm-Dancoff approximation (TDA). [22] The advantage of TDA is that it can help to compensate for problems that arise due to the fact that the XC energy is approximated, in some situations (e.g. excitons [23] and triplet instabilities [24]) and give qualitatively correct results.

2.4 Colour of organic dyes

Colour is ubiquitous in our life and it allows us to visually perceive different objects in the world. Coloured substances, such as dyes and pigments, have been employed throughout human history in the areas of textile, food processing and cosmetics etc. The enormous usage of colours motivates us to understand their origin at a deeper level. The colours of objects are primarily determined by their optical absorption spectra, while it is also possible to form colours by structural interference with light. [25] The first encounter to the science of colours for us is probably the use of organic dyes, such as phenolphthalein, methyl orange and thymol blue as pH indicators during our high school chemistry class. While we know the colour change of the dyes arises because they have different structures at different pH, it is not particularly well understood how these changes in structural (composition and geometry etc.) and ambient conditions (temperature, acidity, polarity) affect the colours of organic dyes, in general. [26] We show an example in Figure 2.2. This figure illustrates the experimental absorption spectra of nile red in a range of solvents. We observe redshifts as well as changes in the line-shapes of the spectra when we change to different solvents, which have different polarities (blueshifts are also possible and it is difficult to predict which way a particular molecule will go [27]). Although cyclohexane and benzene have similar polarities, we observe different spectral line-shapes, suggesting that the explicit solvent-solute interactions are highly important factors in determining the correct absorption spectra.

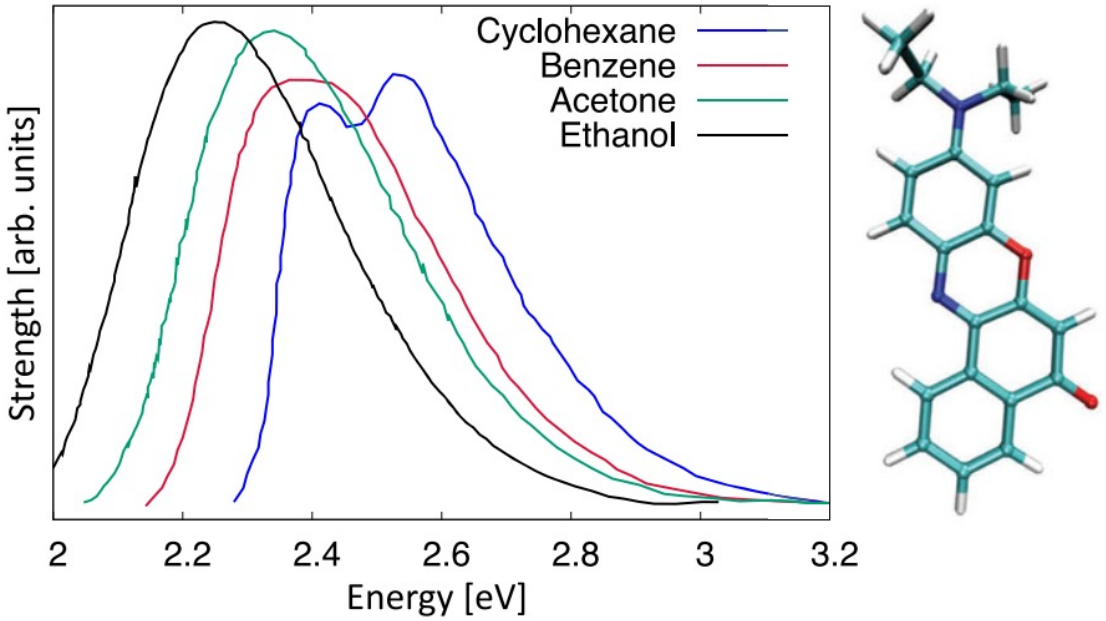


Figure 2.2: Experimental absorption spectra of nile red in different solvents. The static dielectric constants of the solvents are 2.0, 2.3, 20.7, 24.5 for cyclohexane, benzene, acetone, and ethanol, respectively. The observed solvatochromic redshift is found to be correlated with increasing solvent polarity. The experimental data in this plot were taken from Davis and Helzer [31]. This figure was adapted from ref. [28] with permission of AIP publishing

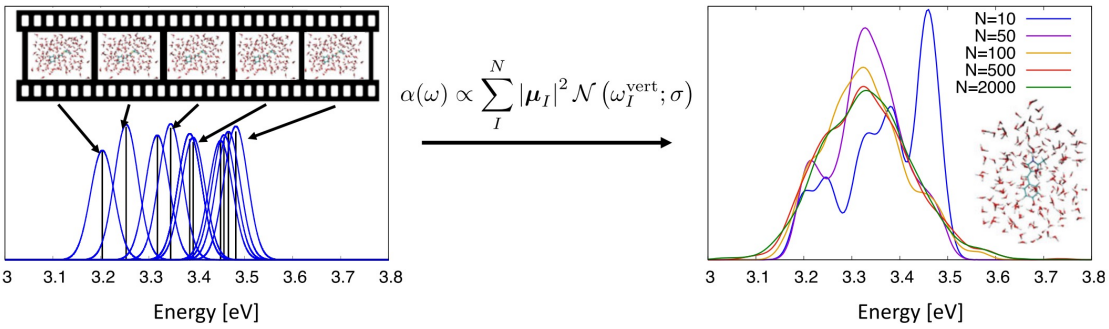


Figure 2.3: Computational spectrum of GFP chromophore anion in water. Figure on the left shows vertical excitation energies and the corresponding oscillator strengths calculated for ten solute-solvent configurations obtained from snapshots of a molecular dynamics trajectory, where each line is broadened by a Gaussian function. Figure on the right shows convergence of the spectrum with respect to the number of snapshots taken from the molecular dynamics simulation, where vertical excitation energies are convoluted with a Gaussian function $\mathcal{N}$ of width σ centered at ω_{vert}. Then, the sum of the Gaussian functions produces an ensemble spectrum, where they refer to this method as the ensemble approach. This figure was adapted from ref. [29] with permission of John Wiley & Sons

2.5 Modelling absorption spectra of molecules

2.5.1 Solvation models

Before we perform calculations on our molecules, a crucial aspect to consider is how to treat the surrounding region of the molecule because organic dyes are dissolved in solvents. One of the most simple ways is to treat the surrounding region as a vacuum, where the molecule is in gas phase, and we perform excited state calculations on a single optimised structure. If we want to do better than that, we can use implicit and explicit solvent models, which are discussed next:

Implicit solvation model

This model treats the solvent environment as a continuous medium, where the excited states are calculated by assuming a separation of timescales. The ground state of the solute is modelled through the static dielectric constant of the solvent, while the excited states are modelled through the optical dielectric constant of the solvent, by assuming that only the fastest electronic degrees of freedom can react to the perturbation caused by an excitation. [28, 29]

Explicit solvation model

The disadvantage of the implicit model is the lack of specific solvent-solute interactions. [29] Ideally, we want to treat all the solvent molecules explicitly at the quantum-mechanical level, but this may not be affordable in computational cost. Instead, we may use a combination of both models, where we treat a large part of the solvent near the solute by an explicit model and the remaining implicitly. Previous studies [28, 30] have shown that the explicit treatment of the solvent is vital to explain the solvatochromic shifts of nile red in n-hexane, toluene, acetone and ethanol, as well as N,N-diethyl-4-nitroaniline in water.

