Development of Density Functional Methods for Electronic Excited States and the Influence of Molecular Structure on Electronic Excited States

by

Peter Andrew Thomas John Fletcher
31448319

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemistry

Lancaster University, Department of Chemistry

January 2021

Abstract

An extensive assessment of six density functional approximations has been undertaken, each of these approximations have their own merits and faults. Range separated hybrids are the best performing for excited state properties of those approximations assessed.

There has been an attempt to generate an attenuated form of PBE (CAM-PBE) which initially had issues which were investigated in detail regarding the dependence of Hartree–Fock exchange energy on approximation performance. This attenuated form of PBE had similar performance to CAM-B3LYP.

The development of a set of benchmark data for excited state geometries and emission energies was undertaken with a wide range of organic molecules due to the lack of such benchmark data existing currently. This means the accuracy of density functional approximations for calculation of such properties is unknown so there is a clear need for this benchmark data to be developed and used to assess the accuracy of these approximations.

The benchmark data for excited state geometries and emission energies was used to assess the performance of a range of density functional approximations for these properties. This assessment has suggested that there are issues when applying current density functional approximations away from the ground state where they have been tuned and optimised. This suggests that there may be some merit in developing specialised density functional approximations for the calculation of excited state properties.

The existing density functional approximations have been used to assist with experimental investigations of porous polymers and in explaining the excited state properties of these polymers. This was done using model systems and has enabled a deeper understanding of the experimental observations.

Acknowledgement

I would like to thank my supervisor – Dr. Michael Peach for his invaluable advice, support and patience during the course of my PhD study. My gratitude extends to the Faculty of Science and Technology for the funding opportunity to undertake my studies at the Department of Chemistry, Lancaster University. I would like to thank my friends, lab mates, colleagues for a fantastically supported time through my PhD study.

Contents

List of Tables

List of Figures

1 Theory and Background

1.1 Quantum Mechanics

Classical mechanics, also known as Newtonian mechanics, is a series of physical concepts, originated by Newton, used to describe the motion of bodies under the influence of a system of forces. Classical mechanics allows us to predict how an object will move in the future or how it moved in the past provided the present state of the object is known - the systems it describes are deterministic. It is accurate when used on large objects (but not at a planetary scale) and speeds which do not approach the speed of light. Classical mechanics can not be accurately applied to microscopic particles, such as electrons, moving at high speed. Quantum mechanics arose to describe the behaviour of such particles and is probabilistic in nature.

1.1.1 Schrödinger Equation

The Schrödinger equation[3], in atomic units, (equation 1) is a partial differential wave equation that describes how a quantum state of a physical system evolves with time. The wave function (Ψ) is the description of the quantum state of a system of one or more quantum particles. The wave function is a function of the degrees of freedom of a chosen set of observables, for example this could be the position coordinates of the particles over position space.

The time-dependent Schrödinger equation is

$$\hat{H}\Psi = \mathrm{i}\frac{\partial}{\partial t}\Psi \tag{1}$$

In quantum chemistry, the time dependence is usually separated and thus the time-independent Schrödinger equation (equation 2) and the time-independent energy and wave function is considered.

$$\hat{H}\Psi = E\Psi \tag{2}$$

The Hamiltonian operator ($\hat{H}$) characterises the total energy, E, of any given wave function. In the case of molecules with n electrons and N nuclei (equation 3) (capitals for nuclear indices and lower case for electron indices), the Hamiltonian contains the nuclear kinetic energy, electronic kinetic energy, electron–nuclei Coulomb attractions, electron–electron Coulomb repulsions and nuclear–nuclear Coulomb repulsions.

$$\hat{H} = \sum_{I=1}^{N} -\frac{\nabla_I^2}{2m_I} - \sum_{i=1}^{n} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{J=1}^{N} \frac{q_J}{r_{iJ}} + \sum_{i<j}^{n} \frac{1}{r_{ij}} + \sum_{I<J}^{N} \frac{q_I q_J}{r_{IJ}} \tag{3}$$

where m_I are the masses of the nuclei, q_I are the charges of the nuclei and r_{ij} is the distance between particles i and j ($r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$). The operator ∇_I is a notation of the first derivative of the position of particle I over X, Y and Z coordinates (which is the velocity of the particles) ($\nabla_I = \left(\frac{\partial}{\partial X} + \frac{\partial}{\partial Y} + \frac{\partial}{\partial Z}\right)$).

If the Born–Oppenheimer approximation is applied we decouple the motion of the nuclei and electrons and the electrons move in the field of the nuclei. This is typically a good approximation because of the large mass difference between electrons and nuclei, this is an important approximation as it makes the equations more solvable. This leads us to the electronic Schrödinger equation (equation 4), where $\hat{H}^0$ (equation 5) is the Hamiltonian we obtain, excluding the kinetic energy of the nuclei, and adding a constant for the internuclear repulsion (V_N) the final term from equation 3 becomes a constant once the nuclei are fixed.

$$E^0\Psi^0 = \hat{H}^0\Psi^0 \tag{4}$$

$$\hat{H}^0 = -\sum_{i=1}^{n} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{J=1}^{N} \frac{q_J}{r_{iJ}} + \sum_{i<j}^{n} \frac{1}{r_{ij}} + V_N \tag{5}$$

The electronic Hamiltonian can thus be specified for an arbitrary molecular system, but the wave function can only be found in simple closed form for one electron systems or simple model systems.

1.2 Hartree–Fock Theory

1.2.1 Hartree Product Wave Function

The electronic Schrödinger equation cannot be solved exactly for real molecular systems due to the electron–electron interaction and the resulting high dimensionality. One approximation that can be made is that electron–electron repulsion is small. For such a case, the electronic Hamiltonian is a sum of one-electron operators[4]

$$\begin{aligned}\hat{H}^0 &= -\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{J=1}^{N}\frac{q_J}{r_{iJ}} + V_N \\ &= \sum_{i=1}^{n}\hat{h}_i + V_N \\ h_i &= -\frac{1}{2}\nabla_i^2 - \sum_{J=1}^{N}\frac{q_J}{r_{iJ}}\end{aligned} \tag{6}$$

and then the wave function is a simple product of one-electron terms (equation 7) - the one electron orbitals. Such a wave function is called a Hartree product wave function,

$$\psi_i = \psi_1\psi_2\cdots\psi_n. \tag{7}$$

Upon substitution of the Hartree product wave function into equation 6, it is found that the n one-electron orbitals can be obtained by solving n one-electron Schrödinger equations (equation 8), and the total energy is then simply the sum of the one-electron energies ϵ_i (equation 9).

$$h_i\psi_i = \epsilon_i\psi_i \tag{8}$$

$$E = \sum_{i=1}^{n}\epsilon_i \tag{9}$$

However, neglecting electron repulsion leads to large errors. Substitution of the Hartree product wave function into the full electronic Hamiltonian $\hat{H}^0$(equation 4) leads to equation 10. As with the one-electron Hamiltonian, each of the n one-electron orbitals needed can be found by solving a one-electron Schrödinger equation, with the electron interactions included through a mean-field potential (equation 11).

$$\left[-\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{J=1}^{N}\frac{q_J}{r_{iJ}} + V_N + \sum_{i<j}\frac{1}{r_{ij}}\right](\psi_1\cdots\psi_n) = E(\psi_1\cdots\psi_n) \tag{10}$$

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{J=1}^{N}\frac{q_J}{r_{iJ}} + V_i^{\text{eff}}(\mathbf{r}) \tag{11}$$

The effective potential for electron i, $V_i^{\text{eff}}(\mathbf{r})$, is the Coulomb interaction with the average charge density of all the other electrons. However, the implications of electron spin are not accounted for, therefore the Hartree product is a poor approximation for the true electronic wave function.

1.2.2 Slater Determinants

The Hartree product wave function violates the Pauli exclusion principle (that all electrons must have unique quantum numbers), as there is nothing to prevent the one-electron orbitals being identical so multiple electrons could occupy the same state. It also violates the Pauli principle (from which the exclusion principle is derived), as the total wave function is not antisymmetric with interchange of any pair of electrons. In order to bypass this behaviour spin functions, α or β, should be included in the trial Hartree product wave function. Wave functions generated this way are generally written as Slater determinants.[5,6] The spin and spatial parts of the orbital can be combined to make a spin-orbital, $\phi(\mathbf{r}, \sigma)$. The general form for a Slater determinant for n electrons using spin orbitals is

$$\psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_n(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(n) & \phi_2(n) & \dots & \phi_n(n) \end{vmatrix} \tag{12}$$

where the prefactor ensures normalisation.

1.2.3 Hartree–Fock Energy

Evaluating the expectation value of the electronic Hamiltonian (equation 5 using the Slater determinant gives the Hartree-Fock energy[7] (the energy associated with a Slater determinant) as equation 13

$$\begin{aligned} E_{\text{HF}} &= \sum_{i=1}^{n} \langle \phi_i | h_i | \phi_i \rangle + \frac{1}{2} \sum_{i,j}^{n} ([\phi_i \phi_i | \phi_j \phi_j] - [\phi_i \phi_j | \phi_j \phi_i]) \\ J &= [\phi_i \phi_i | \phi_j \phi_j] = \int \frac{|\phi_i(1)|^2 |\phi_j(2)|^2}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \mathrm{d}\sigma_1 \mathrm{d}\sigma_2 \\ K &= [\phi_i \phi_j | \phi_j \phi_i] = \int \frac{\phi_i^*(1) \phi_j(1) \phi_j^*(2) \phi_i(2)}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \mathrm{d}\sigma_1 \mathrm{d}\sigma_2 \end{aligned} \tag{13}$$

The Coloumb integral, J, describes the interactions between the electrons and the exchange integral, K, describes the quantum mechanical exchange.

This energy equation needs to be minimised in an iterative process, using the variational principle, it holds that the energy of the trial orbitals is never less than the true energy. In practice the molecular orbitals that are used to construct are varied in order to minimise the energy. Invoking an iterative procedure known as the self–consistent field method solves the problem (figure 1).

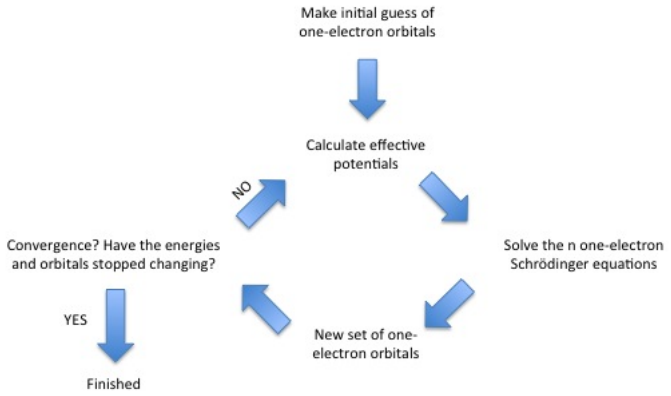


Figure 1: A Schematic of the Self–Consistent Field Method.

The variational minimisation of equation 13 leads to the Hartree–Fock equation, equation 14. The Fock operator, $\hat{f}$, is defined in equation 15.

$$\hat{f}_i \phi_i = \epsilon_i \phi_i \tag{14}$$

$$\hat{f}_1 = \hat{h}_1 + \sum_{j=1}^{n} [\hat{J}_j(1) - \hat{K}_j(1)] \tag{15}$$

The Fock operator for each electron resembles equation 11. There is a one-electron part, $\hat{h}_1$, and an effective potential for the electron, $[\hat{J}_j(1) - \hat{K}_j(1)]$. The one-electron part describes the kinetic energy of the electron and its attraction to the nuclei. The effective potential describes the Coulomb interactions with all of the other electrons via *the Coulomb Operator*, $\hat{J}_j$ and the quantum mechanical exchange via *the exchange operator*, $\hat{K}_j$. The form of the operators in the effective potential mean that each electron only experiences the average interaction of all other electrons.

1.2.4 Quantum–Mechanical Exchange

Accounting for the Pauli principle naturally introduces a quantum-mechanical exchange contribution to the energy. This can be seen when comparing the potential

energy of two interacting electrons of different spin (singlet) and two interacting electrons of the same spin (triplet). The potential energy for paired spin V_s (equation 16) and for parallel spin V_t (equation 17) is

$$V_{\mathrm{s}} = \int \frac{|\psi_1(1)|^2 |\psi_2(2)|^2}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{16}$$

$$V_{\mathrm{t}} = \int \frac{|\psi_1(1)|^2 |\psi_2(2)|^2}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 - \int \frac{\psi_1^*(1)\psi_2(1)\psi_1^*(2)\psi_2(2)}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{17}$$

The V_{s} integral is just the Coulomb interaction between electron densities in one electron orbitals 1 and 2. This integral is a *Coulomb integral*. In the V_{t} case there is an additional term which arises due to the electrons having the same spin. This additional term is an *exchange integral* and only appears when interacting electrons have the same spin. As this exchange integral is positive, the exchange interaction lowers the energy so it is favourable for parallel spins to be in different orbitals.

1.3 Post Hartree–Fock Methods

1.3.1 Electron Correlation

The origin of the missing electron correlation is due to representing the wave function as a single Slater determinant, and so the Hartree–Fock method is a mean field approach. Hartree–Fock theory ignores electron correlation. In the context of wave function theory electron correlation energy is defined as $E_{\mathrm{c}} = E_{\mathrm{exact}} - E_{\mathrm{HF}}$. Electron correlation is classified into two primary contributions: dynamic and static. Dynamic electron correlation represents the correlated instantaneous motions between electrons; in the Hartree–Fock framework the electron experiences only an average interaction due to the other electrons, which does not account for this instantaneous motion. The static electron correlation error stems from the approximation of the wave function as a single Slater determinant being poor when there are several electronic states with similar energies (near degeneracies), this leads to poor descriptions of bond dissociation, the description of delocalised and/or multiple bonds in systems such as ozone.

1.3.2 Configuration Interaction

The correlation error can be reduced by representing the wave function as a linear combination of Slater determinants, each representing a different electron configuration, rather than using just a single Slater determinant. Including more Slater determinants should improve the energy, however not all electron configurations are important.[4] The dynamic correlation error can be improved by considering a large number of electron configurations, all of which contribute a little and the static correlation error can be improved by inclusion of a few key determinants. The simplest method based on this is called configuration interaction (CI). The standard approach to applying CI is to initially perform a HF calculation from which a set of M MOs (where M is the number of basis functions) is obtained. For a closed shell system of n electrons, the lowest $n/2$ MOs will be occupied, the rest are unoccupied/virtual orbitals. Alternative Slater determinants can be formed by replacing one of the occupied MOs with one of the virtual MOs. A determinant with only one such replacement is known as a singly-excited determinant as only one electron is moved. This process is then repeated whereby each electron is excited to each virtual MO, leading to a large number of Slater determinants that can be used for the linear combination representing the many electron wave function. CI assumes that each Slater determinant is fixed and thus the optimal coefficients can be calculated by making use of the variational principle.

This process is not limited to single electron excitations, higher numbers of electrons can be excited, this leads to a hierarchy of CI methods: CIS, CISD, CISDT ... full CI. CIS only includes single excitations, CISD includes single and double excitations, CISDT includes single, double and triple excitations and full CI contains all possible excitations. It should be noted that CIS cannot be used to correct the ground state as the single excitations on their own do not mix with the Hartree–Fock determinant. Full CI is the closest it is possible to get to an exact solution to the electronic Schrödinger equation for an arbitrary molecule, however it becomes difficult to perform for any system larger than around 10 electrons.

Complete active space CI (CAS-CI) is a cheaper method of performing configuration interaction on larger systems. The orbitals are classified into three groups: core (always fully occupied), active (partially occupied orbitals) and virtual (always unoccupied). The many electron wave function is then described as a linear combination of Slater determinants for these spaces. All possible excitations in the active orbitals are allowed and determinants are generated for each possible excitation, however the core and virtual orbitals allow no excitations. This enables a smaller but key section of the orbitals to have full CI performed on it but without having the exceptional cost associated with performing full CI on the full set of

system. However the choice of the active space needs to be carefully consider due to difficulties associated with the inclusion of a high number of electrons in the active orbitals (same issues as full CI).

1.3.3 Coupled Cluster

Coupled cluster (CC) theory[8] is an alternative method of accounting for the electron correlation missing from HF. The wave function for coupled cluster is written via an exponential ansatz

$$\psi = e^{\hat{T}}\phi_0 \tag{18}$$

where ϕ_0 is a reference wave function, usually the Slater determinant from HF molecular orbitals. $\hat{T}$ is the cluster operator, which produces a linear combination of excited determinants based on the reference wave function. The use of the exponential ansatz leads to size extensivity of the solution which is not guaranteed in truncated configuration interaction. Size extensivity is the concept that the method has the correct (linear) scaling with the number of electrons[9]. However, truncated CC is not variational.

The cluster operator is expressed in the form shown in equation 19, where $\hat{T}_1$ is the operator of all single excitations; $\hat{T}_2$ is the operator of all double excitations etc. The form of these two operators is shown in equations 20 and 21, respectively.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots \tag{19}$$

$$\hat{T}_1 = \sum_i \sum_a t_i^a \hat{a}^a \hat{a}_i \tag{20}$$

$$\hat{T}_2 = \frac{1}{4} \sum_{i,j} \sum_{a,b} t_{ab}^{ij} \hat{a}^a \hat{a}^b \hat{a}_j \hat{a}_i \tag{21}$$

Here, $\hat{a}^a$ and $\hat{a}_i$ are creation and annihilations operators, respectively. Creation operators increase the number of particles in a given state by one and annihilation operators decrease the number of particles in a given state by one. The indices i and j stand for occupied orbitals and the indices a, b for unoccupied orbitals. The cluster operator governs the type of excitation (single, double, triple etc). They generate from the reference wave function various excited Slater determinants, which contribute to a multideterminant linear combination wave function, with a magnitude controlled by the cluster amplitudes t. Solving for the unknown

coefficients t is required to find the approximate solution from equation 18.

The exponential cluster operator naturally includes higher excitations at each level of truncation due to the various products and powers of the $\hat{T}_n$ introduced (this can seen in the Taylor series expansion of $\hat{(}e)^T$, for example $\hat{T}_1$ and $\hat{T}_2$ give rise to

$$\hat{e}^T = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \cdots = 1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \cdots \tag{22}$$

which is the origin of the size-extensivity, which is the correct scaling behaviour of the method with the number of electrons[9].

Even though this series is finite in practice due to there being a set number of molecular orbitals (and thus number of possible excitations), the solution of the full series is too costly for systems bigger than around 10 electrons, therefore usually only single and double excitations are included [CCSD]. CCSD performs better[10] than the configuration interaction analogue [CISD] due to the exponential operator in truncated systems (i.e. systems that have > 3 electrons). In order to obtain more accurate results, some treatment of triple excitations needs to be included. This is usually done via estimation of the connected triples using many-body perturbation theory [CCSD(T)][11,12].

1.3.4 Møller–Plesset Perturbation Theory

Møller–Plesset (MP) perturbation theory[13] is a special case of Rayleigh–Schrödinger (RS) perturbation theory[14]. RS perturbation theory considers an unperturbed Hamiltonian operator to which a small perturbation is applied (equation 23). λ is an arbitrary parameter, which controls the size of the perturbation.

$$\hat{H} = \hat{H}_0 + \lambda\hat{V} \tag{23}$$

If the eigenvalues and eigenfunctions (ψ) of the unperturbed Hamiltonian are known, it is possible to represent the desired eigenvalues and eigenfunctions of the full Hamiltonian as a Taylor expansion in λ

$$\begin{aligned} \psi &= \psi_0 + \lambda\psi_i + \lambda^2\psi_{ii} + \cdots \\ E &= E_0 + \lambda E_i + \lambda^2 E_{ii} + \cdots \end{aligned} \tag{24}$$

It is straightforward to derive equations for first-order, second-order etc. corrections to the eigenfunctions and eigenvalues. These corrections allow the approximation of the eigenvalues and eigenfunctions of the full Hamiltonian using just the eigenvalues and eigenfunctions of the zeroth-order Hamiltonian.

In MP perturbation theory, the sum of one-electron Fock operators is the

zeroth-order Hamiltonian (equation 25). Then the perturbation follows the form of equation 26.

$$\hat{H}_0 = \sum_{k=1}^{n} \hat{f}_k \tag{25}$$

$$\hat{V} = \hat{H} - \sum_{k=1}^{n} \hat{f}_k \tag{26}$$

Evaluation of the energy with the Hartree–Fock wave function gives an energy identical to the sum of the zeroth-order and first-order energies in MP perturbation theory

$$E_{\mathrm{HF}} = \langle \psi^{\mathrm{SD}} | H | \psi^{\mathrm{SD}} \rangle = E_0 + E_1 \tag{27}$$

Hartree–Fock is exactly correct to the first order of MP perturbation theory. Thus the first change in energy is seen from the second-order correction (MP2). The second-order correction to the energy is

$$E_2 = \frac{1}{4} \sum_{i,j} \sum_{a,b} \frac{(ij||ab)(ab||ij)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}$$

$$(ij||ab) = \int \frac{\phi_i(1)\phi_j(2)\phi_a(1)\phi_b(2)}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 - \int \frac{\phi_i(1)\phi_j(2)\phi_b^*(1)\phi_a^*(2)}{r_{12}} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{28}$$

It follows that the second-order MP energy (MP2) is simply the sum of the Hartree–Fock energy and the second-order energy correction E_2. Practically, MP2 calculations involve an initial Hartree–Fock calculation to obtain a set of one-electron orbitals (and the associated Hartree–Fock energy) then applying the second-order correction to determine the MP2 energy. One advantage of this method is the relatively low computational cost (can be performed with large systems). However, it is important to note that the reference Hartree–Fock determinant is required to be a good approximation for MP perturbation theory to be a good approximation.

1.4 Localised Basis Sets

A basis set is a set of functions (basis functions) which are used to represent the electronic wave function, in order to enable efficient implementation on computers. This enables the single particle states to be expressed as linear combinations of basis functions.

There are two major types of basis sets used in quantum chemistry, atom-centred (localised) and plane wave basis sets. Localised basis sets build from atomic orbitals as these are a good approximation for molecular orbitals. These basis sets work according to chemical insight and small basis sets give fairly good results. They have some issues though, namely they are non-orthogonal, dependant on atomic positions and are susceptible to basis set superposition error (BSSE). The calculation of interaction energies is especially susceptible to BSSE. For example, when two molecules are brought closer together their basis functions will overlap. Each molecule "borrows" functions from other nearby components, effectively increasing its basis set size and improving the flexibility of the electron description in calculation of derived properties such as energy. If the total energy is minimised as a function of the system geometry, the short-range energies from the mixed basis sets must be compared with the long-range energies from the unmixed sets, and this mismatch introduces an error.

Plane wave basis sets are generally used in periodic systems, where localised basis sets have their own issues. They are dependent on the volume of the simulation cell and independent of the atomic positions (which leads to an elimination of BSSE). They assume periodic boundary conditions which cause them to be less suitable for gas-phase molecular calculations than localised basis sets. Due to large volumes of vacuum needing to be added around each gas-phase molecule to eliminate any interaction between the molecule and its periodic copies, the inclusion of this extra volume will lead to a vastly increased computational cost compared to the localised basis sets due to plane wave basis sets being dependent on the volume of the simulation cell. They are the natural choice for calculation involving periodic boundary conditions such as interactions on a surface.

The calculation cost increases with the size of the basis set therefore the smallest basis set possible is desirable. However, in order to obtain accurate results a large basis set is required; this leads to a compromise between accuracy and cost when selecting a basis set.

1.4.1 Slater-Type Orbitals

Basis functions are typically based on atomic orbitals as these give a good approximation to molecular orbitals. The atomic orbitals from the one electron hydrogen atom (figure 2) can be used to make a basis set. However, there is a possible error introduced when applying these one-electron orbitals to systems with many electrons interacting. Slater proposed a modification to the hydrogenic orbitals, so called Slater-type orbitals,[15] which have effective nuclear charges chosen to account for interelectronic interactions and shielding. Slater-type orbitals have the

general form

$$\chi_{lm}^{\text{Slater}}(r,\theta,\phi) = Nr^l e^{-\alpha r} Y_{lm}(\theta,\phi) \tag{29}$$

where N is a normalisation constant, $r^l e^{-\alpha r}$ is the radial part of the orbital (which is an exponential multiplied by a polynomial) and $Y_{lm}(\theta,\phi)$ is the angular part (here, the spherical harmonics similar to hydrogenic orbitals). The form of Slater-type orbitals means that $n_e \geqslant 1$ which means that the radial part is always positive, thus they lack radial nodes. Slater-type orbitals reproduce the cusp behaviour close to the nuclei.

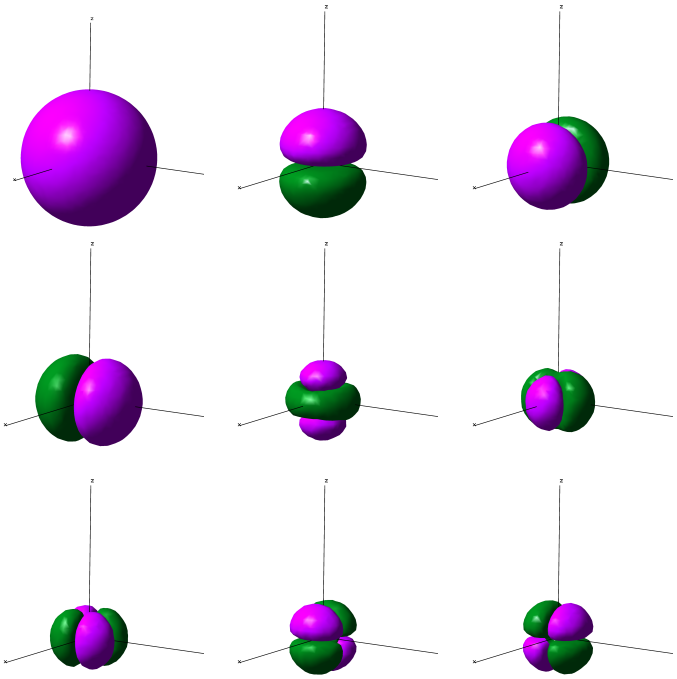


Figure 2: Angular parts of the Hydrogen Atomic Orbitals

1.4.2 Minimal Basis Sets

Practical use of Slater-type orbitals is difficult and inefficient, for example due to difficulties in calculating the 2-electron Coulomb and Exchange integrals, which occur in the Fock matrix. Boys[16] in 1950 noticed that calculating 2-electron integrals for Gaussian orbitals is straightforward due to the Gaussian product rule. Gaussian orbitals have the general form

$$\chi_{lm}^{\mathrm{GTO}}(r,\theta,\phi) = Nr^l e^{-\alpha r^2} Y_{lm}(\theta,\phi) \tag{30}$$

where N is a normalisation constant, $r^l e^{-\alpha r^2}$ is the radial part and $Y_{lm}(\theta,\phi)$ is the angular part. Pople[17] in 1969 suggested fitting n Gaussian orbitals to a single Slater-type orbital which could then be used as a basis set. The result is commonly known as a STO-nG basis set where n has a value of between 2 and 6. These are examples of a minimal basis set, as they only include enough orbitals to represent all the electrons on a neutral atom, e.g 1s for H, 1s 2s 2p for C.

1.4.3 Split Valence Basis Sets

In molecular bonding, it is primarily the valence electrons that take part. Due to this, it is common practice to represent valence orbitals by more than one basis function but still represent the core orbitals by a single basis function. The behaviour of the core electrons does not change much as a result of the environment, for example the 1s orbital of a carbon in a benzene ring is much like the 1s orbital of a carbon atom (comparatively to the valence electrons). Core basis functions are typically comprised of a linear combination of Gaussian functions. Basis sets in which there are more than one basis function for each valence orbital are called valence double/triple/quadruple-ζ basis sets. This refers to how many basis functions each valence orbital is 'split' into.[18]

1.4.4 Polarisation and Diffuse Functions

Polarisation and diffuse functions are additional functions added to improve the flexibility of a basis set. Polarisation functions add atomic orbitals of higher angular momentum to atoms. This helps improve the description of bonding between molecules by allowing polarisation of the electron density on an atom. Diffuse functions improve the description of diffuse electron densities, e.g. anions, by the addition of higher shell functions (e.g. 2s to H).

1.4.5 Correlation-Consistent Basis Sets, cc-pVnZ

Dunning and coworkers developed a set of basis sets starting in 1989,[19] which are designed to converge systematically to the complete-basis-set limit. For first and second row atoms the basis sets are cc-pVnZ where n=D, T etc (D=double, T=triples). The cc-p stands for correlation consistent polarised, this means they include incrementally larger shells of polarisation functions (d, f etc). These can be augmented with additional diffuse functions to improve different calculation types.

1.5 Density Functional Theory

Reliable wave function theory methods are computationally expensive, due to the variational problem in optimising the electron energy in wavefunction theory being $3N$-dimensional (where N is the number of electrons). Therefore, it is useful to consider alternatives in order to reduce this potentially large computational cost when considering chemically relevant systems. Using a density based method is most common alternative to wavefunction based method. Density based methods have a few advantages over wavefunction based methods, they are easier to visualise for systems beyond one electron due to the wavefunction being a $3N$-dimensional entity whereas the density is always 3-dimensional. Useful wavefunction techniques have poor scaling with the number of electrons, which leads to large costs when considering medium sized chemical systems, the scaling is not as poor with density based methods.

1.5.1 Ground-State Density Functional Theory

Density functional theory (DFT) has its basis in the Hohenburg–Kohn theorems set out in 1964.[20] The first theorem states that the ground-state charge density of a system of interacting electrons determines the external potential to within an additive constant (the ground state electronic energy is completely determined by the electron density $\rho(\mathbf{r})$). There are representability problems apparent from this first theorem. Given a trial ground-state charge density that integrates to N electrons, are we sure that the density is coming from an N-electron wave function and sure that it is the ground-state density for an external potential. There is also a problem with how to calculate nontrivial properties such as the ground-state energy of the trial density. The second Hohenburg–Kohn theorem establishes a variational principle, giving that the energy from a trial density is greater than or

equal to the true ground-state energy

$$F[\rho] + \int v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})dr \geqslant E_0 \tag{31}$$

The functional $F[\rho]$ is universal in the sense that it is independent of the external potential. However the form of $F[\rho]$ in general is unknown and therefore it must be approximated in order to obtain a practical theory. Approximations derived purely from the Hohenburg–Kohn theorems are pure density functional or orbital-free theories.

1.5.2 Hohenburg–Kohn–Sham

The Kohn–Sham[21–23] formulation of DFT introduces a fictious system of noninteracting electrons with the same density as the real system. For such a system, a single Slater determinant is sufficient to exactly represent the exact wavefunction. The density is given as

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} |\phi_i(\mathbf{r}, \sigma)|^2 \tag{32}$$

The total energy expression in this formulation is

$$\begin{aligned} E &= T_{\text{s}} + V_{\text{ne}} + J + E_{\text{xc}} \\ &= \sum_{i=1}^{n} \int \phi_i(\mathbf{r}_1)(-\frac{1}{2}\nabla^2)\phi_i(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1 - \sum_{J=1}^{N} \int \frac{q_J\rho(\mathbf{r}_1)}{r_{1J}}\mathrm{d}\mathbf{r}_1 + \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \\ &+ E_{\text{xc}}[\rho] \end{aligned} \tag{33}$$

This equation is similar in form to the Hartree–Fock energy expression, which of course is also based on single determinant. E_{xc} is the exchange-correlation (xc) energy functional; this term contains all the parts of the energy we cannot express exactly (i.e. the quantum effects). The total electronic energy, E, is exact when the exact xc-functional is used. The xc-functional contains not only the exchange and correlation energies but also the difference between the kinetic energy of the interacting and noninteracting systems. The exchange–correlation functional must be approximated as its exact form is unknown (as with $F[\rho]$). Such approximations are known as density functional approximations (DFA). The use of the self-consistent field method to get the orbitals and construct the density from these.

1.5.3 Perdew's 'Jacob's Ladder'

In 2001, Perdew organized these DFAs into a hierarchy, the 'Jacob's Ladder' of DFAs for exchange–correlation[24] (figure 3). The basis of this hierarchy is that as more information about the behaviour of the density is included (e.g. reduced gradient, local kinetic energy), the functionals produced have the potential to be more accurate. The lower rungs of the ladder are pure density functionals, which follow the original basis of Hohenburg–Kohn–Sham DFT, considering only the density as the working variable for describing E_{xc}. As the higher rungs are reached more information is included however this comes at an increased cost of calculation. It should be noted that although there is a tendency to higher accuracy in the higher rungs, there is no guarantee of higher accuracy.

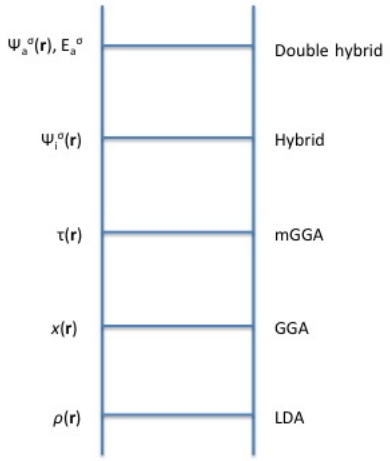


Figure 3: A schematic of Perdew's Jacob's Ladder of DFAs.

1.5.4 Local Density Approximation

Local density approximations (LDA) are the most basic DFA on Jacob's ladder. They express the xc energy in terms of only the density. The form of E_{xc} is derived from the homogeneous/uniform electron gas model. The general form of the LDA E_{xc} (excluding spin effects) is given in equation 34, where ρ is the electron density and ε_{xc} is the exchange–correlation energy per particle from a homogeneous electron gas of charge density ρ.

$$E_{xc}^{LDA}[\rho] = \int \rho(\mathbf{r})\varepsilon_{xc}(\rho)\mathrm{d}\mathbf{r} \tag{34}$$

The exchange–correlation energy is generally a combination of exchange and correlation terms. The exchange-energy density of a homogeneous electron gas is known analytically.[25] The LDA assumes that the exchange-energy in non-homogenous systems can be obtained by applying the homogeneous electron gas exchange-energy pointwise, giving

$$E_{\mathrm{x}}^{\mathrm{LDA}}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(\mathbf{r})^{\frac{4}{3}} \mathrm{d}\mathbf{r} \tag{35}$$

Analytical expressions of the correlation energy for the homogeneous electron gas are only known in the high and low density limits, corresponding to infinitely weak and infinitely strong correlation[26,27]. Highly accurate simulations for the energy of the homogeneous electron gas have been performed for several intermediate values of density, providing accurate values for the correlation energy[28]. The most popular correlation energy LDAs interpolate between these values while reproducing the known behaviour at the high and low density limits. There are several approaches that use different forms to represent E_{c} (the correlation energy). These approaches include Vosko–Wilk–Nusair (VWN)[29] and Perdew–Wang (PW92)[30] approximations to $E_{\mathrm{c}}^{\mathrm{LDA}}$.

1.5.5 Generalised Gradient Approximations

In the LDA, the density at point $\mathbf{r}$ is used to calculate the exchange–correlation energy, with the expressions based on the known behaviour for a homogeneous density. Real systems however have an inhomogeneous spatially varying density $[\rho(\mathbf{r})]$, thus information about how varying the density affects the energy, should be included to improve the accuracy of calculations. The first attempt to do this was via gradient–expansion approximations (GEA)[20]. This approximation tries to systematically include gradient-corrections of the form $|\nabla\rho(\mathbf{r})|, |\nabla\rho(\mathbf{r})|^2, \nabla^2\rho(\mathbf{r})$ etc. to the LDA. The inclusion of low-order gradient corrections had limited success, and the higher order corrections of this type are difficult to calculate. This lead to the concept that instead of a systematic gradient expansion, more general functions of $\rho(\mathbf{r})$ and $\nabla\rho(\mathbf{r})$ should be used. Such approaches are known as generalised gradient approximations and take the general form for spin unpolarised systems as shown in equation 36 and the general form for spin polarised systems as shown in equation 37.