2.5.2 Spectral line-shapes

As mentioned earlier, one of the most simple and inexpensive approaches to compute absorption spectra is to perform excited state calculations on a single optimised structure, in gas phase (we can also do this in the presence of an implicit solvent). By convoluting the vertical excitation energies with a Gaussian function (see Figure 2.3) it is possible to obtain qualitatively correct spectral shapes. The position of the maximum absorption of these spectra is then compared directly to the experimental results, and obtaining the correct position of the maximum absorption is seen as a starting point towards accurate spectral calculations, which is also the approach taken by previous works [28, 32]. However, this method will be challenging if we want to use explicit solvent models. Firstly, the computational cost of optimisation increases rapidly as the system gets larger by including more explicit solvent molecules. Secondly, there is often no unique optimised geometry of such large systems. One solution to this issue is to sample solvent-solute configurations from snapshots of a molecular dynamics (MD) simulation and perform a calculation on each (this is indeed the case in reality because temperature effects mean that the molecule explores many different configurations). Then, we convolute the excitation energies from each configuration using the same technique and average the

spectrum over all snapshots. This is referred to as the ensemble approach.[29] However, the spectra calculated by this approach may be too narrow compared to experimental results due to the lack of vibronic effects, as found in previous works [33, 34]
Therefore, the next challenge is obtaining the correct spectral line-shape. There are mainly two factors that cause the broadening of the spectra—the inhomogeneous solvent broadening and the vibronic states broadening. The former is well accounted for by the sampling of solvent-solute configurations in the ensemble approach. To include the effect of vibronic states on the line-shape, it is necessary to treat the nuclei quantum-mechanically using the Franck-Condon principle. [35] However, this approach is not necessarily better than the ensemble approach because it can also produce narrow spectra due to the lack of sampling of solvent-solute configurations. [29] To fully account for the effects of solvent-solute interactions and vibrations, it may be necessary to use a combination of the two approaches. Different approaches to compute spectral line-shapes are illustrated in Figure 2.4.

2.5.3 Excited state methods

A range of different methods can be potentially used to calculate excited states, but we need to find a good balance between computational accuracy and cost for the method we choose. There are wavefunction-based methods, such as complete active space self-consistent field (CASSCF), [36] coupled cluster (CC) [37] and configuration interaction (CI) [38]. These methods are considered to be highly accurate but limited to small systems (tens of atoms). It would be unaffordable to use these methods for both the solvent and solute molecules. The less expensive methods include configuration interaction singles (CIS), [39] linear response time-dependent density-functional theory (LR-TDDFT) [2] and second-order approximate coupled-cluster (CC2). [40] With these methods, we may go up to thousands of atoms and it is possible to treat the solute molecules and a large region of the solvent at the same level of theory.

2.6 Spectral warping

Apart from considering appropriate models and excited state methods, previous works have also developed a special technique to reduce computational cost. Spectral warping is a numerical technique that transforms the absorption spectra calculated by GGA to the ones calculated by hybrid functionals via a linear shift in the excitation energy, with the aim to achieve hybrid functional accuracy with GGA functional cost. Ge *et al.* (2015) [32] first introduced this technique, which they termed 'colour morphing procedure'. They defined a linear transformation of the excitation energy as:

$$\omega^{\text{trans}} = \alpha\omega + \beta \tag{2.20}$$

where ω is the excitation energy calculated by GGA in a particular snapshot, α and β are fitting parameters which transform the peak positions in the GGA spectrum to match the peak positions in the hybrid spectrum, where the spectrum are calculated from a single snapshot of the molecule in gas phase. While Ge *et al.* found success in

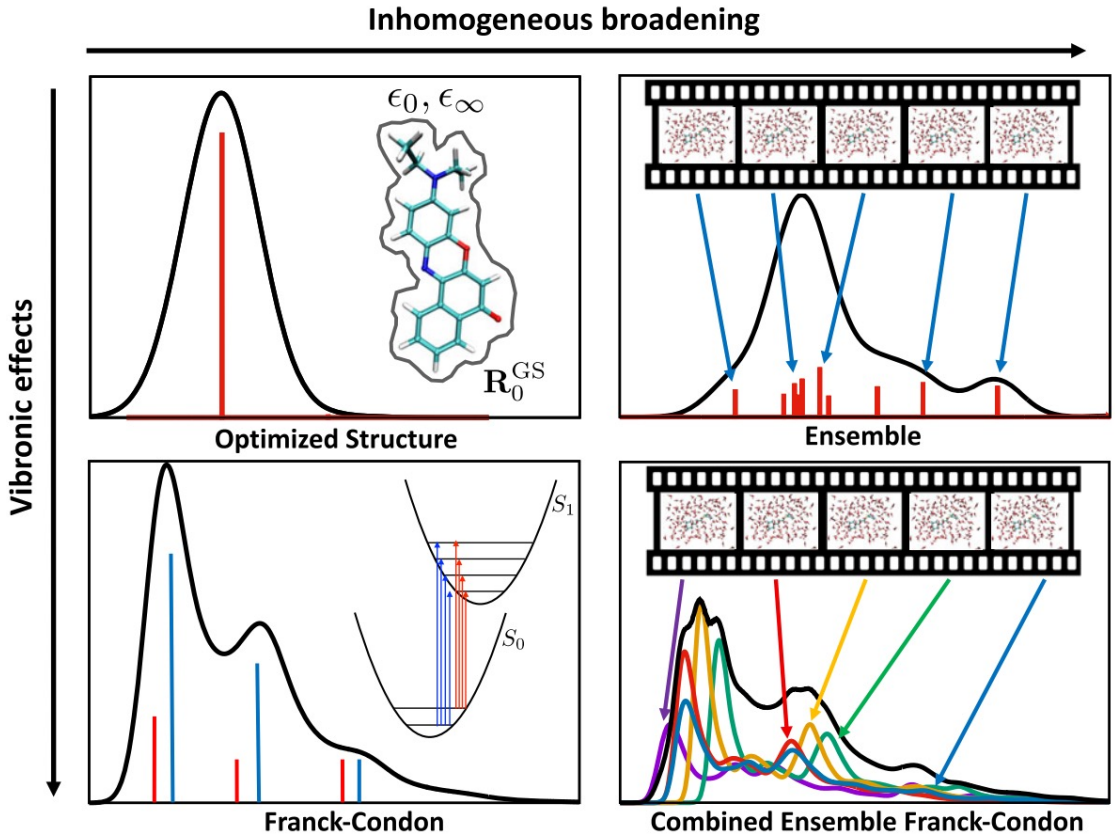


Figure 2.4: Different approaches to compute spectral line-shapes. From left to right, we include the inhomogeneous solvent broadening effect by sampling solvent-solute configurations; from top to bottom, we include the broadening effect of vibronic states by using the Franck-Condon principle. This figure was adapted from ref. [29] with permission of John Wiley & Sons

predicting the colours of anthocyanin using this method, Zuehlsdorff *et al.* (2017) [28] also obtained promising results for nile red, where they used a variant of this method:

$$\omega_i^{\text{trans}} = \omega_i + \delta + \gamma\phi_i^2 \tag{2.21}$$

where δ and γ are fitting parameters and ϕ_i is the average of certain dihedral angles of the molecule in snapshot i (they used index i because they carried out multiple gas phase calculations). If we look more closely, we find that these methods are equivalent if we set $\alpha = 1$ in Ge *et al.*'s definition, and therefore what Zuehlsdorff *et al.* have found, in essence, is that α is very close to 1 for every snapshot and β is linked inherently to the geometry of the system. Therefore, it is particularly useful to adapt the spectral warping method to make improvements for different systems.
However, a natural question arises when we apply spectral warping in practice, that is how do we choose our snapshots to perform both GGA and hybrid functional calculations on, in order to obtain accurate spectral warping functions? This is something that the previous works did not investigate in depth. Ge *et al.* assumed that the spectral warping shift does not change regardless of the solvent environment, and they therefore performed GGA and hybrid functionals calculations on a single snapshot of anthocyanin in gas phase and applied the gas phase spectral warping to the solvated systems. Zuehlsdorff *et al.* had the same idea, apart from the fact that they carried out calculations on a few snapshots of gas phase nile red, rather than one (the set of snapshots they chose was associated with a specific variation of a single dihedral angle of nile red). The problem of performing only one or two gas phase calculations is that, if we have a very flexible molecule, the fitting parameters can change significantly for different snapshots. However, we do not really know how many of these calculations are sufficient. Therefore, our objective is to develop a systematic approach to obtaining accurate spectral warping functions with the least computational effort.

2.7 Colour prediction

Colour prediction is not only useful for practical applications, but it can also serve as a test for the accuracy of the computational approach to optical absorptions in general. Spectral warping, as explained in previous discussions, only shifts the absorption peaks in the spectrum, whereas the line-shape usually remains unchanged (can change the line-shape when more than one important excitation is involved, in our case there is only one such excitation). However, colours are sensitive to both absorption peaks and the line-shape within the visible light range, so the lineshape of the spectrum has to also match those from the experiments to give the desired colours. The colour is predicted following the tristimulus colourimetry theory [32]:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = N \int I_0(\lambda) \mathrm{e}^{-\kappa(\lambda)x} \begin{pmatrix} \overline{x}(\lambda) \\ \overline{y}(\lambda) \\ \overline{z}(\lambda) \end{pmatrix} \mathrm{d}\lambda \tag{2.22}$$

where $X, Y$and Z are the three colour indices, N is a normalisation factor, $I_0(\lambda)$ is the illuminant, $\kappa(\lambda)$ is the absorption coefficient from TDDFT calculations, $\overline{x}(\lambda), \overline{y}(\lambda), \overline{z}(\lambda)$

are tristimulus colour-matching functions that describe the human retina response to red, blue and green colours. [32] The approach taken by previous works is to use the D65 illuminant [41], which represents standard daylight. After X, Y and Z are computed, they can be converted to RGB values to produce the colour.

3 Methods

3.1 Organic dyes

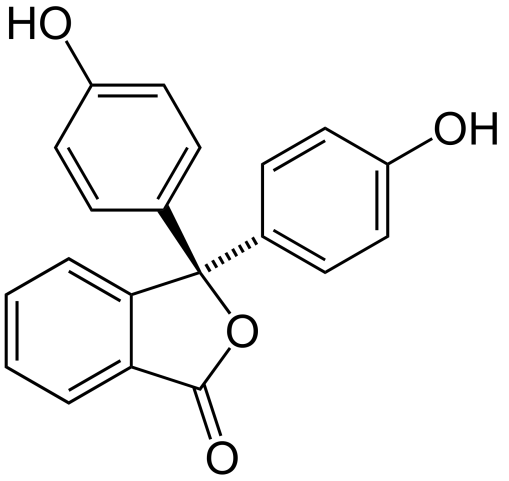
We examine two organic dyes in this work, namely phenolphthalein and thymol blue, which have several charge states with different colours. We investigate three of the states, which are neutral phenolphthalein, neutral thymol blue and phenolphthalein in the doubly negative charged state, shown in Figure 3.1. In neutral phenolphthalein, only the two phenol groups can rotate, whereas in the charged case, all three rings, as well as the CO2 groups are able to rotate. In thymol blue, rotations include those of the phenol rings and the side groups of three carbons.

3.2 Molecular dynamics

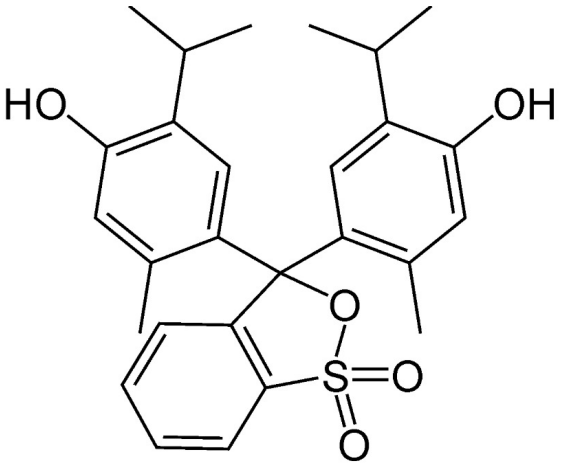
The initial structures of the molecules are taken from PubChem [44]. They are first relaxed using DFT, before being fed into AMBER [45] for molecular dynamics (MD) simulation. The antechamber tool within AMBER is used to generate a force field for the solute molecule and the standard TIP3P force field [46] is used to describe the water. The built-in tools in AMBER are used to solvate the molecule. After that, the MD starts. First, a structural relaxation is performed on the solvated system. Then, the system is heated from 0 to 300 K over 20 ps, before finally equilibrating the volume at 300 K for 400 ps. Once all of these steps have been performed, the production run begins. In all runs, the time step is 2 fs, whereas the usual time step for these simulations is 0.5 fs. A relatively large time step is used because we want to have a longer MD run with the same computational cost, in order to explore configurational spaces of our systems more efficiently. We achieve this by freezing the vibrations of bonds involving hydrogen using the SHAKE algorithm because they are the highest frequency vibrations in our system. We choose to freeze the hydrogen bonds because we do not expect the precise positions of the hydrogen atoms to significantly affect the electronic structure of our systems. The full treatment of vibronic effects is investigated in other studies, [33] and we will not worry about it here, although the low frequency vibrations of the entire system already manifest themselves by changing configurations in MD. Finally, snapshots are taken from the production run and used as the inputs for TDDFT calculations.