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int f(\rho, \nabla\rho)\mathrm{d}\mathbf{r} \tag{36}$$

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\mathrm{d}\mathbf{r} \tag{37}$$

There is a large variety of choice for the functional $f(\rho, \nabla\rho)$ leading to a large number of GGAs. A broader discussion of the construction of functionals can be found in chapter 2. Example GGAs include: PBE[31] (using Perdew, Burke and Ernzerhof's exchange and correlation functionals), BLYP (combining Becke's exchange functional[32] and Lee, Yang and Parr's correlation functional[33]) and BP86 (combining Becke's exchange functional[32] and Perdew's 1986 gradient corrected correlation functional[34]).

1.5.6 Meta-Generalised Gradient Approximations

GGA functionals are a notable improvement over LDA for many chemically relevant properties but they can still be improved. One approach is via dependence not only on the electon density and its derivatives but also on the Kohn–Sham kinetic-energy density, $\tau(\mathbf{r})$

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_i |\nabla\phi_i(\mathbf{r})|^2 \tag{38}$$

The addition of the kinetic energy density allows additional constraints on E_{xc} to be satisfied, for example a finite exchange potential at the nucleus. An example meta-GGA is Tao–Perdew–Staroverov–Scuseria (TPSS).[35]

1.5.7 Hybrid Functionals

The functionals covered above are all semi-local functionals, i.e. they primarily depend on ρ and its derivatives. Hybrid functionals incorporate a certain amount of non-local HF exchange. Becke introduced this hybrid approach in 1993[36] based on arguments developing a functional based in the adiabatic correction formalism. This first hybrid was the so-called 'half and half' functional (it includes 50% HF exchange). Hybrid functionals improve several molecular properties such as atomisation energies and bond lengths[37]. Hybrid xc functionals are written as a combination of the HF exact exchange functional, $E_{\mathrm{x}}^{\mathrm{HF}}$ (the Hartree–Fock exchange evaluated using the Kohn Sham orbitals), and any number of the exchange and correlation density functionals discussed above (LDA, GGAs). The general form for E_{xc} in hybrid functionals is shown in equation 39

$$E_{\mathrm{xc}} = a_0 E_{\mathrm{x}}^{\mathrm{HF}} + (1 - a_0)E_{\mathrm{x}}^{\mathrm{DFA}} + E_{\mathrm{c}}^{\mathrm{DFA}} \tag{39}$$

The most widely used hybrid functional is B3LYP,[38,39]. This incorporates 20% HF exchange ($a_0 = 0.2$) and uses the Becke 88 gradient correction exchange functional ($E_{\mathrm{x}}^{\mathrm{B88X}}$), the Lee, Yang and Parr correlation functional[33] ($E_{\mathrm{c}}^{\mathrm{GGA}}$) and the Vosko–Wilk–Nusair local density approximation[29] to the correlation functional ($E_{\mathrm{c}}^{\mathrm{LDA}}$). The xc functional for B3LYP is defined as

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = 0.8E_{\mathrm{x}}^{\mathrm{LDA}} + 0.72E_{\mathrm{x}}^{\mathrm{B88X}} + 0.2E_{\mathrm{x}}^{\mathrm{HF}} + 0.19E_{\mathrm{c}}^{\mathrm{VWN}} + 0.81E_{\mathrm{c}}^{\mathrm{LYP}} \tag{40}$$

Other notable hybrid functionals include PBE0[40] and M06[41] amongst others. PBE0 involves mixing PBE exchange and HF exchange in a 3 to 1 ratio, i.e. $a_0 = 0.25$, 25% HF exchange, along with full PBE correlation. The form of the PBE0 xc functional is therefore

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = 0.25E_{\mathrm{x}}^{\mathrm{HF}} + 0.75E_{\mathrm{c}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{PBE}} \tag{41}$$

1.5.8 Range-Separated Hybrids

When the a_0 in equation 39 is a constant, the hybrid is called a conventional or global hybrid. It is possible to generate a more general mixing of exchange, where the mixing is not constant but some more general function. One approach uses the interelectron distance r_{12} to determine the ratio of exchange missing. Using such an approach, the exchange can be partitioned into short-range and long-range components. This is typically done by splitting the Coulomb operator using the standard error function, erf (equation 42). The left hand term is for short-range interactions and the right hand term is for long-range interactions. The parameter ω defines the range of the separation.

$$\frac{1}{r} = \frac{1 - \mathrm{erf}(\omega r)}{r} + \frac{\mathrm{erf}(\omega r)}{r} \tag{42}$$

Given a set ω value, the HF and DFA exchange energies can be split into short-range (SR), long-range (LR) and full-range (FR) components, following the scheme in equations 43 and 44.

$$E_{\mathrm{x}}^{\mathrm{HF}} = E_{\mathrm{x}}^{\mathrm{SR\text{-}HF}}(\omega) + E_{\mathrm{x}}^{\mathrm{LR\text{-}HF}}(\omega) + E_{\mathrm{x}}^{\mathrm{FR\text{-}HF}}(\omega) \tag{43}$$

$$E_{\mathrm{x}}^{\mathrm{DFA}} = E_{\mathrm{x}}^{\mathrm{SR\text{-}DFA}}(\omega) + E_{\mathrm{x}}^{\mathrm{LR\text{-}DFA}}(\omega) + E_{\mathrm{x}}^{\mathrm{FR\text{-}DFA}}(\omega) \tag{44}$$

When the short-range, long-range and full-range parts are mixed separately, a general expression for the E_{xc} of a range-separated hybrid functional is obtained,

$$\begin{aligned} E_{\mathrm{xc}} = {} & aE_{\mathrm{x}}^{\text{SR-HF}}(\omega) + (1-a)E_{\mathrm{x}}^{\text{SR-DFA}}(\omega) + bE_{\mathrm{x}}^{\text{LR-HF}}(\omega) + (1-b)E_{\mathrm{x}}^{\text{LR-DFA}}(\omega) \\ & + cE_{\mathrm{x}}^{\text{FR-HF}}(\omega) + (1-c)E_{\mathrm{x}}^{\text{FR-DFA}}(\omega) + E_{\mathrm{c}}^{\text{DFA}} \end{aligned} \tag{45}$$

The choice of the parameters a, b and c in equation 45 can lead to several different applications of range-separated hybrids. Setting $b = 0$, removes the long-range HF exchange. This reduces computational cost greatly for extended systems, enabling hybrid DFT calculations on bulk metals. A functional of this type developed by Heyd, Scuseria and Ernzerhof[42] (HSE) has been shown to be effective for solid-state studies.[43]

Setting $b = 1$, where long-range exchange is completely described by HF, gives a completely different application of range-separated hybrid functionals. This scheme means the long-range HF exchange serves as an asymptotic correction to the exchange potential. This method has been shown to improve some properties[44]. An example of this type of range-separated hybrid functional is LC-ωPBE,[45] which has parameters $c = 0$, $b = 1$, $a = 0$ and $\omega = 0.4$. Another functional of this type is ωB97XD[46], which has parameters $c = 0.22$, $b = 1$, $a = 0$ and $\omega = 0.2$.

The thermochemical performance of a functional can be influenced by the long-range correction, thus fitting the parameters (a, b) independently between the values of 0 and 1 (i.e. not enforcing $a + b = 1$) could lead to better thermochemical accuracy[47]. This would remove the asymptotic behaviour of having $a+b = 1$. One functional of this type is CAM-B3LYP,[48] the parameters used in this functional are $c = 0.19$, $b = 0.46$, $a = 0$ and $\omega = 0.33$.

1.6 Excited States of Molecules

The properties of excited states, in particular their energies, character and geometries, are important for explaining electronic spectra. They are of particular relevance for applications such as dye-sensitised solar cells, luminescent transition metal complexes and bioimaging[49]. However, it is difficult to explain the experimental results of investigation of these properties, therefore computational methods are used to rationalise the results. The most commonly used methods for the calculation of excited state properties are the linear response formulations of wave function-based methods and density-based methods of ground state calculations. The most commonly used of these methods is linear response time-dependent DFT (TDDFT).

1.6.1 Time-Dependent Density Functional Theory

The linear response formulation of time-dependent density functional theory (TDDFT)[50,51] is the most widely used method for the calcuclation of excited state properties for medium/large molecular systems[52]. There are two major components for the foundation of TDDFT: the Runge–Gross Theorem[53,54] and the time-dependent Kohn–Sham equation[54].

1.6.1.1 Runge–Gross Theorem

The Runge–Gross Theorem[53,54] is seen as the time-dependent analogue of the first Hohenburg–Kohn theorem and such is the formal basis of the time-dependent Kohn–Sham formalism. The derivation of the Runge–Gross theorem begins with the time-dependent Schrödinger equation (equation 1) where the time-dependent Hamiltonian has the form given in equation 46. The operators $\hat{T}(\mathbf{r})$, $\hat{V}_{\mathrm{ee}}(\mathbf{r})$ and $\hat{V}_{\mathrm{ne}}(\mathbf{r})$ correspond to the kinetic energy operator and potentials corresponding to the electron–electron repulsion and the electron–nuclei attraction, respectively.

$$\hat{H}(\mathbf{r},t) = \hat{T}(\mathbf{r}) + \hat{V}_{\mathrm{ee}}(\mathbf{r}) + \hat{V}_{\mathrm{ne}}(\mathbf{r}) + \hat{V}_{\mathrm{ext}}(t) \tag{46}$$

$$\hat{V}_{\mathrm{ext}}(t) = \sum_{i}^{n} \hat{v}(\mathbf{r}_i, t) \tag{47}$$

The term $\hat{V}_{\mathrm{ext}}(t)$ is the time-dependent external potential, which is given by the sum of time-dependent one-particle potentials, equation 47. The number of electrons, n, is constant with time. The Runge–Gross theorem assumes $\hat{V}_{\mathrm{ext}}(t)$ is Taylor expandable around intial time t_0. The theorem[53] states that two solutions $\Psi(t)$ and $\Psi'(t)$ to the time-dependent Schrödinger equation which evolve from a fixed initial state $[\Psi_0]$ under the influence of the potentials $v(\mathbf{r},t)$ and $v'(\mathbf{r},t)$ respectively, always lead to different electron densities $\rho(\mathbf{r},t)$ and $\rho'(\mathbf{r},t)$, provided the two potentials $v(\mathbf{r},t)$ and $v'(\mathbf{r},t)$ differ by more than a purely time-dependent function $[C(t)]$ (equation 48).

$$v(\mathbf{r},t) \neq v'(\mathbf{r},t) + C(t) \tag{48}$$

The proof of this theorem has two parts. The first part shows a one-to-one mapping between an external potential and a current density, the second part shows a one-to-one mapping between a current density and an electron density. Thus overall there is a one-to-one correspondence between an external potential and an electron density. Demonstrating the one-to-one correspondence between an exter-

nal potential and a current density is accomplished by expanding two potentials (equation 48) in a Taylor series around t_0, which differ by more than a purely time-dependent function[53] Assuming there exists some non-negative integer k such that equation 49 is satisfied,

$$\frac{\partial^k}{\partial t^k}[v(\mathbf{r},t) - v'(\mathbf{r},t)]\bigg|_{t=t_0} \neq \text{constant} \tag{49}$$

Consider the current densities arising from each potential, the current density is given by equation 50. The system corresponding to each potential differs from the other by their single-body potential, the equation-of-motion for the difference between the two current densities is given in equation 51 where $\rho(\mathbf{r},0)$ is the initial electron density.

$$j(\mathbf{r},t) = \frac{1}{2i}[\Psi^*(\mathbf{r},t)\nabla\Psi(\mathbf{r},t) - \nabla\Psi^*(\mathbf{r},t)\Psi(\mathbf{r},t)] \tag{50}$$

$$\begin{aligned}\frac{\partial}{\partial t}\{j(\mathbf{r},t) - j'(\mathbf{r},t)\}_{t=0} &= -\mathrm{i}\left\langle \Psi_0 \middle| \left[\hat{j}(\mathbf{r},t), \{\hat{H}(0) - \hat{H}'(0)\}\right] \middle| \Psi_0 \right\rangle \\ &= -\mathrm{i}\left\langle \Psi_0 \middle| \left[\hat{j}(\mathbf{r},t), \{v(\mathbf{r},0) - v'(\mathbf{r},0)\}\right] \middle| \Psi_0 \right\rangle \\ &= -\rho(\mathbf{r},0)\nabla\{v(\mathbf{r},0) - v'(\mathbf{r},0)\}\end{aligned} \tag{51}$$

If at time t_0 the potentials differ by more than just a constant, the first derivative of each of the currents must differ. This leads to a difference in currents when $t > t_0$. A similar relation for higher derivatives can also be derived (equation 52).

$$\frac{\partial^{k+1}}{\partial t^{k+1}}\{j(\mathbf{r},t) - j'(\mathbf{r},t)\}_{t=0} = -\rho(\mathbf{r},0)\nabla\frac{\partial^k}{\partial t^k}\{v(\mathbf{r},t) - v'(\mathbf{r},t)\}_{t=0} \tag{52}$$

Since equation 49 is valid and the potentials are both Taylor expandable about t_0, there exists a positive integer, k, such that $-\rho(\mathbf{r},0)\nabla\{v(\mathbf{r},0) - v'(\mathbf{r},0)\} \neq 0$ (equation 53). This establishes the one to one mapping of potentials and current densities.

$$j(\mathbf{r},t) \neq j'(\mathbf{r},t) \tag{53}$$

The one-to-one mapping of current densities and electron densities still needs to be proven. This connection is made by initially taking the gradient of equation 52 and the continuity of the Schrödinger equation which leads to equation 54.

$$\frac{\partial^{k+2}}{\partial t^{k+2}}\left\{\rho(\mathbf{r},t)-\rho'(\mathbf{r},t)\right\}_{t=0}=\nabla\cdot\left[-\rho(\mathbf{r},0)\nabla\frac{\partial^{k}}{\partial t^{k}}\left\{v(\mathbf{r},t)-v'(\mathbf{r},t)\right\}_{t=0}\right] \tag{54}$$

The right hand side of equation 54 must be non-zero for some integer, k, such that the density difference is non-zero. The proof of this is through contradiction. Setting $f(\mathbf{r})=\frac{\partial^k}{\partial t^k}\{v_{\text{ext}}(\mathbf{r},t)-v'_{\text{ext}}(\mathbf{r},t)\}|_{t=0}$ and considering equation 55.

$$\int f(\mathbf{r})\nabla\cdot\left[\rho_0(\mathbf{r})\nabla f(\mathbf{r})\right]\mathrm{d}\mathbf{r}=\int\left\{\nabla\cdot\left[f(\mathbf{r})\rho_0(\mathbf{r})\nabla f(\mathbf{r})\right]-\rho_0(\mathbf{r})|\nabla f(\mathbf{r})|^2\right\}\mathrm{d}\mathbf{r} \tag{55}$$

The first term on the right hand side of equation 55 is recognisable as a surface integral at $r=\infty$. This surface integral decays at least as fast as $-1/r$, such that it vanishes. The other term on the right hand side of equation 55 will be less than zero, which means the left hand side will be non-zero somewhere. Provided $\nabla f(\mathbf{r})$ is non zero somewhere, $\nabla(\rho_0\nabla f(\mathbf{r}))$ can not vanish everywhere. As a result of this the densities $\rho(\mathbf{r},t)$ and $\rho'(\mathbf{r},t)$ will differ by more than a time dependent phase factor.

1.6.1.2 Time-Dependent Kohn–Sham Equation

The Runge–Gross theorem[53] is valid for arbitrary time-dependent potentials, which enables comparison between the interacting system and a fictitious non-interacting system with the same time-dependent density. Considering a system of interacting particles with a time-dependent density $\rho(\mathbf{r},t)$ and a non-interacting system with the same time-dependent density, the one-to-one mapping of densities and potentials means that a local effective potential $v_{\text{KS}}[\rho](\mathbf{r},t)$ for the non-interacting system gives the same density as the interacting system. This was shown by van Leeuwen.[54]

The time-dependent Kohn–Sham equation as defined in reference 55 has the form shown in equation 56 with $\rho(\mathbf{r},t)$ the density of both the interacting and non-interacting system (equation 57).

$$\mathrm{i}\frac{\partial\varphi_j(\mathbf{r},t)}{\partial t}=\left[-\frac{\nabla^2}{2}+v_{\text{KS}}[\rho](\mathbf{r},t)\right]\varphi_j(\mathbf{r},t) \tag{56}$$

$$\rho(\mathbf{r},t)=\sum_{j=1}^{N}|\varphi_j(\mathbf{r},t)|^2 \tag{57}$$

Due to the one-to-one mapping of the density and potentials as discussed above, the potential $v_{\mathrm{KS}}[\rho](\mathbf{r},t)$ is determined from this density (equation 57) and takes the form shown in equation 58, where $v_{\mathrm{H}}(\mathbf{r},t)$ is a the time-dependent Hartree potential and $f_{\mathrm{xc}}(\mathbf{r},t)$ is the exchange-correlation kernel (i.e. the second derivative of $\mathrm{E}_x c$ with respect to ρ).

$$\begin{aligned} v_{\mathrm{KS}} &= v_{\mathrm{ext}}(\mathbf{r},t) + v_H(\mathbf{r},t) + f_{\mathrm{xc}}(\mathbf{r},t) \\ v_{\mathrm{H}}[\rho](\mathbf{r},t) &= \int \frac{\rho(\mathbf{r}',t')}{|\mathbf{r}-\mathbf{r}'|} \end{aligned} \tag{58}$$

This exchange–correlation kernel is the analogue of the exchange–correlation functional in ground state DFT, hence the exact form is unknown but when it is known the TDDFT equation will give exact results. The first approximation made is usually the adiabatic local density approximation (ALDA), which has the non-local time-dependent exchange–correlation kernel replaced with a time-independent local one due to the density varying slowly with time. This approximation leads to the ability to use standard ground-state xc functionals in TDDFT.