3.3 Time-dependent density-functional theory

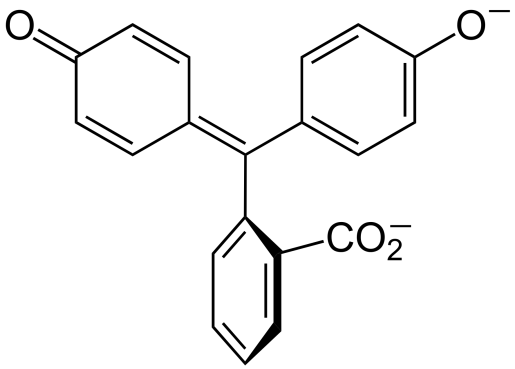
We use the NWChem code [47] for our TDDFT calculations. First, the built-in COSMO model is used to generate the implicit water solvent, where the static dielectric constant



(a) Skeletal formula of neutral phenolphthalein, where the solution is colourless



(b) Skeletal formula of neutral thymol blue, where the solution is yellow



(c) Skeletal formula of phenolphthalein in the doubly negative charged state, where the solution is pink. This state occurs between pH 8.2 and pH 10.0, where the molecule loses two protons

Figure 3.1: Skeletal formulae of organics dyes of interest. Images taken with permission from ref. [42, 43]

is 78.54. This is followed by a TDDFT calculation, where the Tamm-Dancoff approximation is used. [22] This approximation method is chosen because it is found to be the more computationally efficient one compared to the only other method supported by NWChem, namely the random phase approximation [48, 49, 50]), after trial tests of both methods. These methods yield very similar accuracies during the tests and this is also known to be true in general, as stated in the manual of NWChem. We perform around 4000 TDDFT calculations for each system in an implicit water solvent. We also carry out approximately 200 calculations for both phenolphthalein systems in an explicit water solvent including solvent molecules that are within 2.6 Å of the solute. For each model used, half of the calculations is carried out with PBE while the other half with B3LYP. The cc-pVDZ basis set is used for all calculations (this basis set has a suitable computational cost for the duration of our study).
For each calculation, ten possible singlet excitations, along with their oscillator strengths and frequencies, are extracted from the outputs (the triplet states are forbidden in these systems). We further define a selection criterion that only keeps the results of our primary interest, which is as follows: for systems that are known to have absorptions in the visible range (thymol blue and charged phenolphthalein), only examine the excitations within the visible range and pick the strongest one out of those because typically there is one excitation that is much stronger than the others.; for systems that do not have absorptions in the visible range (neutral phenolphthalein), look at the five lowest excitations (closest to visible range) and pick the strongest one out of those. By using this criterion, we can always pick out one excitation from each snapshot. We only care about the strongest one because our objective is to obtain the correct position of the maximum absorption, as explained in section 2.5.

3.4 Spectral warping

We define our spectral warping in a similar fashion to Zuehlsdorff *et al.*'s equation [28], shown in section 2.6:

$$\omega_i^{\text{trans}} = \omega_i + \beta_i \tag{3.1}$$

It is worth noting that our β here is snapshot dependent, which can be seen as the sum of δ and $\gamma\phi_i^2$ in Zuehlsdorff *et al.*'s equation. We tried Ge *et al.*'s equation initially, but it was found that for our results the α in Ge *et al.*'s equation is very close to 1 for all the systems that we investigate, which means that we do not need to multiply the excitation energy by a factor and a linear shift is sufficient. Moreover, the way we define our spectral warping implies that we use the same transformation for all the peaks in the same snapshot. It is possible to define an individual transformation for every peak in the spectrum, but this approach is not particularly relevant to the systems we are investigating because there is only one strong peak in the visible range.
We will obtain β from calculations using the implicit model, in contrast to previous works where β is obtained from gas phase calculations. We assume for a neutral system both methods give similar results because there are not many long-range Coulomb interactions and the presence of an implicit solvent should not affect the results. For a charged system

however, the use of implicit solvent is vital because it screens the long-range interactions, which can have a large effect on the excitation energies. [28]

4 Results and Discussion

Table 1 shows the excitation energies calculated using different combinations of models and functionals, as well as the spectral warped and experimental results. In general, the use of implicit models results in overestimation of the excitation energies, while PBE calculations yields lower energies than B3LYP ones. The failure of the implicit model may be due to the exclusion of the strong solvent-solute interactions caused by hydrogen bonding in water. The lower energies produced by PBE are commonly known as deficiencies of semi-local functionals. [12]
We also find that our spectral warping method works relatively well for neutral phenolphthalein because the warped excitation energy is similar to that found in the experiment, whereas for charged phenolphthalein, the difference between our calculation and experiment is much more significant. This result shows that the use of an explicit model can at least for some cases correctly account for the specific solvent-solute interactions and the polarisation effects of solute in the ground state. [28] For charged phenolphthalein, we do not have a definite explanation of the discrepancy, but one potential reason could be the basis set we are using (cc-pVDZ) does not include diffuse functions. It is generally assumed that DFT calculations involving anions require basis sets with diffuse functions in order to obtain reliable results. However, including diffuse functions for large systems such as the ones we are considering would increase the computational cost as well as introducing self-consistent field convergence problems. [51]

Table 1: Calculated average excitation energies compared to experimental values for neutral phenolphthalein (PhPh0), phenolphthalein -2 (PhPh-2) and neutral thymol blue (ThBl0), all in eV. The experimental values are extracted from ref. [52, 53]

	Implicit PBE	Implicit B3LYP	Explicit PBE	Explicit Warped	Experiment
PhPh0	3.82	4.65	3.61	4.45	4.49
PhPh-2	2.31	2.41	2.33	2.43	2.21
ThBl0	2.47	2.68	-	-	2.26

4.1 Correlation between PBE and B3LYP energies

Figure 4.1 shows the difference between the B3LYP energies of individual snapshots and their average against that difference of the PBE energies. It is evident that the B3LYP and PBE energies are strongly correlated, despite a significant number of outliers. We notice the horizontal and vertical axes are very comparable in scale, which means the slope of a fit line (corresponding to the α parameter in the spectral warping function) will be approximately 1 and the PBE and B3LYP energies only differ by a constant (the β parameter). The physical meaning of this constant is thought of as the difference

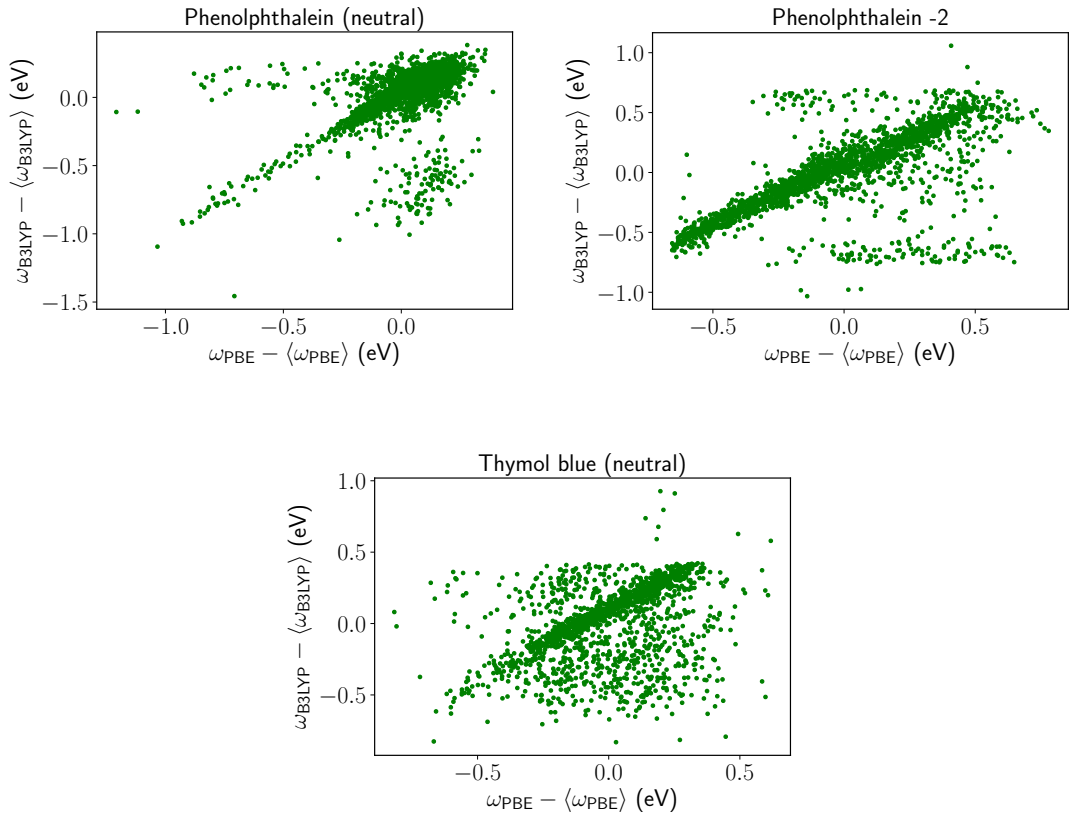


Figure 4.1: Correlation between B3LYP and PBE energies, where triangular bracket indicates the average of 2000 calculations. Axes show differences between energies of one snapshot and the average of 2000 snapshots because we want to see how much each energy deviates from the average

in accuracy between the HOMO-LUMO transitions of the dyes calculated by PBE and B3LYP functionals.

4.2 Correlation between PBE energies and spectral warping

We now consider the quantity of our interest—the spectral warping shift. It can be readily calculated by manipulating Figure 4.1, where we subtract the values on the horizontal axis from the values on the vertical axis, shown in Figure 4.2. We observe an elliptical band of snapshots centred at the zero-point of the vertical axis, which implies there is no correlation between the spectral warping shifts and PBE energies. This is somehow within our expectation because the PBE and B3LYP energies only differ by a constant and the difference between them (the shift) does not depend on the energy itself. Hence, we see the flat ellipse. We plot our results this way because we want to check whether they are correlated. If this is true, it is not necessary to perform the expensive B3LYP calculations for all 2000 snapshots because we can carry out an initial screening of the snapshots and pick out the 'good' ones which can produce accurate spectral warping functions.
We also observe that a small number of snapshots form a downward slope which roughly equals -1 and these snapshots correspond to the flat band in Figure 4.1. These snapshots produce comparable B3LYP energies but highly different PBE energies and are considered outliers. We have not established what links to these outliers.

4.3 Density of snapshots

Figure 4.3 contains the same data as Figure 4.2, but as a histogram, which makes analysis easier. It is evident from Figure 4.3 that, for all three systems, the majority of the snapshots produce spectral warping shifts which are very close to the average value of the 2000 snapshots because we observe peaks at the zero-point of the horizontal axis. In other words, if we merely perform random sampling on our snapshots and average over these random samples we would obtain relatively good spectral warping shifts. We also fit a Gaussian distribution to our results because by doing so, we can make an estimate of the number of snapshots that we need to randomly sample, in order to achieve a given accurary of the spectral warping shift. The fitted distributions have similar standard deviations, which may be due to the fact that the systems we are considering have similar physical sizes and structures. We also notice that these distributions have tails and some are more significant than others, which do not fit a single Gaussian model. Therefore, these fits can only give a ballpark estimate, for instance, for neutral phenolphthalein the standard deviation is 0.0679, that means we require around 50 snapshots to achieve a result within 0.01 eV ($(0.0679/0.01)^2 \approx 50$).
To demonstrate how well the random sampling method works, we have plotted the error of the spectral warping shift as a function of the number of snapshots sampled, in Figure 4.4. We observe that the absolute error decays rapidly with increasing sample size and we use absolute error here because we do not really care if the errors are negative or positive. In reality, the magnitude of the errors would be smaller because the positive