1.6.1.3 Linear Response Formalism

The cost of calculating the full solution to the time-dependent Kohn–Sham equation can be very high for medium size systems. In order to reduce this cost the linear response of the system can be used as an approximation to the full solution. Linear response will produce exact excitation energies when using the exact exchange-correlation kernel[56]. Calculating the linear response can be reached using perturbation theory[57].

The Hohenburg–Kohn theorem gives that the initial ground-state is determined by the ground-state density ρ_0, thus the time-dependent density $\rho(\mathbf{r},t)$ is a functional of the external potential only ($\rho(\mathbf{r},t) = \rho[v_{\mathrm{ext}}](\mathbf{r},t)$). The Runge–Gross theorem implies that the functional $\rho[v_{\mathrm{ext}}]$ can be inverted; that the external potential is a functional of the density. Within perturbation theory, when a small perturbation $v_1(\mathbf{r},t)$ is applied, the functional $\rho[v_{\mathrm{ext}}]$ can be expanded as a Taylor series (equation 59)[57] where the subscripts indicate the order of the perturbation and $\rho_0(\mathbf{r},t)$ is the ground state density.

$$\rho(\mathbf{r},t) - \rho_0(\mathbf{r},t) = \rho_1(\mathbf{r},t) + \rho_2(\mathbf{r},t) + \rho_3(\mathbf{r},t) + \cdots \tag{59}$$

The first order response is shown in equation 60 where χ is the density response of the interacting system.[57] Due to the Hohenburg–Kohn theorem, the intial po-

tential v_0 is a functional of the ground-state density ρ_0, thus the response function χ is also a functional of the ground-state density.

$$\begin{aligned} \rho_1(\mathbf{r},t) &= \iint \chi(\mathbf{r},t,\mathbf{r}',t')v_1(\mathbf{r}',t')\mathrm{d}\mathbf{r}'\mathrm{d}t' \\ \chi(\mathbf{r},t,\mathbf{r}',t') &= \left.\frac{\delta\rho\left[v_{\mathrm{ext}}\right](\mathbf{r},t)}{\delta v_{\mathrm{ext}}(\mathbf{r}',t')}\right|_{v_0} \end{aligned} \tag{60}$$

The Runge–Gross theorem holds for non-interacting particles in external potentials $v_s(\mathbf{r},t)$, thus the functional $\rho(\mathbf{r},t) = \rho[v_s](\mathbf{r},t)$ can be inverted to $v_s(\mathbf{r},t) = v_s[\rho](\mathbf{r},t)$. The Kohn–Sham response function is given by equation 61.[57]

$$\chi_s(\mathbf{r},t,\mathbf{r}',t') = \left.\frac{\delta\rho[v_s](\mathbf{r},t)}{\delta v_s(\mathbf{r}',t')}\right|_{v_s[\rho_0]} \tag{61}$$

Due to the Runge–Gross theorem holding for both the interacting and non-interacting particles; a unique functional $v_s[v_{\mathrm{ext}}]$ can be made such that the time-dependent density of the non-interacting and interacting particles is identical. The potential $v_s(\mathbf{r},t)$ corresponsing to a given $v_{\mathrm{ext}}(\mathbf{r},t)$ is the Kohn–Sham potential (equation 58)[57].

Applying the functional chain rule, the functional derivative of v_s with respect to v_{ext} enables a link between the interacting response function (equation 60) to its non-interacting counterpart (equation 62).[57]

$$\chi(\mathbf{r},t,\mathbf{r}',t') = \iint \left.\frac{\delta\rho(\mathbf{r},t)}{\delta v_s(\mathbf{x},\tau)}\frac{\delta v_s(\mathbf{x},\tau)}{\delta v_{ext}(\mathbf{r}',t')}\right|_{v_0} \mathrm{d}\mathbf{x}\mathrm{d}\tau \tag{62}$$

Taking the functional derivative of equation 58 with repsect to the external potential and inserting it into equation 62 gives equation 63 where $\chi_s(\mathbf{r},t,\mathbf{r}',t')$ is the Kohn–Sham response function (equation 61) and $f_{xc}[\rho_0](\mathbf{r},t,\mathbf{r}',t')$ is the exchange–correlation kernel (equation 64).[57]

$$\begin{aligned} \chi(\mathbf{r},t,\mathbf{r}',t') &= \chi_s(\mathbf{r},t,\mathbf{r}',t') + \iiiint \chi_s(\mathbf{r},t,\mathbf{x},\tau) \\ &\times \left(\frac{\delta(\tau-\tau')}{|\mathbf{x}-\mathbf{x}'|} + f_{\mathrm{xc}}[\rho_0](\mathbf{x},\tau,\mathbf{x}',\tau')\right)\chi(\mathbf{x}',\tau',\mathbf{r}',t')\mathrm{d}\mathbf{x}\mathrm{d}\tau\mathrm{d}\mathbf{x}'\mathrm{d}\tau' \end{aligned} \tag{63}$$

$$f_{\mathrm{xc}}[\rho_0](\mathbf{r},t,\mathbf{r}',t') = \left.\frac{\delta v_{\mathrm{xc}}[\rho](\mathbf{r},t)}{\delta\rho(\mathbf{r}',t')}\right|_{\rho_0} \tag{64}$$

Equation 63 relates the non-interacting and interacting systems, thus is the key equation of TDDFT.[57] In order to get the linear response of the density, equation 63 is inserted into equation 60 giving equation 65. The effective potential $v_{\mathrm{KS},1}(\mathbf{r}',t')$ holds the external perturbation $v_1(\mathbf{r},t)$, the Hartree Coulomb potential and the unknown exchange–correlation function.

$$\begin{aligned} \rho_1(\mathbf{r},t) &= \iint \chi_s(\mathbf{r},t,\mathbf{r}',t')v_{s,1}(\mathbf{r}',t)\mathrm{d}\mathbf{r}'\mathrm{d}t' \\ v_{s,1}(\mathbf{r},t) = v_1(\mathbf{r},t) &+ \int \frac{\rho_1(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + \int\int f_{\mathrm{xc}}[\rho_0](\mathbf{r},t,\mathbf{r}',t')\rho_1(\mathbf{r}',t')\mathrm{d}\mathbf{r}'\mathrm{d}t' \end{aligned} \tag{65}$$

Taking the Fourier transform with respect to time gives the frequency dependent linear density response (equation 66). The frequency-dependent Kohn–Sham response function, which is expressed as a sum over all states (equation 67) where f_k is the occupation number of the ground state Kohn–Sham orbital $\psi_k(\mathbf{r})$ with orbital energy ϵ_k.[57]

$$\begin{aligned} \rho_1(\mathbf{r},\omega) = &\int \chi_s(\mathbf{r},\mathbf{y},\omega)v_1(\mathbf{y},\omega)\mathrm{d}\mathbf{y} \\ &+ \iint \chi_s(\mathbf{r},\mathbf{y},\omega)\left(\frac{1}{|\mathbf{y}-\mathbf{y}'|} + f_{xc}[\rho_0](\mathbf{y},\mathbf{y}',\omega)\right)\rho_1(\mathbf{y}'\omega')\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{y}' \end{aligned} \tag{66}$$

$$\chi_s(\mathbf{r},\mathbf{r}',\omega) = \sum_{j,k}(f_k - f_j)\frac{\psi_j(\mathbf{r})\psi_k^*(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_k(\mathbf{r}')}{\omega - (\epsilon_j - \epsilon_k) + i\eta} \tag{67}$$

1.6.1.4 Matrix Formulation of Linear Response

Casida[58] showed that equation 67 can be transformed into a matrix representation. To begin expand equation 67 into equation 68 where i takes values from 1 to N, representing the occupied orbitals, and a takes values from N+1 to ∞, representing the virtual orbitals.

$$\begin{aligned}\chi_s(\mathbf{r},\mathbf{r}',\omega) &= \sum_{j,k}(f_k - f_j)\frac{\psi_j(\mathbf{r})\psi_k^*(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_k(\mathbf{r}')}{\omega - (\epsilon_j - \epsilon_k) + i\eta} \\ &= \sum_{k=1}^{N}\sum_{j=1}^{\infty}\frac{\psi_j(\mathbf{r})\psi_k(\mathbf{r}')\psi_k^*(\mathbf{r})\psi_j^*(\mathbf{r}')}{\omega - (\epsilon_j - \epsilon_k)} - \sum_{k=1}^{N}\sum_{j=1}^{\infty}\frac{\psi_k(\mathbf{r})\psi_j(\mathbf{r}')\psi_j^*(\mathbf{r})\psi_k^*(\mathbf{r}')}{\omega - (\epsilon_j - \epsilon_k)} \\ &= \sum_{i,a}\left(\frac{\psi_a(\mathbf{r})\psi_i(\mathbf{r}')\psi_i^*(\mathbf{r})\psi_a^*(\mathbf{r}')}{\omega - (\epsilon_a - \epsilon_i)} - \frac{\psi_i(\mathbf{r})\psi_a(\mathbf{r}')\psi_a^*(\mathbf{r})\psi_i^*(\mathbf{r}')}{\omega - (\epsilon_a - \epsilon_i)}\right)\end{aligned} \tag{68}$$

Setting P_{ai} and P_{ia} as defined in equation 69 lets the linear response be written as equation 70.

$$\begin{aligned}\mathrm{P}_{ai} &= \frac{\int \psi_i(\mathbf{r}')\psi_a^*(\mathbf{r}')v_{s,1}(\mathbf{r}',\omega)\mathrm{d}\mathbf{r}'}{\omega - (\epsilon_a - \epsilon_i)} \\ \mathrm{P}_{ia} &= \frac{\int \psi_a(\mathbf{r}')\psi_i^*(\mathbf{r}')v_{s,1}(\mathbf{r}',\omega)\mathrm{d}\mathbf{r}'}{-(\omega + (\epsilon_a - \epsilon_i))}\end{aligned} \tag{69}$$

$$\rho_1(\mathbf{r},\omega) = \sum_{i,a}\psi_a(\mathbf{r})\psi_i^*(\mathbf{r})\mathrm{P}_{ai} + \psi_i(\mathbf{r})\psi_a^*(\mathbf{r})\mathrm{P}_{ia} \tag{70}$$

Rearranging equation 69 gives equation 71. The Hartree and exchange–correlation peotentials can be written as equation 72 which allows definition of matrix elements v_{ai} (equation 73) and $\mathrm{K}_{kl,mn}$ (equation 74).

$$\begin{aligned}(\omega - (\epsilon_a - \epsilon_i))\mathrm{P}_{ai} &= \int \psi_i(\mathbf{r}')\psi_a^*(\mathbf{r}')v_{s,1}(\mathbf{r}',\omega)\mathrm{d}\mathbf{r}' \\ (\omega + (\epsilon_a - \epsilon_i))\mathrm{P}_{ia} &= -\int \psi_a(\mathbf{r}')\psi_i^*(\mathbf{r}')v_{s,1}(\mathbf{r}',\omega)\mathrm{d}\mathbf{r}'\end{aligned} \tag{71}$$

$$f_{H_{\mathrm{xc}}}(\mathbf{r},\mathbf{r}',\omega) = \frac{1}{|\mathbf{r}-\mathbf{r}'|} + f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\omega) \tag{72}$$

$$v_{ai} = \int \psi_i(\mathbf{r})v_1(\mathbf{r},\omega)\psi_a^*(\mathbf{r})\mathrm{d}\mathbf{r} \tag{73}$$

$$\mathrm{K}_{kl,mn} = \int\int \psi_k(\mathbf{r})\psi_l^*(\mathbf{r}) f_{H_{\mathrm{xc}}}(\mathbf{r},\mathbf{r}',\omega)\psi_m(\mathbf{r}')\psi_n^*(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{74}$$

This gives two matrix forms of the linear response of the density depending if using v_{ai} (equation 75) or v_{ia} (equation 76).

$$\sum_{j,b}\left\{\left[\delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i - \omega) + \mathrm{K}_{ai,bj}\right]\mathrm{P}_{bj} + \mathrm{K}_{ai,bj}\mathrm{P}_{jb}\right\} = -v_{ai} \tag{75}$$

$$\sum_{j,b}\left\{\left[\delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i - \omega) + \mathrm{K}_{ai,jb}\right]\mathrm{P}_{jb} + \mathrm{K}_{ai,bj}\mathrm{P}_{bj}\right\} = -v_{ia} \tag{76}$$

This enables the construction of a compact notation of equations 75 and 76 (equation 77) with definitions shown in equation 78.

$$\left[\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^* & \mathbf{A}^*\end{pmatrix} - \omega\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right]\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} = \begin{pmatrix}-v_{ai}\\ -v_{ia}\end{pmatrix} \tag{77}$$

$$\begin{aligned}\mathrm{X}_{jb} &= \mathrm{P}_{jb}\\ \mathrm{Y}_{jb} &= \mathrm{P}_{bj}\\ \mathrm{A}_{ia,jb} &= \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \mathrm{K}_{ai,jb}\\ \mathrm{B}_{ia,jb} &= \mathrm{K}_{ia,bj}\end{aligned} \tag{78}$$

When equation 77 is solved at an excitation energy, the right hand side becomes 0. Thus the excitation frequencies must satisfy a non-Hermitian pseudo-eigenvalue problem (equation 79).

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^* & \mathbf{A}^*\end{bmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} = \omega\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \tag{79}$$

The matrix $\mathbf{B}$ couples the positive and negative eigenvalue solutions, which can be interepted as excitation and de-excitation energies[59]. Assuming that this coupling is small, the $\mathbf{B}$ matrix can be set to zero. This leads to a complete decoupling of the excitation and de-excitations. The resulting Hermitian eigenvalue equation (equation 80) can be solved for the excitation energies, doing this is referred to as the Tamm–Dancoff approximation[60,61].

$$\mathbf{AX} = \omega\mathbf{X} \tag{80}$$

1.6.1.5 Properties and Limitations of TDDFT

Results from TDDFT are sensitive to the choice of xc functional as with ground-state DFT, especially between local/GGA functionals and hybrid functionals.[62,63] This means the reliability of TDDFT calculations should be checked against benchmark calculations/experimental data. Ground state xc functionals can be used in TDDFT due to the ALDA shown in section 1.6.2.3.

Even with the approximate functionals used in TDDFT, the excitation energies obtained via TDDFT are generally accurate. The reason for this is that the difference between the Kohn–Sham orbital energies are usually good approximations for the excitation energies.

TDDFT still has problems with certain classes of excitations: Rydberg states, charge-transfer states[64–67] and molecules with extended π-systems.[68,69] Rydberg states are states in which one electron has been excited into a high energy virtual orbital. Charge transfer states are states in which the electron 'moves' a long distance, effectively generating a positively charged and negatively charged part of the molecule. Charge transfer states are usually underestimated using TDDFT; this is again due to incorrect long-range behaviour.

The problem with Rydberg states can be linked to incorrect long-range behaviour of standard XC functionals, as they decay faster than $1/r$. Range-separated functionals that correct this long-range behaviour and should lead to improved description of Rydberg states.

1.7 Overview of Thesis

Chapter 2 describes the assessment, development and tuning of density-functional approximations over ground state and excited state properties. Specifically looking at hybrid functionals of both global and range-separated types. Chapter 3 is concerned with the development of an excited state geometry and emission energy benchmark and the implementation of this benchmark to assess the performance of several density functional approximations for emission energies. Chapter 4 contains a broader applications of TDDFT techniques to porous materials in order to aid understanding of interesting experimental properties of these materials. All DFT and TDDFT calculations were performed using Gaussian 09[70] and all RI-CC2 calculations were performed using TURBOMOLE V6.6[71].

2 Benchmarking and Development of Functionals

The Kohn–Sham framework developed provides a universal E_{xc} that can provide the exact exchange correlation energy in terms of the density. However, the exact form of E_{xc} is unknown, therefore it is necessary to develop useable approximations to E_{xc}. There are two complementary approaches in the development of such approximations:

1. Use exact conditions that it is known E_{xc} must satisfy, to constrain the form of the functional (see reference 72 for some such conditions).

2. Use known exact data that E_{xc} should reproduce to constrain the form.

The use of exact conditions generally has more success in solid-state physics/condensed matter applications[72]. These conditions tend to be more relevant for extended systems than molecular systems as the conditions are generally valid limiting cases of E_{xc} with respect to some scaling or limiting property of the density, e.g. uniform density limit[72]. Such conditions are relevant for solid-state physics/condensed matter, but are less so for molecular systems. Fewer conditions are known that are specifically relevant for molecular systems. Regardless of the origin of the functional, approximate $E_textxcs$ seldom perform uniformly for solid state and molecular applications.

Perdew, Burke and Erzenhof[31] developed a GGA (PBE) to give an accurate description of the response of uniform gas and correct density scaling. Becke[32] developed an exchange GGA (B88 exchange) to get the correct asymptotic behaviour of the exchange energy density. Lee, Yang and Parr[33] developed an approximation for the correlation energy (LYP). BLYP is a commonly used GGA which combines B88 exchange and LYP correlation. PBE is a commonly used GGA which combines PBE exchange and correlation.

The exact data approach is more commonplace when considering approximate E_{xc}s for chemical applications. Typically some physically motivated form is chosen that satisfies some fundamental properties of E_{xc}, and then parameters are introduced that allow it to be calibrated against reference data. A key parameter in the development of hybrid functionals is that controlling the amount of exact orbital exchange (Hartree–Fock exchange, HFx). For example, B3LYP[38] was fitted to reproduce a series of atomisation energies, ionisation potentials and proton affinities accurately; the functional form has been adjusted to minimise the error of these values relative to reference data. The optimisation leads to good performance for the property/reference data that the functional is optimised against but is not

a guarantee of good performance for other properties. This 'exact' data is either derived from experiment or high-level wave function based benchmark calculations in atomic and molecular systems, so is directly relevant to the types of systems that will be studied

The Minnesota family of functionals, developed by the Truhlar group, have complicated forms that are parameterised against various benchmark sets. These families of functionals have different functional forms for different properties. They are based on one underlying GGA and are parameterised to perform well for different properties, no single Minnesota functional is the most robust, it depends on the property of interest. However, assessments of these functionals against other properties raises questions about the accuracy of these functionals for properties not related to those which have been used to parameterise the functionals. These functionals aim for a balanced description for both main group and transition metal chemistry. Comparison to functionals that are optimised solely for main group chemistry on main group chemistry may lead to the poor performance seen with Minnesota functionals. One of families of Minnesota functionals is the Minnesota 06 family[41], based on the GGA M06-L[73]. Global hybrid functionals have been parameterised based on this GGA (M06[41] with 27% HFx, M06-2X[41] with 54% HFx and M06-HF[74,75] with 100% HFx) for accurate performance on different properties. For example, M06-2X is good for main group thermochemistry but less good for transition metal thermochemistry; whereas M06 is fairly good for both main group and transition metal thermochemistry (worse than M06-2X for main group)

Koopmans theorem[76] states that the first ionisation energy is equal to the negative of the HOMO energy. The ionisation energy calculated via the neutral and positively charged species should be equal to this (this is referred to as Koopmans condition). This condition is not usually held in most commonly used DFT approximations, in some cases having a several eV difference between these two values. Baer, Neuhauser and Livshits[77,78] (BNL) parameterised an approximation which satisfies the Koopmans condition.