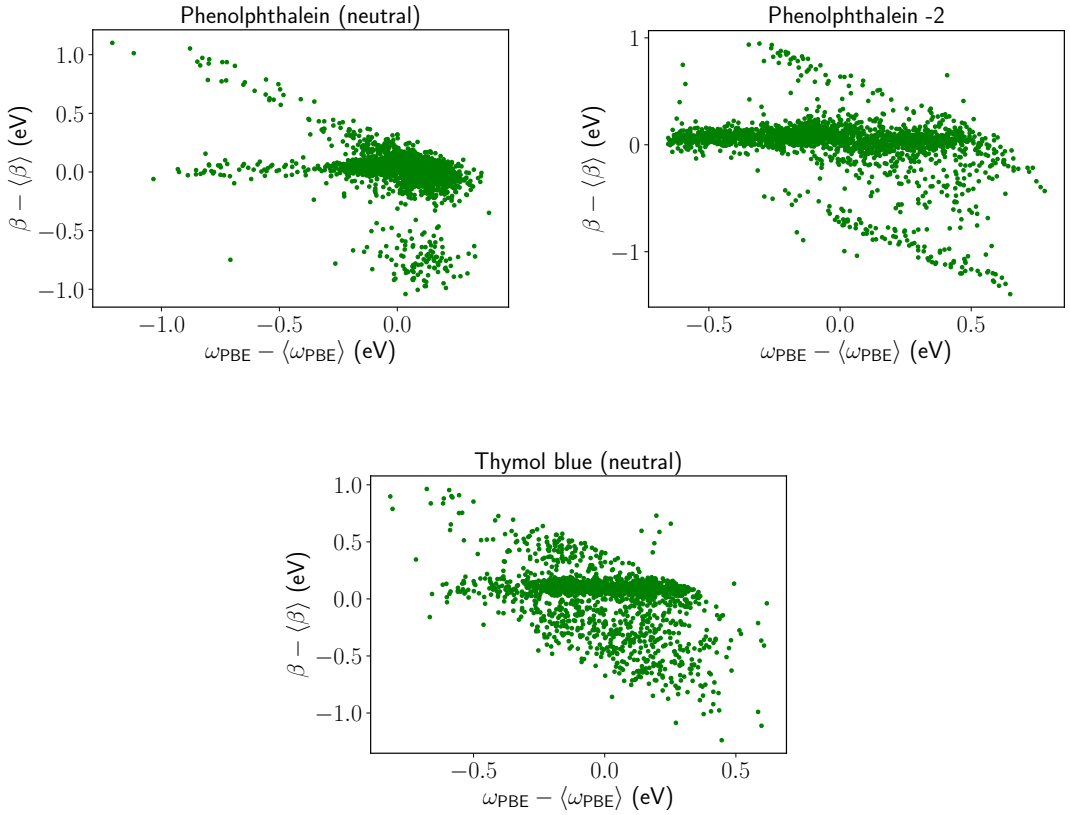


Figure 4.2: Spectral warping shifts are plotted against PBE energies, where the triangular bracket indicates the average of 2000 calculations.

and negative errors cancel out and the number of snapshots required would also be smaller. We also notice the errors of the phenolphthalein -2 are significantly larger than others and our reasoning for this is the same as our explanation for Table 1, i.e. due to not including diffuse functions in our basis set.

4.4 Correlation between dihedral angles and spectral warping

Figure 4.5 plots the spectral warping of individual snapshots against the variation of a single dihedral angle in that snapshot, for 2000 calculations, where the dihedral angles are defined by numbered atoms shown in the structures of the molecules. The correlation between the dihedral angles and spectral warping has a highly similar structure to that between PBE energies and spectral warping, namely a flat band. This result suggests that the dihedral angle does not affect spectral warping in our systems, whereas Zuehlsdorff *et al.* found that the spectral warping shifts vary as the dihedral angle squared for nile red. [28] It is worth noting the fact that nile red has only one dihedral angle, while our systems have multiple dihedral angles and Figure 4.5 only shows two typical plots (the trend is similar for other dihedral angles). The failure in finding a correlation between spectral warping and individual dihedral angles may be due to the fact that the bond rotations in our systems are coupled and it is necessary to view the rotations as a whole, rather than looking at changes in individual dihedral angles. Accounting for the effects of coupled rotations of bonds on spectral warping is a good avenue for future work.

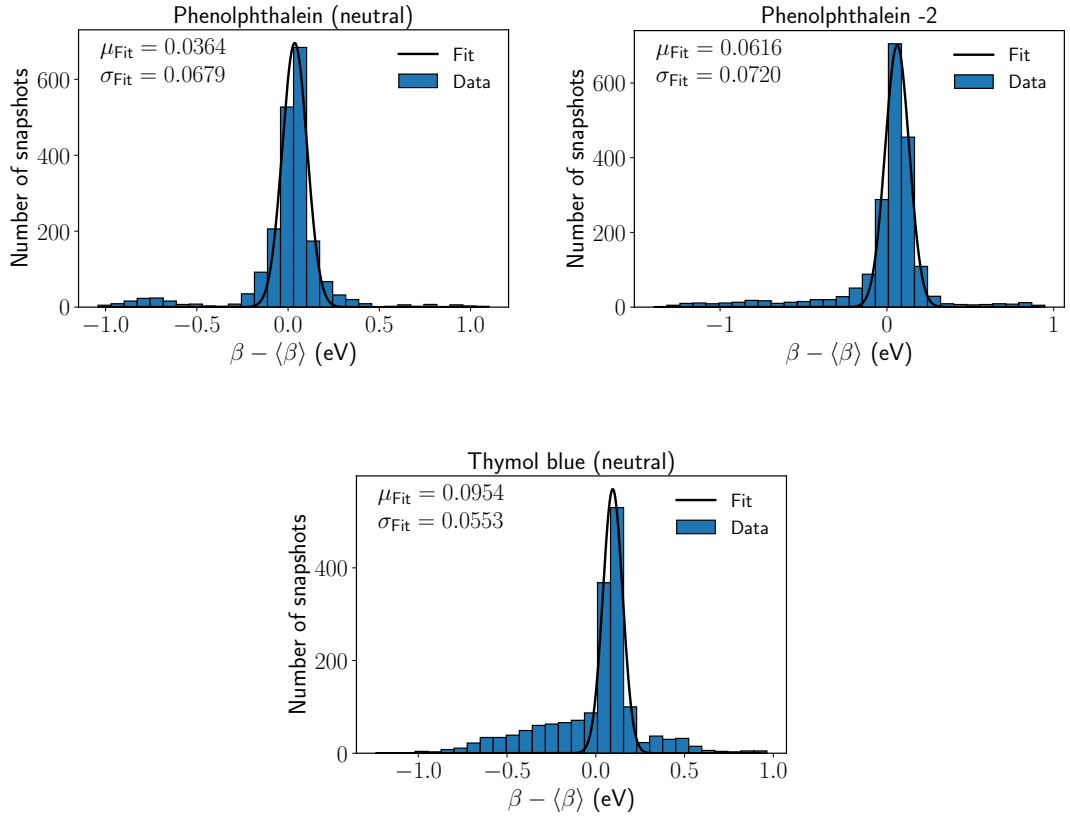


Figure 4.3: Number of snapshots are plotted against spectral warping shifts, with a bin width of 0.048 eV. Gaussian distributions are fitted to each system

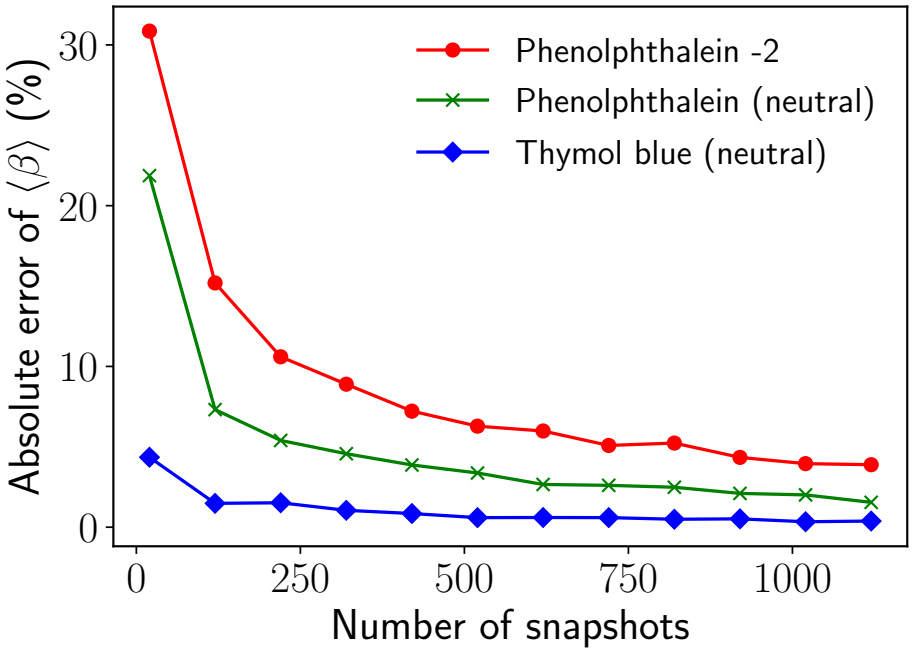
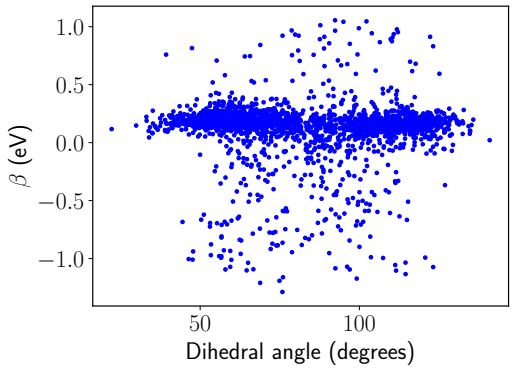
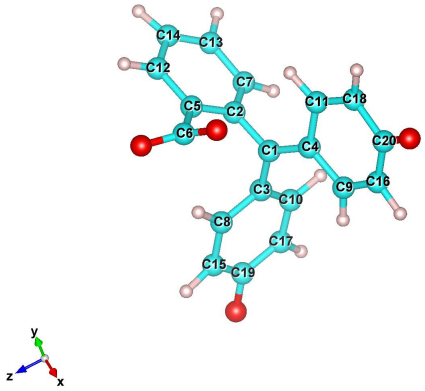


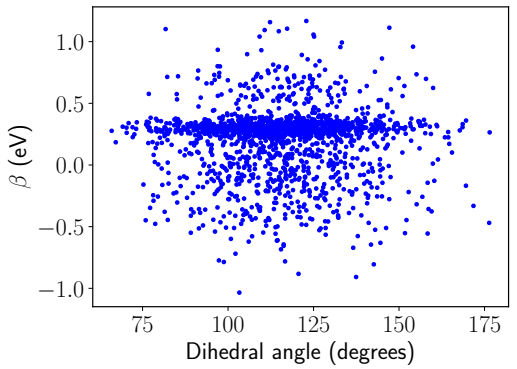
Figure 4.4: Absolute percentage error of spectral warping shift for a given number of randomly sampled snapshots compared to the average shift of all snapshots



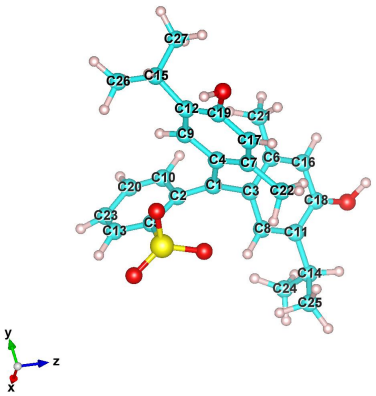
(a) Dihedral angle defined by C3-C1-C2-C5 of phenolphthalein -2



(b) Structure of phenolphthalein -2



(c) Dihedral angle defined by C9-C12-C15-C26 of neutral thymol blue



(d) Structure of neutral thymol blue

Figure 4.5: Spectral warping shifts plotted as a function of dihedral angles. Structures are visualised by VESTA [54]

5 Conclusions

In this study, we have calculated the excited states of neutral phenolphthalein, neutral thymol blue and charged phenolphthalein in an implicit and explicit solvent using PBE and B3LYP functionals and investigated closely the accuracy of the spectral warping method, a numerical technique used in excited state calculations in order to reduce computational costs. While our application of the spectral warping method gives results in good agreement with experiments for neutral phenolphthalein, it fails for the case of charged phenolphthalein. It is also shown that the spectral warping shifts are uncorrelated to the excitation energies calculated by the PBE functional and also uncorrelated to the variation of individual dihedral angles of the system. One useful way to extract the average spectral warping shift is to perform calculations on randomly sampled solvent-solute configurations from snapshots of a molecular dynamics simulation for the system of interest. We find that around 50-100 snapshots are appropriate to achieve an accuracy of 0.01 eV. For future work, we could investigate the effects of coupled bond rotations on the spectral warping functions.

References

[1] F. Giustino. *Materials modelling using density functional theory : properties and predictions*. Oxford University Press, Oxford, 2014.

[2] C. Ullrich and Z. Yang. A brief compendium of time-dependent density functional theory. *Brazilian Journal of Physics*, 44(1):154–188, 2014.

[3] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, 1964.

[4] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136(15):150901, 2012.

[5] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[6] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.

[7] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, May 1981.

[8] Lee A. Cole and J. P. Perdew. Calculated electron affinities of the elements. *Phys. Rev. A*, 25:1265–1271, Mar 1982.

[9] J. D. Patterson. Density-functional theory of atoms and molecules: Robert g. parr and weitao yang . oxford university press, new york, and clarendon press, oxford. *Annals of Nuclear Energy*, 16(11):611–611, 1989.

[10] J. P. Perdew, K. Burke, and M Ernzerhof. Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]. *Phys. Rev. Lett.*, 78(7):1396–1396, 1997.

[11] Timothy J. Giese and Darrin M. York. Density-functional expansion methods: Evaluation of lda, gga, and meta-gga functionals and different integral approximations. *The Journal of Chemical Physics*, 133(24):244107, 2010.

[12] M. H. N. Assadi and D. A. H. Hanaor. Theoretical study on copper's energetics and magnetism in tio 2 polymorphs. *J. Appl. Phys*, 113(23), 2013.

[13] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem*, 98(45):11623–11627, 1994.

[14] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988.

[15] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988.

[16] E. Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997–1000, Mar 1984.

[17] C. A. Ullrich and I. V. Tokatly. Nonadiabatic electron dynamics in time-dependent density-functional theory. *Phys. Rev. B*, 73:235102, Jun 2006.

[18] M. Thiele, E. K. U. Gross, and S. Kümmel. Adiabatic approximation in nonperturbative time-dependent density-functional theory. *Phys. Rev. Lett.*, 100:153004, Apr 2008.