The standard approach for the development of functionals attempts to have multiple properties described with reasonable accuracy across multiple molecules with a single functional form and a single set of parameters. A key part in the assessment of approximate E_{xc}s is investigating for which properties they display good performance and which properties they display poor performance.

In order to optimise and assess the performance of approximate E_{xc}s, there exist benchmark sets of various properties ranging from thermochemical (for example atomisation energies) to geometric (for example diatomic bond lengths) to absorptions (for example triplet absorptions). These benchmark sets have a large

quantity of high level theoretical/experimentally derived values for the property that the benchmark set is interested in. These benchmark sets are used to optimise/train the parameters in E_{xc} approximations. However, the use of these sets to optimise does not guarantee the performance of the approximation on other properties (the approximation is not guaranteed to be robust). Therefore, the use of other benchmark sets to assess the performance on other properties is another important use of these benchmark sets, as it enables comparison of the performance of approximations against the computational cost (where is the additional cost of more complicated approximated necessary for accurate description of properties).

Grimme and coworkers have carried out a large amount of work in the field of assessing the performance of approximate E_{xc}s. They combined various benchmark sets into a series of databases, which cover thermochemical, kinetic and non-covalent interaction-based properties. These databases are then used to assess the performance of a large range of approximate E_{xc}s. The first database is the GMTKN24[79] (which collected 24 benchmark sets), the second is the so called GMTKN30[80] (which expanded the database to 30 benchmark sets) and the most recent database is the GMTKN55[81] database (which has expanded the database to 55 benchmark sets). This most recent benchmark database is separated by the various property types in the assessment performed by Grimme. This leads to a large amount of information about the performance of various approximate E_{xc}s across many different properties along with all the properties combined. It should be noted however that all the properties in the Grimme benchmark databases are ground state properties.

2.1 Benchmarking and Assessment of Functionals

Benchmarking against data not used when developing the functional (i.e. data not included in the tuning set) can test the versatility of functionals and their performance. Peach et al. performed an assessment of CAM-B3LYP in 2006[82]. The assessment included a range of properties including: atomisation energies and ionisation potentials, reaction barriers, diatomic bond lengths, harmonic vibrational wavenumbers, electronic polarisabilities and excitation energies. The assessment compared B3LYP and CAM-B3LYP, and their behaviour relative to reference values. The assessment of several functionals for the same set of properties and molecules is beneficial as it can suggest the type of functionals (i.e. GGA, global hybrid, range-separated hybrid) that will perform well for each property assessed, along with giving an overall idea of the general performance of each functional on an equivalent set of data.

An assessment of a series of 6 functionals has been undertaken over a range

of properties. The functionals tested are B3LYP, PBE0, B97-2, CAM-B3LYP, LC-ωPBE and ωB97-XD. These functionals are all hybrid functionals, but are separated into two main categories, global hybrid functionals and range-separated hybrid functionals. These functionals differ in the way the HFx is introduced, global hybrids have a constant percentage of Hartree–Fock exchange (HFx) included and the range-separated hybrids scale the percentage of HFx included by the interelectron distance (r); the parameters that scale the HFx percentage are the full range HFx (a), the long range HFx ($a+b$) and the scaling parameter (ω). The global hybrids are B3LYP, PBE0 and B97-2 and the range-separated hybrids are CAM-B3LYP, LC-ωPBE and ωB97-XD. PBE0, B3LYP and CAM-B3LYP were chosen, as there is comparison to results obtained in reference 82 and thus can be used to verify the calculations undertaken are consistent. LC-wPBE was chosen as it is a range-separated hybrid based on PBE, as PBE0 is based on PBE. B97-2 and wB97-XD were chosen to expand the scope of the functionals tested with functionals that had not been assessed in reference 82. This will give a good idea of the differences between the underlying functionals along with which properties it is important to undertake the additional computational cost associated with the range-separated hybrids, which it has little to no affect on the accuracy and which properties it may be detrimental.

B3LYP is a global hybrid with 20% HFx; CAM-B3LYP is a range separated hybrid based on B3LYP which scales the HFx percentage with a=19%, $a+b$=65% and ω=0.33a_0^{-1}. PBE0 is a global hybrid with 25% HFx; LC-ωPBE is a range separated hybrid based on PBE0 which scales HFx with a=0%, $a+b$=100% and ω=0.40a_0^{-1}. B97-2 is a global hybrid with 21% HFx; ω-B97-XD is a range separated hybrid based on B97-2 which scales HFx with a=22.2%, $a+b$=100% and ω=0.2a_0^{-1}.

The properties chosen to assess the functionals are split into two main categories: ground state properties and excited state properties. The ground state properties assessed are: ionisation potentials, electron affinities, atomisation energies, diatomic bond lengths, vibrational frequencies, isotropic electric polarisabilities and reaction barriers. These are the same properties as those investigated in reference 82. The excited state properties assessed are a series of singlet absorptions of organic molecules and a series of triplet absorptions of organic molecules. Details about each property will be given as they are introduced.

2.1.1 Thermochemical Data

The assessment of the functionals will begin with considering some ground state thermochemical properties, these are properties related to the energetics of the atoms or molecules studied, such as the ionisation energies, electron affinities and atomisation energies.

2.1.1.1 Ionisation Energy

The ionisation energy of an atom (or molecule) is defined as the energy required to remove a single electron from an atom (or molecule), if not otherwise specified it refers to the most easily removed electron. The set of atoms and molecules for which we consider are listed in table 1. They are the same molecules studied in reference 82. The ionisation energies are calculated using the energy difference between of a neutral species and a positively charged species (one with one electron removed). All the ionisation potentials are quoted in eV, and the size of the ionisation energies ranges between 5-20 eV.

Table 1: Systems whose ionisation potentials were studied.

Li, Be, B, C, N, O, F, Na, Mg, Al, CH_4, NH_3, H_2, HF, HCl, C_2H_2, C_2H_4, CO, N_2, Cl_2, ClF

The calculated results are compared to reference values from reference 1. The reference values are derived from experiment, vibrational effects have been removed from the experimental values. The errors (calculated − reference) for B3LYP, CAM-B3LYP and PBE0 are compared to the errors found in references 82 and 1. Following reference 1, we use MP2/6-31G* geometries. Also following reference 1, the 6-311+G(3df,2p) is used. The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the functionals studied are presented in table 2.

Table 2: Mean error (d), mean absolute error ($|d|$) and standard deviation (σ) for ionisation potentials (all values in eV), reference values taken from reference 1.

	Reference		Reference		Reference				
	B3LYP	B3LYP	CAM-B3LYP	CAM-B3LYP	PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/eV	0·00	0·00	0·11	0·10	−0·02	−0·02	0·07	−0·11	−0·09
$\|d\|$/eV	0·17	0·17	0·17	0·18	0·18	0·18	0·20	0·17	0·19
σ/eV		0·20		0·21		0·18	0·25	0·25	0·29

The mean errors and mean absolute errors for ionisation potentials are reproduced for B3LYP, CAM-B3LYP and PBE0. This shows that the calculations being performed are consistent with previous results.

B3LYP and CAM-B3LYP

Despite the mean error in the ionisation potentials increasing from 0.00 eV to 0.10 eV between B3LYP and CAM-B3LYP, the mean absolute error remains almost constant, 0.17 eV to 0.18 eV respectively. The mean errors reflect that the range separation tends to increase the ionisation potentials, leading to an average overestimation with CAM-B3LYP. The mean absolute error and standard deviation

shows that the spread for the two functionals is similar. Thus the inclusion of range-separation leads to no change in the accuracy of ionisation potentials calculated, however B3LYP has ionisation potentials spread fairly uniformly around zero (as many are underestimated as overestimated) whereas CAM-B3LYP predicts higher ionisation potentials.

PBE0 and LC-ωPBE

The observations made with B3LYP and its range-separated analogue are also seen between PBE0 and LC-ωPBE. The mean error increases from -0.02 eV to 0.07 eV between PBE0 and LC-ωPBE; again highlighting the range-correction showing a tendency to overestimate ionisation potentials. The mean absolute error also remains similar, 0.18 eV to 0.20 eV respectively.

B97-2 and ωB97-XD

There is a very small decrease in the mean error between B97-2 and ωB97-XD, from -0.11 eV to -0.09 eV and the mean absolute error and standard deviation remain similar. This is contrary to the results for the other range-separations as ωB97-XD still underestimates the ionisation potentials, but this reflects that less extra exact exchange is put into the functional compared to the other range-separated hybrids.

Overall Comparison

The errors for B3LYP and PBE0 are comparable but the errors for B97-2 are increased compared to the other two global hybrid functionals. This suggests that B97-2 has a tendency to underestimate ionisation energies as compared to B3LYP and PBE0.

Conversely to the case of the global hybrid functionals, the range-separated hybrids all show similar mean errors and mean absolute errors. This suggests that all the range-separated hybrids are of an equivalent accuracy for the calculation of ionisation potentials. However it is important to bear in mind that the range-separated hybrids overestimate (apart from ωB97-XD) whereas the hybrids tend to underestimate ionisation potentials. All the functionals tested perform well for ionisation energies; there is not a 'stand out' functional for ionisation energies due to the good performance of all functionals tested.

2.1.1.2 Electron Affinities

The electron affinity of an atom (or molecule) is the energy change when an electron is added to the neutral species. It is a measure of how favourable it is for the neutral atom (or molecule) to gain an electron. The set of atoms and molecules studied

are listed in table 3. They are the same molecules studied in reference 82. All the electron affinities are quoted in eV, the reference values of the electron affinities lies between 0–4 eV.

Table 3: Systems whose electron affinities were studied.

C, O, F, Si, P, S, Cl, CH, CH_2, CH_3, NH, NH_2, OH, SiH
SiH_2, SiH_3, PH, PH_2, SH, O_2, NO, CN, PO, S_2, Cl_2

The calculated results are compared to reference values from reference 1. The reference values are from experiment. The geometries are from reference 1. The errors for B3LYP and PBE0 are compared to previous errors from reference 1. The basis set used in reference 1 was 6-311+G(3df,2p), the same basis set is used for the electron affinities calculated. The mean error, mean absolute error and standard deviations for the functionals studied are given in table 4.

Table 4: Errors for electron affinities (all values in eV).

	Reference B3LYP	B3LYP	Reference CAM-B3LYP	CAM-B3LYP	Reference PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/eV	−0·01	−0·01		0·04	−0·05	−0·04	0·00	−0·09	0·00
$\|d\|$/eV	0·09	0·09		0·10	0·13	0·13	0·13	0·11	0·08
σ/eV		0·12		0·13		0·22	0·15	0·09	0·09

B3LYP and CAM-B3LYP

As with the ionisation potentials, the electron affinity errors between B3LYP and CAM-B3LYP remain almost constant, with a small increase in mean error (underestimation with B3LYP and overestimation with CAM-B3LYP) but mean absolute errors and standard deviations remain constant (same spread of values). This again suggests that the range-separation has no overall detrimental or beneficial effect on the accuracy of electron affinities calculated.

PBE0 and LC-ωPBE

The range separation of PBE0 shows a slight improvement to the mean errors (−0.04 eV to 0.00 eV), again eliminating the underestimation of the global hybrid, however the mean absolute error is unchanged. Thus again the range-separation has little to no influence on the overall accuracy of calculated electron affinities.

B97-2 and ωB97-XD

The long range correction to B97-2 shows an improvement to the mean errors (−0.09 eV to 0.00 eV), again eliminating the underestimation of the hybrid, and a

small improvement in the mean absolute error (0.11 eV to 0.08 eV). This suggests that unlike the other range-separations, there is a small benefit to using ωB97-XD in the case of electron affinities, again due to lower quantity of extra exact exchange in this functional.

Overall Comparison

The three global hybrid functionals have a defined hierarchy for the mean errors, B3LYP then PBE0 then B97-2 reflecting increased underestimation but the mean absolute errors are similar. Therefore as with ionisation potentials the choice between these three functionals seems arbitrary for the accuracy of the calculation of electron affinities.

As with the standard hybrid functionals the mean errors for the RSH show a hierarchy, LC-ωPBE and ωB97-XD then CAM-B3LYP (0.00 eV, 0.00 eV to 0.04 eV); the mean absolute errors also show a hierarchy ωB97-XD then CAM-B3LYP then LC-ωPBE (0.08 eV to 0.10 eV to 0.13 eV). The best performing range-separated hybrid and the overall best performing functional of those tested for electron affinity is ωB97-XD.

2.1.1.3 Atomisation Energies

An atomisation energy is defined as the energy difference between the molecule and the constituent atoms i.e. for benzene, the energy difference between a benzene molecule and 6 isolated carbon and hydrogen atoms. The set of molecules studied are listed in table 5. They are the same molecules studied in reference 82. Atomisation energies quoted in kcal mol^{-1}. The range of values of the reference values is between 20 and 1300 kcal mol^{-1}.

Table 5: Systems whose atomisation energies were studied.

Acetamide, acetic acid, acetone, acetyl chloride, acetyl fluoride, acrylonitrile, $AlCl_3$, allene, aziridine BCl_3, BeH, benzene, BF_3, bicyclobutane C_2H_2, C_2H_3, C_2H_4, C_2H_5, C_2H_6, CCH, CF_3CN, CF_4, CH, CH_2, CH_2CHF, CH_3, $(CH_3)_2CH$, $(CH_3)_3C$, CH_3CH_2O, CH_3Cl, CH_3CO, CH_3O, CH_3OH, CH_4, CHF_3, Cl_2, ClF, ClNO, CO, CO_2, cyclobutene, cyclopropene, dimethylamine, dimethylether, ethanol, ethylchloride, F_2, F_2O, formic acid, furan, H_2, H_2CO, H_2COH, H_2O, H_2O_2, HCl, HCO, HF, HOCl, isobutane, isopropanol, ketene, Li_2, LiF, LiH, methyl cyanide, methyl ethylether, methyl formate, methyl nitrite, methylamine, methylene cyclopropane, N_2, N_2O, Na_2, NaCl, NF_3, NH_3, nitromethane, NO_2, O_3, OH, oxirane, propane, propylchloride, propyne, pyridine, pyrrole, *trans*-ethylamine, trimethylamine, 2-butyne, vinylchloride

The calculated results are compared to reference values from reference 1. The reference values are experimental values. The basis set used in reference 1 was 6-311+G(3df,2p), the same basis set is used for the atomisation energies calculated. The errors for B3LYP, CAM-B3LYP and PBE0 are compared to previous

errors[1,82]. The mean error, mean absolute error and standard deviations for the functionals studied are in table 6. The mean errors and mean absolute errors for atomisation energies are reproduced for B3YLP, CAM-B3LYP and PBE0. This shows that the calculations being performed are consistent with previous results.

Table 6: Errors for atomisation energies (all values in kcal mol^{-1}).

	Reference		Reference		Reference				
	B3LYP	B3LYP	CAM-B3LYP	CAM-B3LYP	PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/kcal mol^{-1}	−2·7	−2·8	1·5	1·5	2·3	2·3	0·4	0·3	0·0
$\|d\|$/kcal mol^{-1}	3·5	3·5	3·3	3·2	4·4	4·4	3·2	2·9	1·8
σ/kcal mol^{-1}		3·5		3·7		5·4	4·1	3·6	2·5

B3LYP and CAM-B3LYP

Despite the mean error in the atomisation energies showing a large change from −2.8 kcal mol^{-1} to 1.5 kcal mol^{-1} between B3LYP and CAM-B3LYP, highlighting an average underestimation with B3LYP and overestimation with CAM-B3LYP, the mean absolute error only shows a small decrease between B3LYP and CAM-B3LYP, 3.5 kcal mol^{-1} to 3.2 kcal mol^{-1} respectively. Thus the range-separation leads to overestimation compared to the underestimation of B3LYP but has minimal effect on the overall spread of the errors observed.

PBE0 and LC-ωPBE

The mean error decreases by a significant amount from 2.3 kcal mol^{-1} to 0.4 kcal mol^{-1} between PBE0 and LC-ωPBE. Overestimation is seen with PBE0 and is almost eliminated with LC-ωPBE. The mean absolute error also decreases by a significant amount from 4.4 kcal mol^{-1} to 3.2 kcal mol^{-1} respectively. The range-separated functional shows a decrease in the errors thus an apparent increase in the accuracy of atomisation energies calculated by a decrease in the spread of the values along with an elimination of the overestimation seen with PBE0.

B97-2 and ωB97-XD

There is a small decrease in the mean error between B97-2 and ωB97-XD (0.3 kcal mol^{-1} to 0.0 kcal mol^{-1}), both functionals seem to on average get the atomisation energies about correct, no over/underestimation, and a larger decrease in the mean absolute error between B97-2 and ωB97-XD (2.9 kcal mol^{-1} to 1.8 kcal mol^{-1}). This is a similar trend to that seen with PBE0 and LC-ωPBE, thus the range-separation again shows an improvement to the accuracy of atomisation energies by a decrease in the spread of the values.