[19] M. Thiele and S. Kümmel. Hydrodynamic perspective on memory in time-dependent density-functional theory. *Phys. Rev. A*, 79:052503, May 2009.

[20] H. Lehmann. Über eigenschaften von ausbreitungsfunktionen und renormierungskonstanten quantisierter felder. *Il Nuovo Cimento*, 11(4):342–357, Apr 1954.

[21] M. E. Casida. *Recent Advances in Density Functional Methods*. WORLD SCIENTIFIC, 1995.

[22] S. Hirata and M. Head-Gordon. Time-dependent density functional theory within the tamm–dancoff approximation. *Chem. Phys. Lett.*, 314(3):291 – 299, 1999.

[23] Zeng-hui Yang, Yonghui Li, and Carsten A. Ullrich. A minimal model for excitons within time-dependent density-functional theory. *The Journal of Chemical Physics*, 137(1):014513, 2012.

[24] Mark E. Casida, Fabien Gutierrez, Jingang Guan, Florent-Xavier Gadea, Dennis Salahub, and Jean-Pierre Daudey. Charge-transfer correction for improved time-dependent local density approximation excited-state potential energy curves: Analysis within the two-level model with illustration for h2 and lih. *The Journal of Chemical Physics*, 113(17):7062–7071, 2000.

[25] Wikipedia contributors. Structural coloration — Wikipedia, the free encyclopedia, 2019. [Online; accessed 8-April-2019].

[26] O. B. Malcoğlu, A. Calzolari, R. Gebauer, D. Varsano, and S. Baroni. Dielectric and thermal effects on the optical properties of natural dyes: A case study on solvated cyanin. *J. Am. Chem. Soc.*, 133(39):15425–15433, 2011. PMID: 21905678.

[27] T. J. Zuehlsdorff, P. D. Haynes, F. Hanke, M. C. Payne, and N. D. M. Hine. Solvent effects on electronic excitations of an organic chromophore. *J. Chem. Theory Comput.*, 12(4):1853–1861, 2016. PMID: 26967019.

[28] T. J. Zuehlsdorff, P. D. Haynes, M. C. Payne, and N. D. M. Hine. Predicting solvatochromic shifts and colours of a solvated organic dye: The example of nile red. *J. Chem. Phys.*, 146(12):124504, 2017.

[29] T. J. Zuehlsdorff and C. M. Isborn. Modeling absorption spectra of molecules in solution. *Int. J. Quantum Chem.*, 119(1):e25719, 2019.

[30] Andrzej Eilmes. Solvatochromic probe in molecular solvents: implicit versus explicit solvent model. *Theor. Chem. Acc.*, 133(9):1538, Jul 2014.

[31] M. M. Davis and H. B. Helzer. Titrimetric and equilibrium studies using indicators related to nile blue a. *Anal. Chem.*, 38(3):451–461, 1966.

[32] X. Ge, L. Timrov, S. Binnie, A. Biancardi, A. Calzolari, and S. Baroni. Accurate and inexpensive prediction of the color optical properties of anthocyanins in solution. *J. Phys. Chem. A*, 119(16), 2015.

[33] T. J. Zuehlsdorff and C. M. Isborn. Combining the ensemble and franck-condon approaches for calculating spectral shapes of molecules in solution. *J. Chem. Phys.*, 148(2):024110, 2018.

[34] C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker, and T. J. Martínez. Electronic absorption spectra from mm and ab initio qm/mm molecular dynamics: Environmental effects on the absorption spectrum of photoactive yellow protein. *J. Chem. Theory Comput.*, 8(12):5092–5106, 2012. PMID: 23476156.

[35] J. Franck and E. G. Dymond. Elementary processes of photochemical reactions. *Trans. Faraday Soc.*, 21:536–542, 1926.

[36] F. Jensen. *Introduction to Computational Chemistry.* Chichester, England: John Wiley and Sons, 2007.

[37] I. Shavitt and R. J. Bartlett. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory.* Cambridge University Press, 2002.

[38] C. D. Sherrill and H. F. Schaefer. The configuration interaction method: Advances in highly correlated approaches. volume 34 of *Advances in Quantum Chemistry*, pages 143 – 269. Academic Press, 1999.

[39] J. A. Pople J. B. Foresman, M. Head-Gordon and M. J. Frisch. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem*, 1992.

[40] O. Christiansen, H. Koch, and P. Jørgensen. The second-order approximate coupled cluster singles and doubles model cc2. *Chem. Phys. Lett.*, 243(5):409 – 418, 1995.

[41] *CIE Standard Colorimetric System*, chapter 3, pages 63–114. John Wiley Sons, Ltd, 2006.

[42] Wikipedia contributors. Phenolphthalein — Wikipedia, the free encyclopedia, 2019. [Online; accessed 9-April-2019].

[43] Wikipedia contributors. Thymol blue — Wikipedia, the free encyclopedia, 2018. [Online; accessed 9-April-2019].

[44] Kim S., Chen J., Cheng T., Gindulyte A., He J., He S., Li Q., Shoemaker B. A., Thiessen P. A., Yu B., Zaslavsky L., Zhang J., and Bolton E. E. Pubchem 2019. *Nucleic Acids Res. 47(D1):D1102-1109*, 2019.

[45] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, V.W.D. Cruzeiro III, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, J. Liu C. Lin, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, and P.A. Kollman. Amber 2018. *University of California, San Francisco*, 2018.

[46] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[47] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong. Nwchem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun. 181, 1477*, 2010.

[48] D. Pines and D. Bohm. A collective description of electron interactions. i. magnetic interactions. *Phys. Rev.*, 82:625–634, Jun 1951.

[49] D. Pines and D. Bohm. A collective description of electron interactions: Ii. collective vs individual particle aspects of the interactions. *Phys. Rev.*, 85:338–353, Jan 1952.

[50] D. Pines and D. Bohm. A collective description of electron interactions: Iii. coulomb interactions in a degenerate electron gas. *Phys. Rev.*, 92:609–625, Nov 1953.

[51] A. Bauzá, D. Quiñonero, P. M. Deyà, and A. Frontera. Is the use of diffuse functions essential for the properly description of noncovalent interactions involving anions? *J. Phys. Chem. A*, 117(12):2651–2655, 2013. PMID: 23465073.

[52] W. R. Orndorff, R. C. Gibbs, and S. Alice McNulty. The absorption spectra of phenolphthalein, isophenolphthalein and of diphenylphthalide1,2. *J. Am. Chem. Soc.*, 48(7):1994–2002, 1926.

[53] T. Fujii and K. Toriumi. Ultraviolet–visible absorption spectra of thymol blue during the sol–gel–xerogel transition of tetraethyl orthosilicate: in situ probe of microchemical environment. *J. Chem. Soc. Faraday Trans.*, 89:3437–3441, 1993.

[54] K. Momma and F. Izumi. Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr., 44, 1272-1276*, 2011.