Overall Comparison

There is a well-defined hierarchy of the hybrid functionals, B97-2 has the lowest error and PBE0 and B3LYP have similar errors. The errors suggest that the hybrid of choice for atomisation energies is B97-2.

As with the standard hybrid functionals the range-separated functionals show a hierarchy, ωB97-XD to LC-ωPBE to CAM-B3LYP. The most accurate range-separated hybrid studied for atomisation energy is ωB97-XD due to its low errors compared to the other range-separated hybrids. The functional of choice of those tested for atomisation energies is ωB97-XD due to the low errors.

2.1.1.4 Classical Reaction Barriers

The barriers of a series of simple reactions were studied. The reaction barrier is the energy difference between the reactants and the transition state of the reaction. The set of reactions studied are listed in table 7. They are the same reactions as those studied in reference 82. The units are kcal mol^{-1}, typical values of classical reaction barriers are 0–20 kcal mol^{-1}.

Table 7: Systems whose classical reaction barriers were studied.

$CH_3 + H_2 \rightarrow CH_4 + H$
$OH + CH_4 \rightarrow CH_3 + H_2O$
$H + H_2 \rightarrow H_2 + H$
$OH + NH_3 \rightarrow H_2O + NH_2$
$HCl + CH_3 \rightarrow Cl + CH_4$
$OH + C_2H_6 \rightarrow H_2O + C_2H_5$
$F + H_2 \rightarrow HF + H$
$O + HCl \rightarrow OH + Cl$
$NH_2 + CH_3 \rightarrow CH_4 + NH$
$NH_2 + C_2H_5 \rightarrow C_2H_6\ NH$
$NH_2 + CH_4 \rightarrow CH_3 + NH_3$
$C_5H_8 \rightarrow C_5H_8$
$H_2 + Cl \rightarrow H + HCl$
$CH_4 + H \rightarrow CH_3 + H_2$
$H_2O + C_2H_5 \rightarrow OH + C_2H_6$
$OH + CH_3 \rightarrow O + CH_4$
$PH_2 + H_2 \rightarrow H + PH_3$
$H_2 + HS \rightarrow H + H_2S$
$OH + Cl \rightarrow O + HCl$
$NH_3 + C_2H_5 \rightarrow NH_2 + C_2H_6$
$CH_3 + NH_3 \rightarrow NH_2 + CH_4$

The calculated results are compared to reference values from reference 1. The reference values are from a combination of experimental and theoretical data (see reference 1 and references within). The errors for B3LYP, CAM-B3LYP and PBE0 are compared to previous errors. The mean error, mean absolute error and standard deviations for the functionals studied are in table 8.

Table 8: Errors for classical reaction barriers (all values in kcal mol^{-1}).

	Reference		Reference		Reference				
	B3LYP	B3LYP	CAM-B3LYP	CAM-B3LYP	PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/kcal mol^{-1}	−3·7	−3·6	−2·6	−2·5	−4·3	−4·3	0·0	−3·5	−2·4
$\|d\|$/kcal mol^{-1}	3·8	3·7	2·7	2·8	4·3	4·3	1·2	3·6	2·5
σ/kcal mol^{-1}		2·4		2·1		1·6	1·7	2·3	1·6

The mean errors and mean absolute errors for the reaction barriers are reproduced for B3LYP, CAM-B3LYP and PBE0. This shows that the calculations being performed are consistent.

B3LYP and CAM-B3LYP

The mean errors decrease between B3LYP and CAM-B3LYP by a significant amount from -3.6 kcal mol^{-1} to -2.5 kcal mol^{-1}, reaction barriers are still underestimated but to a lesser degree. The mean absolute errors also decrease by a significant amount from 3.7 kcal mol^{-1} to 2.8 kcal mol^{-1}. This error data suggests that the range-separation has a beneficial effect on the accuracy of the reaction barriers calculated.

PBE0 and LC-ωPBE

The trend observed between B3LYP and CAM-B3LYP is seen between PBE0 and LC-ωPBE. The mean error decreases by a significant amount between PBE0 and LC-ωPBE (-4.3 kcal mol^{-1} to 0 kcal mol^{-1}), underestimation is eliminated by use of LC-ωPBE. The mean absolute errors decrease by a significant amount between PBE0 and LC-ωPBE (4.3 kcal mol^{-1} to 1.2 kcal mol^{-1}). This suggests again that the range-separation has a benefit to the accuracy of reaction barriers calculated.

B97-2 and ωB97-XD

The trend seen with B3LYP and CAM-B3LYP and with PBE0 and LC-ωPBE is observed between B97-2 and ωB97-XD. The mean error decreases from -3.5 kcal mol^{-1} to -2.4 kcal mol^{-1} and the mean absolute error decreases from 3.6 kcal mol^{-1} to 2.5 kcal mol^{-1}. The range-separation against shows improvement of the accuracy of reaction barriers calculated.

Overall Comparison

The three hybrid functionals show a defined hierarchy with B97-2 and B3LYP providing similar accuracy results and with PBE0 giving less accurate answers. This suggests that of the three hybrids either B97-2 or B3LYP should be used for calculation of reaction barriers.

The three range-separated hybrids also show a hierarchy with LC-ωPBE being the most accurate and CAM-B3LYP and ωB97-XD showing similar accuracy. This suggests that the range-separated hybrid that should be used for the calculation of reaction barriers is LC-ωPBE. As shown above the range-separated hybrids have lower errors that the standard hybrid functionals therefore the functional of choice of those studied is LC-ωPBE for calculation of reaction barriers.

2.1.2 Other Properties

The assessment continues by considering a series of other ground state properties, not directly related to the relative energies of the atoms or molecules (for example related to the geometries).

2.1.2.1 Geometry Optimisations - Diatomic Bond Lengths

The bond lengths of a series of diatomic molecules were optimised. The set of molecules studied are listed in table 9. They are the same molecules studied in reference 82. Bond lengths quoted in Å. The reference bond lengths fall in the range of 1–3.5 Å.

Table 9: Systems whose bond lengths were studied.

Li_2, LiNa, LiK, Na_2, NaK, K_2, N_2, NP, NAs, P_2, PAs, As_2, F_2, FCl, FBr, Cl_2, ClBr, Br_2, LiF, LiCl, NaF, NaCl, NaBr, KF, KCl, BCl, BBr, AlF, AlCl, AlBr, CO, CS, CSe, SiO, SiS, SiSe, GeO, GeS

The calculated results are compared to reference values from reference 1. The reference values are experimental values. The errors for B3LYP, CAM-B3LYP and PBE0 are compared to previous errors[1,82]. The mean error, mean absolute error and standard deviations for the functionals studied are in table 10.

Table 10: Errors for diatomic bond lengths (all values in Å).

	Reference B3LYP	B3LYP	Reference CAM-B3LYP	CAM-B3LYP	Reference PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/Å	0·013	0·013	−0·007	−0·006	0·007	0·007	−0·001	0·017	0·002
$\|d\|$/Å	0·017	0·017	0·011	0·013	0·014	0·015	0·027	0·023	0·015
σ/Å		0·017		0·018		0·023	0·039	0·035	0·019

The reference error values, both mean and mean absolute, are reproduced, thus the calculations performed are consistent with those performed previously.

B3LYP and CAM-B3LYP

The mean errors decrease between B3LYP and CAM-B3LYP (0.013 Å to −0.006 Å respectively), overestimation when using B3LYP is eliminated with use of CAM-B3LYP. The mean absolute errors also decrease, albeit by a smaller amount, between B3LYP and CAM-B3YLP (0.017 Å to 0.013 Å respectively). The range-separation slightly improves the accuracy of bond lengths calculated.

PBE0 and LC-ωPBE

The mean errors follow the trend seen in B3LYP and CAM-B3LYP, i.e. a decrease in error between PBE0 and LC-ωPBE (0.007 Å to −0.001 Å respectively). The overestimation when using PBE0 is eliminated with LC-ωPBE. The trend observed in mean absolute errors between PBE0 and LC-ωPBE is the opposite to that seen between B3LYP and CAM-B3LYP. Namely the error increases between PBE0 and LC-ωPBE (0.015 Å to 0.027 Å respectively), increase in spread of values (standard deviation also increases between PBE0 and LC-ωPBE). This suggests that range-separation has a detrimental effect on the accuracy of bond lengths.

B97-2 and ωB97-XD

The trend observed between B3LYP and CAM-B3LYP is again seen between B97-2 and ωB97-XD. The mean error decreases between B97-2 and ωB97-XD (0.017 Å to 0.002 Å respectively), again overestimation is eliminated. The mean absolute error also decreases but by a smaller amount (0.023 Å to 0.015 Å), resulting in a decrease in the spread of values (the standard deviation also decreases between B97-2 and ωB97-XD). This decrease in error suggests that range-separation corresponds to an increase in accuracy of bond lengths calculated.

Overall Comparison

The mean errors between the three hybrids show a hierarchy, with PBE0 to B3LYP to B97-2. The mean absolute errors also show this trend however it is not as pronounced. The data suggests that the most accurate hybrid tested for bond lengths is PBE0.

The mean errors between the three range-separated hybrids show a hierarchy, with LC-ωPBE to ωB97-XD to CAM-B3LYP. However the mean absolute errors show a completely different trend, with CAM-B3LYP to ωB97-XD to LC-ωPBE. This error data suggests that the most accurate range-separated hybrid for bond lengths is CAM-B3LYP or ωB97-XD due to the similarities in the errors seen. The functional of choice of those tested for bond lengths is CAM-B3LYP due to the low errors observed and small spread of values.

2.1.2.2 Vibrational Frequencies of Diatomic Molecules

The harmonic vibrational wavenumbers (the vibration frequency divided by the speed of light in a vacuum) of a series of diatomic molecules were studied. The set of diatomics studied are listed in table 11. They are the same molecules studied in reference 82. Harmonic vibrational wavenumbers quoted in cm^{-1}, the

Table 11: Systems whose harmonic vibrational wavenumbers were studied.

Li_2, LiNa, LiK, Na_2, NaK, K_2, N_2, NP, NAs, P_2, PAs, As_2, F_2, FCl, FBr, Cl_2, ClBr, Br_2, LiF, LiCl, NaF, NaCl, NaBr, KF, KCl, BCl, BBr, AlF, AlCl, AlBr, CO, CS, CSe, SiO, SiS, SiSe, GeO, GeS

vibrational wavenumbers studied fall between 90 and 2400 cm^{-1}. The calculated results are compared to reference values from reference 1. The reference values are experimentally derived values. The errors for B3LYP, CAM-B3LYP and PBE0 are compared to previous errors. The mean error, mean absolute error and standard deviations for the functionals studied are in table 12.

Table 12: Errors for harmonic vibrational wavenumbers (all values in cm^{-1}).

	Reference B3LYP	B3LYP	Reference CAM-B3LYP	CAM-B3LYP	Reference PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/cm^{-1}	6	4	34	33	24	23	51	15	27
$\|d\|/cm^{-1}$	22	22	37	37	30	30	56	29	43
σ/cm^{-1}		34		47		43	61	41	53

The reference error values, both mean and mean absolute, are reproduced, thus the calculations performed are consistent with those performed previously [1,82].

B3LYP and CAM-B3LYP

Both the mean and mean absolute errors increase between B3LYP and CAM-B3LYP. The mean errors increase by a large amount between B3LYP and CAM-B3LYP (4 cm^{-1} to 33 cm^{-1}), CAM-B3LYP overestimates by a significant amount, and the mean absolute errors increase by a smaller but still significant amount between B3LYP and CAM-B3LYP (22 cm^{-1} to 37 cm^{-1}). This suggests that the range-separation has a detrimental effect on the accuracy of calculated vibrational wavenumbers.

PBE0 and LC-ωPBE

The trend observed between B3LYP and CAM-B3LYP is seen between PBE0 and LC-ωPBE. The mean error increases by a large amount between PBE0 and LC-ωPBE (23 cm^{-1} to 51 cm^{-1}). The mean absolute errors increase by a large amount between PBE0 and LC-ωPBE (30 cm^{-1} to 56 cm^{-1}). This suggests again that the range-separation has a detrimental effect on the accuracy of calculated vibrational wavenumbers.

B97-2 and ωB97-XD

The trend observed between B3LYP and CAM-B3LYP and between PBE0 and LC-ωPBE is seen between B97-2 and ωB97-XD. The mean error increases by a

large amount between B97-2 and ωB97-XD (15 cm^{-1} to 27 cm^{-1}). The mean absolute errors increase by a large amount between B97-2 and ωB97-XD (29 cm^{-1} to 43 cm^{-1}). This suggests that the range-separation has a detrimental effect on the accuracy of calculated vibrational wavenumbers.

Overall Comparison

The three hybrids show a well-defined hierarchy of accuracy for the calculation of harmonic vibrational wavenumbers. The errors increase from B3LYP to B97-2 to PBE0. This suggests that the most accurate hybrid tested is B3LYP as it has the lowest errors. It is noted in literature that B3LYP gives particularly good harmonic vibrational wavenumbers so B3LYP being the best of the three hybrids for harmonic vibrational wavenumbers is expected[83].

The mean errors between the three range-separated hybrids show a hierarchy, with ωB97-XD to CAM-B3LYP to LC-ωPBE. However the mean absolute errors show a different trend, with CAM-B3LYP to ωB97-XD to LC-ωPBE. This error data suggests that the most accurate range-separated hybrid for harmonic vibrational wavenumbers is CAM-B3LYP or ωB97-XD due to the similarities in the errors seen and the large increase in error with LC-ωPBE.

2.1.2.3 Isotropic Electric Polarisabilities

The isotropic electric polarisabilities (a measure of the amount the electron density of a molecule can be distorted by an external electric field) of a series of molecules were examined. The set of molecules examined are listed in table 13. They are the same molecules studied in reference 82. The units are atomic units. The range of reference polarisabilities is 6–30 au.

Table 13: Systems whose isotropic electric polarisabilities were studied.

HF, F_2, CO, N_2, CH_4, CO_2, C_2H_4, PH_3, H_2O, H_2S, SO_2, HCl, Cl_2

The calculated results are compared to reference values from reference 1. The reference values are from Brückner doubles (BD) calculations using an augmented Sadlej basis set. The errors for B3LYP, CAM-B3LYP and PBE0 are compared to previous errors. The mean error, mean absolute error and standard deviations for the functionals studied are in table 14.

Table 14: Errors for isotropic electric polarisabilities (all values are in au).

	Reference		Reference		Reference				
	B3LYP	B3LYP	CAM-B3LYP	CAM-B3LYP	PBE0	PBE0	LC-ωPBE	B97-2	ωB97-XD
d/au	0·36	0·45	0·15	0·24	0·03	0·14	−0·25	0·05	0·14
$\lvert d \rvert$/au	0·45	0·51	0·30	0·37	0·25	0·31	0·35	0·28	0·31
σ/au		0·43		0·36		0·38	0·41	0·41	0·39

The reference errors are not reproduced with the calculations performed. This is due to a different basis set being used for the calculations, Sadlej[84] was used in the reference calculations compared to daug-cc-pVTZ. The daug-cc-pVTZ basis set has more basis functions than the previously used Sadlej basis set. For example C in the Sadlej is made up of 20 functions (10s, 6p and 4d); C in daug-cc-pVTZ is made up of 34 functions (20s, 7p, 4d, 3f). As there are more basis functions the results of the calculations are closer to the basis set limit. The errors are larger than the reference errors despite of this because the reference values being compared to are calculated using the Sadlej basis set not the daug-cc-pVTZ. As the same basis set is used for all the calculations the error introduced from comparing to reference values calculated the Sadlej basis set is consistent for all functionals tested, therefore comparison between the functionals can be made.

B3LYP and CAM-B3LYP

The mean errors decrease between B3LYP and CAM-B3LYP by a significant amount from 0.45 au to 0.24 au. The mean absolute errors also decrease by a significant amount from 0.51 au to 0.37 au. This error data suggests that the range-separation has a beneficial effect to the accuracy of polarisabilities calculated. Long range behaviour is important for polarisabilities when using BLYP based functionals

PBE0 and LC-ωPBE

The trend seen between B3LYP and CAM-B3LYP is not seen between PBE0 and LC-ωPBE. The mean errors increase between PBE0 and LC-ωPBE by a significant amount from 0.14 au to -0.25 au, LC-ωPBE underestimates and PBE0 overestimates. The mean absolute errors are almost constant, 0.31 au to 0.35 au respectively. The range-separation overcorrects the overestimation seen with PBE0, which leads to underestimation.

B97-2 and ωB97-XD

The trend seen between PBE0 and LC-ωPBE is observed between B97-2 and ωB97-XD. The mean errors increase between B97-2 and ωB97-XD from 0.05 au to 0.14 au, the mean absolute errors remain almost constant between B97-2 and ωB97-XD from 0.28 au to 0.31 au. This suggests that the range-separation leads to a slight decrease in accuracy.

Overall Comparison

The three hybrids show a well-defined hierarchy of accuracy for the calculation of polarisabilities. The errors increase from B97-2 to PBE0 to B3LYP. This suggests that the most accurate hybrid for the calculation of polarisabilities tested is B97-2 as it has the lowest errors.

The range-separated hybrids all show similar error values. This suggests that the choice between which of the three range-separated hybrids is fairly arbitrary when accurate calculation of the polarisability is wanted. The polarisability is expected to be linked to the excited state energy accuracy as they are both related to changes in the electron density).

2.1.3 Ground State Properties Summary

A summary table of the mean error and mean absolute error for all the ground state properties assessed with each functional is shown in table 15. This will enable easier comparison between the functionals studied across all the ground state properties and thus show what features of a functional is important for each property and across all properties.

Table 15: Summary of errors for all functionals tested across ground state properties.

Property		B3LYP	CAM-B3LYP	PBE0	LC-ωPBE	B97-2	ωB97-XD
Ionisation	d/eV	0·00	0·10	−0·02	0·07	−0·11	−0·09
Energy	$\|d\|$/eV	0·17	0·18	0·18	0·20	0·17	0·19
Electron	d/eV	−0·01	0·04	−0·05	0·00	−0·09	0·00
Affinity	$\|d\|$/eV	0·09	0·10	0·13	0·13	0·11	0·08
Atomisation	d/kcal mol^{-1}	−2·7	1·5	2·3	0·4	0·3	0·0
Energies	$\|d\|$/kcal mol^{-1}	3·5	3·2	4·4	3·2	2·9	1·8
Diatomic	d/Å	0·013	−0·006	0·007	−0·001	0·017	0·002
Bond Lengths	$\|d\|$/Å	0·017	0·013	0·015	0·027	0·023	0·015
Vibrational	d/cm^{-1}	4	33	23	51	15	27
Wavenumbers	$\|d\|$/cm^{-1}	22	37	30	56	29	43
Isotropic	d/au	0·45	0·24	0·14	−0·25	0·05	0·14
Polarisabilities	$\|d\|$/au	0·51	0·37	0·31	0·35	0·28	0·31
Reaction	d/kcal mol^{-1}	−3·6	−2·5	−4·3	0·0	−3·5	−2·4
Barriers	$\|d\|$/kcal mol^{-1}	3·7	2·8	4·3	1·2	3·6	2·5

There are 2 main differences between the global hybrid functionals and the range separated hybrid functionals: the underlying functional (e.g. BLYP for B3LYP and CAM-B3LYP) and the percentage of HFx contained in the functional. Global hybrids have a constant percentage of HFx included, B3LYP has 20% HFx, PBE0 has 25% HFx and B97-2 has 21% HFx. The percentage in the range separated hybrids is not so straightforward to ascertain as it varies dependent

on the interelectron distance. The scaling of the HFx depends on a number of parameters and these are different in the three range separated hybrids. The parameters are the full range HFx (a) which is a constant quantity of HFx across all distances, the long range HFx ($a + b$) which is the maximum percentage HFx allowed at long distances (asymptotic value) and the scaling factor (ω) which governs how rapidly the HFx percentage is attenuated over distance. CAM-B3LYP has $a = 19\%$, $a + b = 65\%$ and $\omega = 0.33\mathrm{a}_0^{-1}$, LC-$\omega$PBE has $a = 0\%$, $a + b = 100\%$ and $\omega = 0.4\mathrm{a}_0^{-1}$ and ωB97-XD has $a = 22.2\%$, $a+b = 100\%$ and $\omega = 0.2\mathrm{a}_0^{-1}$. These two main differences mean that it is possible to understand the influence of the percentage of HFx on the calculated values (the actual values and the accuracy), along with which underlying functional performs better for each property.

Ionisation potentials show higher calculated values with range separated hybrids compared to global hybrids across all the functionals. This is seen in the increase in the mean error values between the global and range separated hybrids. Thus increasing the percentage of HFx increases the calculated ionisation potentials. There is no change in the accuracy of the calculated values as the mean absolute errors remain almost constant between the global and range separated hybrids (there is no change to the spread of the data points).

Electron affinities again show higher calculated values with range separated hybrids compared to global hybrids across all the functionals. Therefore, as with ionisation potentials, increasing the percentage of HFx increases the calculated ionisation potentials. The mean absolute errors remain constant between B3LYP and CAM-B3LYP and PBE0 and LC-ωPBE, thus there is no change in the accuracy of the calculated values with these functionals. However, there is a decrease in the mean absolute error between B97-2 and ωB97-XD, thus an increase in the accuracy of the calculated values. The differences in the trends seen here are due to underlying differences in the functionals.

Atomisation energies show a different story to that shown by ionisation potentials and electron affinities, atomisation energies are more dependent on the underlying functional that the percentage of HFx. There is an increase in the calculated values between B3LYP and CAM-B3LYP (mean error increases), thus a higher percentage of HFx leads to higher calculated atomisation energies. There is a decrease in the calculated values between PBE0 and LC-ωPBE (mean error decreases), thus a higher percentage of HFx leads to lower calculated atomisation energies. The atomisation energies calculated by B97-2 and ωB97-XD remain almost constant showing little to no influence from increasing the HFx in the functional. The spread decreases between the global hybrids and the range-separated hybrids, thus a higher percentage of HFx leads to improved accuracy.

Diatomic bond lengths show lower calculated values with range-separated hy-

brids compared to global hybrids, due to a decrease in the mean errors. Thus increasing the percentage of HFx decreases the calculated bond lengths. The spread decreases between the global and range separated hybrids for B3LYP and CAM-B3LYP along with B97-2 and ωB97-XD but increases between PBE0 and LC-ωPBE. The differences in the trends in accuracy are due to underlying differences in the functionals. Vibrational wavenumbers show a higher calculated value with range-separated hybrids compared to global hybrids, thus more HFx leads to an increase in calculated values. As all the global hybrids overestimate the vibrational wavenumbers, the range-separated hybrids make this overestimation worse. The spread of the data is higher with range-separated hybrids compared to global hybrids, thus there is a decrease in the accuracy with more HFx. Vibrational wavenumbers are a property that is worsened with more HFx (global hybrids perform better than range-separated hybrids).

Isotropic polarisabilities again shows a greater dependence on the underlying functional than the percentage of HFx. There is a decrease in the calculated values between B3LYP and CAM-B3LYP and PBE0 and LC-ωPBE, thus more HFx decreases the calculated values. There is a increase in the calculated values between B97-2 and ωB97-XD, thus more HFx increase the calculated values. The spread decreases between B3LYP and CAM-B3LYP, thus more HFx improves the accuracy. However there is little to no change in the spread between PBE0 and LC-ωPBE and B97-2 and ωB97-XD.

Reaction Barriers show an increase in the calculated values when HFx is increased due to the global hybrid mean errors being lower than those of the range-separated hybrids. The spread decreases with higher HFx as the mean absolute error decreases between the global hybrids and the range-separated hybrids, thus there is higher accuracy when more HFx is included in the functional.

A ranking of the performance (lowest error is 0, highest error is 5) of all the hybrids across the individual properties along with an overall ranking is shown in table 16.

Table 16: Ranking of the performance of the functionals tested across ground state properties.

Property	B3LYP	CAM-B3LYP	PBE0	LC-ωPBE	B97-2	ωB97-XD
Ionisation Potentials	0	3	2	5	1	4
Electron Affinities	1	2	5	4	3	0
Atomisation Energies	4	3	5	2	1	0
Diatomic Bond Lengths	3	0	2	5	4	1
Vibrational Wavenumbers	0	3	2	5	1	4
Isotropic Polarisabilities	5	4	1	2	0	1
Reaction Barriers	4	2	5	0	3	1
Totals /35	17	17	22	23	13	11

Using the ranking of each functionals performance for each property is it possible to tell which functional performs the best across the ground state properties (by summing the ranking, the lowest sum is the best performing functional). The ranking of performance across all the ground state properties is: ωB97-XD $>$ B97-2 $>$ B3LYP $\sim$ CAM-B3LYP $>$ PBE0 $>$ LC-ωPBE. The underlying functional is important to the general performance across ground state properties with the ranking of general performance separated by the underlying functional (B97 is best, BLYP is second and PBE based is worst).

2.1.4 Excited State Energies

An electronic excited state of a molecule is when an electron is moved from the ground state configuration into a higher energy excited state, here restricted to single excitations, where 1 electron is excited, although it is possible to examine double excitations etc., all the molecules considered are closed shell ground state molecules. There are two types of excitations that are studied, singlet and triplet excitations. A singlet excitation is when the electron retains its spin as it is excited (i.e. in a closed shell system the total spin angular momentum is conserved in the excitation, same number of α and β spin electrons) and a triplet excitation the spin of the electron changes as it is excited. A schematic for the two types of excitation studied is shown in figure 4. Excitation energies can be used as another property to benchmark functionals.

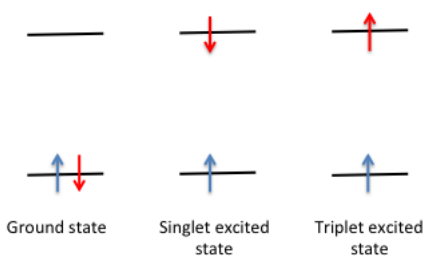


Figure 4: Schematic of singlet and triplet excitations.

2.1.4.1 Singlet Excitations

The set of singlet excitations studied is the same as those in reference 47. The molecules and excitations are listed in table 17. The calculated values are compared to reference excitation energies from reference 47 and references within.

Table 17: Singlet excitations that were studied.

Dipeptide $[n_1 \rightarrow \pi_2^*, \pi_1 \rightarrow \pi_2^*, n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*]$, β-dipeptide $[n_1 \rightarrow \pi_2^*, \pi_1 \rightarrow \pi_2^*, n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*]$
Tripeptide $[\pi_1 \rightarrow \pi_2^*, \pi_2 \rightarrow \pi_2^*, \pi_1 \rightarrow \pi_2^*, n_1 \rightarrow \pi_3^*, n_2 \rightarrow \pi_3^*, n_1 \rightarrow \pi_2^*, n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*, n_3 \rightarrow \pi_3^*]$
Acene (n=1–5) $[B_{2u}, B_{3u}]$, PP $[1B_2, 2A_1, 2B_2, 3A_1]$, DMABN [B, A] PA oligomer (n=2–5) $[B_u]$
N_2 $[\Pi_u, \Sigma_u^+, \Pi_u, \Sigma_g^+, \Delta_u, \Sigma_u^-, \Pi_g]$, CO $[F\Sigma^+, E\Pi, C\Sigma^+, B\Sigma^+, D\Delta, I\Sigma^-, A\Pi]$, HCl $[\Pi]$
H_2CO [A_2, A_2, B_1, B_2, A_1, B_2, A_2]

The excitations are split into 3 types, Local, Rydberg and Charge-Transfer. The excitations of each type are noted in table 18. Local excited states are states in which the excitation is localised i.e. the electron does not move a long distance. Rydberg excited states are states in which the electron is excited to a high lying virtual orbital. Charge-Transfer excited states are states in which the electron 'moves' a long distance, effectively generating a partially positively and negatively charged part of the molecule.

Table 18: Categorisation of singlet excitations.

Local	Rydberg	Charge Transfer
Dipeptide $[n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*]$	N_2 $[\Pi_u, \Sigma_u^+, \Pi_u, \Sigma_g^+]$	Dipeptide $[n_1 \rightarrow \pi_2^*, \pi_1 \rightarrow \pi_2^*]$
β-dipeptide $[n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*]$	CO $[F\Sigma^+, E\Pi, C\Sigma^+, B\Sigma^+]$	β-dipeptide $[n_1 \rightarrow \pi_2^*, \pi_1 \rightarrow \pi_2^*]$
Tripeptide $[n_1 \rightarrow \pi_1^*, n_2 \rightarrow \pi_2^*$	H_2CO [A_2, A_2	Tripeptide $[\pi_1 \rightarrow \pi_2^*, \pi_2 \rightarrow \pi_2^*$
$n_3 \rightarrow \pi_3^*]$	B_2, A_1, B_2]	$\pi_1 \rightarrow \pi_2^*, n_1 \rightarrow \pi_3^*, n_2 \rightarrow \pi_3^*, n_1 \rightarrow \pi_2^*]$
PP [$1B_2$, $2A_1$]		PP [$2B_2$, $3A_1$]
DMABN [B]		DMABN [A]
Acene (n=1–5) $[B_{2u}, B_{3u}]$		HCl [Π]
PA oligomer (n=2–5) $[B_u]$		
N_2 $[\Delta_u, \Sigma_u^-, \Pi_g]$		
CO $[D\Delta, I\Sigma^-, A\Pi]$		
H_2CO [B_1, A_2]		

The mean and mean absolute errors and standard deviations for each group of excitation for the 3 hybrid and the 3 range-separated hybrid functionals are shown in table 19.

Table 19: Errors for singlet excitations (all values are in eV).

Excitation Type		B3LYP	CAM-B3LYP	PBE0	LC-ωPBE	B97-2	ωB97-XD
Local	d	−0·12	0·02	−0·04	0·14	−0·05	0·04
	$\lvert d \rvert$	0·21	0·21	0·19	0·23	0·19	0·19
	σ	0·26	0·29	0·26	0·27	0·27	0·26
Rydberg	d	−1·08	−0·48	−0·87	−0·02	−0·90	−0·61
	$\lvert d \rvert$	1·08	0·48	0·81	0·14	0·90	0·61
	σ	0·23	0·18	0·28	0·17	0·22	0·22
Charge Transfer	d	−1·34	−0·18	−1·11	0·11	−1·25	−0·13
	$\lvert d \rvert$	1·36	0·27	1·14	0·68	1·27	0·25
	σ	0·86	0·31	0·75	0·81	0·83	0·31

Local Excitations

The mean error of the set of local excitations decreases between B3LYP and CAM-B3LYP (−0.12 eV to 0.02 eV). However the mean absolute errors remain constant (0.21 eV). This suggests that the range-separation shows no overall improvement or deterioration in the accuracy of the excitation energies for local excitations, however it eliminates the underestimation of B3LYP. This is understandable as a local excitation does not have an obvious long-range influence.

The mean error of the local excitations increases between PBE0 and LC-ωPBE (−0.04 eV to 0.14 eV), again eliminating but overcorrecting the average underestimation but overcorrects it. The mean absolute error also increases slightly (0.19 eV to 0.23 eV). This shows that this range-separation (LC-ωPBE) has a small detrimental effect to the accuracy of local excitations.

The mean error is almost constant between B97-2 and ωB97-XD (−0.05 eV to 0.04 eV), again eliminating and overcorrecting the average the underestimation.

The mean absolute error is constant between the hybrid and range-separated hybrid (0.19 eV). This again shows that the long range-separation shows no improvement to the local excitations due to their nature.

The three hybrid functionals show a well-defined hierarchy for the mean errors, B3LYP with the largest, then B97-2 and PBE0 with similar mean errors. The mean absolute errors for the three hybrid functionals are similar. This suggests that the three hybrids are all of similar accuracy for the calculation of local excitations.

The three range-separated hybrids have a well-defined hierarchy for the mean errors from LC-ωPBE to ωB97-XD to CAM-B3LYP. The mean absolute errors of the three range-separated hybrids are all similar. This again suggests that not one of the range-separated hybrids is more accurate than the others for the calculation of local excitations.

Rydberg Excitations

The errors decrease between B3LYP and CAM-B3LYP. The mean error improves from -1.08 eV to -0.48 eV and the mean absolute error drops from 1.08 eV to 0.48 eV. This shows that the range-separation leads to an increase in the accuracy of Rydberg excitations, this has been noted in previous literature[47] and the results confirm it. Both functionals underestimate all of the Rydberg excitations studied.

The errors again decrease between PBE0 and LC-ωPBE. The mean error improves significantly from -0.87 eV to -0.02 eV and the mean absolute error drops from 0.81 eV to 0.14 eV. This shows that the range-separation leads to an increase in the accuracy of Rydberg excitations; this is again concurrent with the effect of range-separation on Rydberg excitations. The LC-ωPBE range separation has eliminated the underestimation from PBE0. 100% long range HFx is important for the accurate calculation of Rydberg states.

The errors again decrease between B97-2 and ωB97-XD. The mean error drops from -0.90 eV to -0.61 eV and the mean absolute error drops from 0.90 eV to 0.61 eV. This shows that again the range-separation leads to an increase in the accuracy of Rydberg excitations.

The three hybrid functionals show a well-defined hierarchy for the errors, B3LYP with the largest, then B97-2 and PBE0 with similar errors. This suggests that B3LYP is less accurate than the other two hybrids for Rydberg excitations.

The three range-separated hybrids have a well-defined hierarchy for the errors, ωB97-XD with the largest, then CAM-B3LYP then LC-ωPBE. This suggests that the most accurate range-separated hybrid tested is LC-ωPBE. The error values observed reflects that the accuracy of Rydberg excitations is almost completely dependent on the amount of HF exchange included at long range (LC-ωPBE has the highest amount with $a + b = 1$ and $\omega = 0.4$).

Charge Transfer Excitations

The errors decrease between B3LYP and CAM-B3LYP. The mean error drops from -1.34 eV to -0.18 eV (underestimation from B3LYP is almost eliminated) and the mean absolute error drops from 1.36 eV to 0.27 eV. This shows that the range-separation leads to an increase in the accuracy of charge transfer excitations; this has been noted in previous literature[47].

The errors again decrease between PBE0 and LC-ωPBE. The mean error drops from -1.11 eV to 0.11 eV (again underestimation is eliminated but overcorrected) and the mean absolute error drops from 1.14 eV to 0.68 eV. This shows that the range-separation leads to an increase in the accuracy of Charge Transfer excitations; this is again concurrent with the effect of range-separation on charge transfer excitations[47].

The errors again decrease between B97-2 and ωB97-XD. The mean error drops from -1.25 eV to -0.13 eV and the mean absolute error drops from 1.27 eV to 0.25 eV. This shows that the range-separation leads to an increase in the accuracy of charge transfer excitations.

The three hybrid functionals show a well-defined hierarchy for the errors, B3LYP with the largest, then B97-2 then PBE0 with the smallest errors. This suggests that PBE0 is the hybrid of choice of those tested for Charge Transfer excitations.

The three range-separated hybrids have a well-defined hierarchy for the errors, LC-ωPBE with the largest, then CAM-B3LYP and ωB97-XD with similar errors. This suggests that LC-ωPBE is the least accurate range-separated hybrid of those tested for charge-transfer excitations.

2.1.4.2 Triplet Excitations

The set of triplet excitations studied are the same as those studied in reference 85. This set of triplet excitations is usually referred, in literature, to as the Thiel set due to the senior author of the benchmark set[86–88]. The molecules in this set are listed in table 20. The molecules are a series of small organic molecules.

Table 20: Thiel set of triplet excitations.

Ethene [B_{1u}], Butadiene [B_u, A_g], Hexatriene [B_u, A_g], Octatetraene [B_u, A_g], Cyclopropene [B_2, B_1], Cyclopentadiene [B_2, A_1]
Norbornadiene [A_2, B_2], Benzene [B_{1u}, E_{1u}, B_{2u}, E_{2g}], Napthalene [B_{2u}, B_{3u}, B_{1g}, B_{2u}, B_{3u}, A_g, A_g, B_{1g}, B_{1g}, A_g], Furan [B_2, A_1]
Pyrrole [B_2, A_1], Imidazole [A′, A′, A″, A′], Pyridine [A_1, B_1, B_2, A_1, A_2, B_2], Tetrazine [B_{3u}, A_u, B_{1g}, B_{1u}, B_{2u}, B_{2g}, A_u, B_{1u}]
Formaldehyde [A_2, A_1], Acetone [A_2 A_1], Benzoquinone [B_{1g}, A_u, B_{1u}, B_{3g}], Formamide[A″, A′], Acetamide[A″, A′], Propanamide [A″, A′]

The literature values for B3LYP and CAM-B3LYP errors are used from reference 85. Calculations of some of these molecules have been undertaken and the

error values are reproduced so the calculations performed are consistent and the reference errors can be used for comparison to the other functionals tested, PBE0, LC-ωPBE, B97-2 and ωB97-XD.

There is a well-documented problem with the triplet excitation energies calculated using TDDFT when the excitation is associated with an orbital transition that has low stability, such low stability excitations have a HF stability of less than 2 eV as reported in references 85,89,90. The method employed to correct this error is the Tamm Dancoff approximation to TDDFT (TDA-TDDFT). Further information about the stability problem for triplet excitations can be seen in references 85,89,90. Both TDDFT and TDA-TDDFT have been used for the calculation of the triplet excitation energies. The mean errors from the 6 functionals using both methods are summarised in table 21.

Table 21: Errors for triplet excitations (all values in eV).

	B3LYP		CAM-B3LYP		PBE0		LC-ωPBE		B97-2		ωB97-XD	
	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA
d/eV	−0·43	−0·26	−0·39	−0·16	−0·48	−0·26	−0·48	−0·14	−0·34	−0·17	−0·29	−0·09
$\|d\|$/eV	0·43	0·27	0·40	0·18	0·48	0·26	0·53	0·22	0·34	0·20	0·30	0·14

TDDFT vs TDA-TDDFT

The reference error values for B3LYP and CAM-B3LYP show a large improvement by applying the TDA-TDDFT method. Further information can be seen in references 85,89,90. This result is confirmed with the 4 additional functionals tested. Both the mean error and mean absolute errors drop drastically between TDDFT and TDA-TDDFT. The excitations that show the largest improvement with the TDA-TDDFT method are those with low HF stabilities (less than or around 2 eV) as observed previously[85,89,90]. The TDA-TDDFT method should be used for triplet excitations.

B3LYP and CAM-B3LYP

The mean and mean absolute errors for B3LYP and CAM-B3LYP for the TDDFT method are comparable (−0.43 eV to −0.39 eV and 0.43 eV to 0.40 eV). This suggests that for TDDFT the range-separation shows no improvement to the accuracy of excitation energies calculated. The mean and mean absolute errors for B3LYP and CAM-B3LYP for the TDA-TDDFT method decrease going from B3LYP to CAM-B3YLP (−0.26 eV to −0.16 eV and 0.27 eV to 0.18 eV). This shows that for the more accurate method (TDA-TDDFT) the range-separation shows an improving effect on the accuracy of the excitations calculated. Both methods underestimate triplet excitations (TDDFT and TDA-TDDFT).

PBE0 and LC-ωPBE

The TDDFT mean and mean absolute errors for PBE0 and LC-ωPBE follow the same trend as B3LYP and CAM-B3LYP, i.e. the errors are similar for both functionals. This again suggests that the range-separation has no effect on the accuracy of triplet excitations calculated with TDDFT. The TDA-TDDFT errors and mean absolute errors also follow the trend seen with B3LYP and CAM-B3LYP, although the difference between the two functionals is much smaller (0.04 eV for mean absolute error). Both methods still underestimate triplet excitations.

B97-2 and ωB97-XD

The two methods again show the same trend as with the other functionals. TDDFT has similar errors for both the hybrid and range-separated hybrid. TDA-TDDFT has slightly improved error for the range-separated hybrid with a similar improvement seen between PBE0 and LC-ωPBE (0.06 eV for mean absolute error). Both methods still underestimate triplet excitations.

Overall Comparison

The three hybrids show a well-defined hierarchy for the errors, B3LYP and PBE0 with similar and largest errors and B97-2 with the smallest errors. This suggests that the hybrid of choice of those studied for triplet excitation energies is B97-2.

The three range-separated hybrids show a well-defined hierarchy for the errors, LC-ωPBE with the largest error, then CAM-B3LYP and B97-2 with similar and the smallest errors. This suggests that the range-separated hybrid of choice of those studied for triplet excitation energies is CAM-B3LYP and B97-2.

2.1.4.3 Excited State Properties Summary

A summary table of the mean error and mean absolute error for all the absorptions assessed with each functional is shown in table 22. This will enable easier comparison between the functionals studied across all the absorptions and thus show what features of a functional is important for absorptions.

Table 22: Summary of Errors for all functionals tested across absorptions (all values in eV).

Property		B3LYP	CAM-B3LYP	PBE0	LC-ωPBE	B97-2	ωB97-XD
Singlet	d	−0·12	0·02	−0·04	0·14	−0·05	0·04
Local	$\lvert d\rvert$	0·21	0·21	0·19	0·23	0·19	0·19
Singlet	d	−1·08	−0·48	−0·87	−0·02	−0·90	−0·61
Rydberg	$\lvert d\rvert$	1·08	0·48	0·81	0·14	0·90	0·61
Singlet	d	−1·34	−0·18	−1·11	0·11	−1·25	−0·13
Charge Transfer	$\lvert d\rvert$	1·36	0·27	1·14	0·68	1·27	0·25
Triplet	d	−0·43	−0·39	−0·48	−0·48	−0·34	−0·29
TDDFT	$\lvert d\rvert$	0·43	0·40	0·48	0·53	0·34	0·30
Triplet	d	−0·26	−0·16	−0·26	−0·14	−0·17	−0·09
TDA-TDDFT	$\lvert d\rvert$	0·27	0·18	0·26	0·22	0·20	0·14

Singlet local excitations show similar performance across all the hybrid functionals tested. The calculated values with the range-separated hybrids are higher than those calculated with global hybrids (mean error is higher for range-separated hybrids), thus the calculated values are increased with higher percentages of HFx. The spread of the data remains constant across all the functionals tested. Singlet local excitations are a property which are calculated to a similar degree of accuracy regardless of the functional used.

Singlet Rydberg excitations show significantly lower calculated values with global hybrids compared to range-separated hybrids. The values of Rydberg excitations are greatly underestimated by global hybrids and this is corrected to some degree by the use of range-separated hybrids (more HFx is better for Rydberg states). There are higher calculated values with higher HFx. The spread of the data decreases between global hybrids and range-separated hybrids, thus more HFx increases the accuracy.

Singlet charge transfer excitations show lower calculated values with global hybrids compared to range-separated hybrids. A higher percentage of HFx increases the calculated values. With global hybrids again underestimating the values of these excitations, more HFx is required for these excitations as with Rydberg excitations. The spread decreases between global hybrids and range-separated hybrids, again showing that increasing the percentage of HFx improves the accuracy.

Triplet TDDFT excitations remain on average fairly constant between the global and range-separated hybrids of each underlying functional. There is some difference between the underlying functionals, with B97 performing best, BLYP in the middle and PBE performing worse. There is a greater influence seen when the

triplet excitations are calculated using TDA-TDDFT, the calculated values are increased with more HFx (range-separated hybrid values are higher than global hybrid values). The spread is also decreased between the global hybrids and range-separated (more HFx has lower spread). Again there is a dependence on the underlying functional with the trend seen when using TDDFT preserved. There is a larger effect seen when using TDA-TDDFT than between the global and range-separated hybrids (there is some improvement seen with the range-separared hybrids and TDA-TDDFT). This follows from the known triplet stability issues, which it is known that TDA-TDDFT can improve[85,89,90].

A ranking of the performance (lowest error is 0, highest error is 5) of all the hybrids across the individual properties along with an overall ranking is shown in table 23.

Table 23: Ranking of the performance of the functionals tested across excited state properties.

Property	B3LYP	CAM-B3LYP	PBE0	LC-ωPBE	B97-2	ωB97-XD
Singlet Local	1	1	0	2	0	0
Singlet Rydberg	5	1	3	0	4	2
Singlet Charge Transfer	5	1	3	2	4	0
Triplet TDDFT	3	2	4	5	1	0
Triplet TDA-TDDFT	5	1	4	3	2	0
Totals /25	19	6	14	12	11	2

There is some definite advantage to having a range-separated hybrid compared to a global hybrid for excited state calculations as all the range-separated hybrids outperform their global counterparts. There also seems to be some advantage to having a percentage of full range HFx included in the range-separated hybrid as CAM-B3LYP and ωB97-XD outperform LC-ωPBE. The performance of LC-ωPBE is lacking compared to the other range-separated hybrids. The performance of LC-ωPBE may be improved by the changing of parameters that govern this range-separation i.e. attenuation in a similar way to CAM-B3LYP ($a + b \neq 100\%$).

2.2 CAM-ωPBE

We have seen that range-separated hybrids offer significant advantages over conventional global hybrid functionals for the study of excited states, and may offer reasonable performance for ground state properties. Notably however, LC-ωPBE often overcorrects the underlying PBE functional, giving rise to overestimated excitation energies, along with other problems. This is attributed to the amount of long-range HFx as RSH with less than 100% long range HFx do not display

this same overestimation, such as CAM-B3LYP. Therefore, a new range-separated was attempted to be generated by changing the range-separation parameters of the LC-ωPBE range-separated hybrid to be like those used in CAM-B3LYP ($a = 0.19$, $b = 0.46$ and $\omega = 0.33$) . This functional will be referred to as CAM-ωPBE. CAM-ωPBE was initially tested on the various properties tested in _ and the errors are compared to those from CAM-B3LYP and LC-ωPBE. The mean and mean absolute errors for the three functionals are in table 24.

It would be expected that the errors for this new functional would be a little higher than those for CAM-B3LYP and LC-ωPBE due to not optimising the parameters of the CAM- type correction. This is not what is seen; the errors are infact much larger for CAM-ωPBE apart from the errors in the classical reaction barriers. The errors in the classical reaction barriers are similar to those from the LC-ωPBE functional. This is potentially an artifact of the range-separation LC-ωPBE being a large improvement to the errors of the reaction barriers over PBE0. It is obvious though that optimisation of the 3 parameters is required to improve the performance of this new functional.

Table 24: Errors for CAM-ωPBE.

Ionisation Potentials	CAM-B3LYP	LC-ωPBE	CAM-ωPBE
d/eV	0·10	0·07	−0·57
$\|d\|$/eV	0·18	0·20	0·57
σ/eV	0·21	0·25	0·32
Atomisation Energies			
d/kcal mol^{-1}	1·5	0·4	29·6
$\|d\|$/kcal mol^{-1}	3·2	3·2	31·4
σ/kcal mol^{-1}	3·7	4·1	25·7
Diatomic Bond Lengths			
d/Å	−0·006	−0·001	−0·032
$\|d\|$/Å	0·013	0·027	0·066
σ/Å	0·018	0·039	0·066
Classical Reaction Barriers			
d/kcal mol^{-1}	−2·5	0·0	−0·4
$\|d\|$/kcal mol^{-1}	2·8	1·2	1·9
σ/kcal mol^{-1}	2·1	1·7	2·4
Vibrational Wavenumbers			
d/cm^{-1}	33	51	131
$\|d\|$/cm^{-1}	37	56	131
σ/cm^{-1}	47	61	117
Isotropic Polarisabilities			
d/au	0·24	−0·25	−1·13
$\|d\|$/au	0·37	0·35	1·13
σ/au	0·36	0·41	0·66
Electron Affinities			
d/eV	0·04	0·00	−0·95
$\|d\|$/eV	0·10	0·13	0·95
σ/eV	0·13	0·15	0·22

2.2.1 Calibration of CAM-ωPBE

The calibration of the CAM-ωPBE functional was carried out on a small subset of the atomisation energies in a similar manner to reference 91 (the reduced subset is listed in table 25) and the best performing (lowest error parameter sets) are carried forward to the complete set of properties.

Table 25: Reduced set of atomisation energies that were studied.

C_2H_2, C_2H_4, CH_3NH_2, CH_3OH, CH_4, CO, CO_2, H_2O, H_2O_2, LiH, N_2, N_2O, NH_3, NO_2

The a and ω parameters were varied and the b parameter is varied under the constraint that $a + b = 0.65$ as with CAM-B3LYP. The parameter a was varied from 0.04 to 0.40 in increments of 0.04 and the parameter ω was varied from 0.1 a_0^{-1} to 0.8 a_0^{-1} in increments of 0.1 a_0^{-1}. The mean absolute errors were plotted against these two parameters, and this can be seen in figure 5.

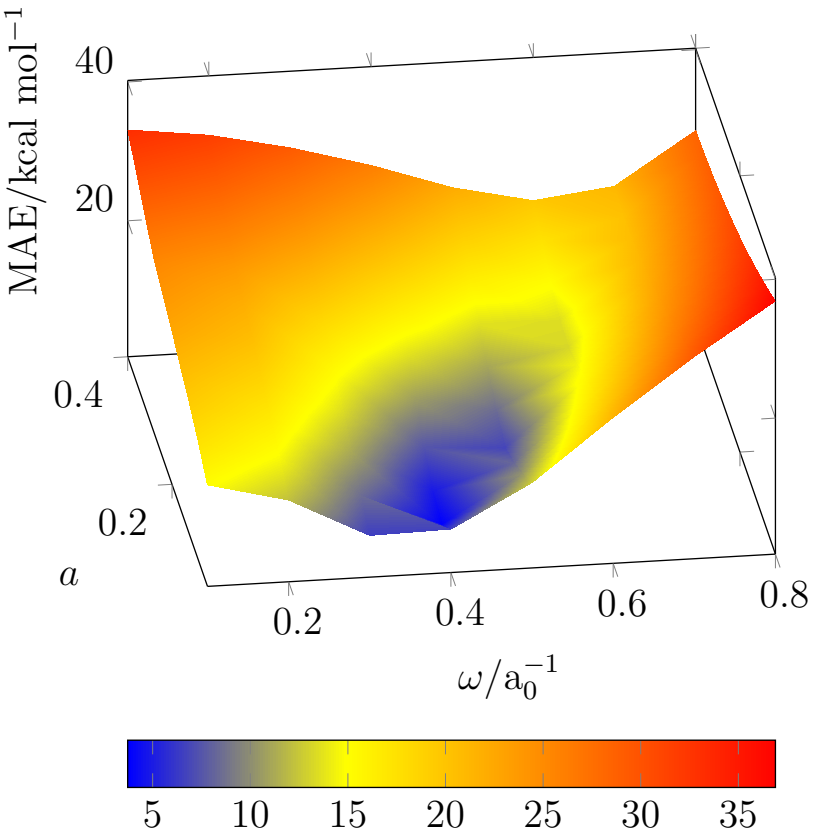


Figure 5: Plot of a and ω against mean absolute error.

The lowest error (of the parameters tested) occurs at $a = 0.08$ and $\omega = 0.4$ a_0^{-1}. These parameters are carried forward and evaluated for the full set of properties. The range-separated functional using these parameters is referred to as CAM-ωPBEa. The errors from this functional are compared to the CAM-ωPBE functional as well as LC-ωPBE and CAM-B3LYP. The errors are shown in table 26.

Table 26: Errors for CAM-ωPBEa.

Ionisation Potentials	CAM-B3LYP	LC-ωPBE	CAM-ωPBE	CAM-ωPBEa
d/eV	0·10	0·07	−0·57	−1·17
$\|d\|$/eV	0·18	0·20	0·57	1·16
σ/eV	0·21	0·25	0·32	0·24
Atomisation Energies				
d/kcal mol^{-1}	1·5	0·4	29·6	3·0
$\|d\|$/kcal mol^{-1}	3·2	3·2	31·4	5·4
σ/kcal mol^{-1}	3·7	4·1	25·7	6·5
Diatomic Bond Lengths				
d/Å	−0·006	−0·001	−0·032	0·016
$\|d\|$/Å	0·013	0·027	0·066	0·046
σ/Å	0·018	0·039	0·066	0·082
Classical Reaction Barriers				
d/kcal mol^{-1}	−2·5	0·0	−0·4	−1·7
$\|d\|$/kcal mol^{-1}	2·8	1·2	1·9	2·1
σ/kcal mol^{-1}	2·1	1·7	2·4	1·8
Vibrational Wavenumbers				
d/cm^{-1}	33	51	131	58
$\|d\|$/cm^{-1}	37	56	131	63
σ/cm^{-1}	47	61	117	71
Isotropic Polarisabilities				
d/au	0·24	−0·25	−1·13	0·27
$\|d\|$/au	0·37	0·35	1·13	0·44
σ/au	0·36	0·41	0·66	0·51
Electron Affinities				
d/eV	0·04	0·00	−0·95	−1·22
$\|d\|$/eV	0·10	0·13	0·95	1·22
σ/eV	0·13	0·15	0·22	0·17

It is noticeable that the calibrated functional, CAM-ωPBEa, has hugely improved errors over some properties (atomisation energies, diatomic bond lengths, vibrational frequencies and polarisabilities) compared to CAM-ωPBE. The error of the reaction barriers has remained somewhat similar between CAM-ωPBE and CAM-ωPBEa. The errors for ionisation potentials and electron affinities have increased between CAM-ωPBE and CAM-ωPBEa, this is contrary to what is expected due to the optimisation of the parameters. The energy of hydrogen was tested as part of the calibration. A lone hydrogen atom should have single point energy of -0.5 Ha. The results from a selected set of these calibration calculations are shown in table 27. This shows that there is a potential problem in varying parameters due to the range of hydrogen energies obtained. CAM-B3LYP has a hydrogen atom energy of -0.4988 Ha, the variance in the hydrogen atom energy

suggest that the parameters used may be fundamentally flawed.

Table 27: Hydrogen atom energies from the calibration of CAM-ωPBE, all values in Ha.

ω/a_0^{-1}	$a = 0.08$	$a = 0.16$	$a = 0.24$	$a = 0.32$	$a = 0.40$
0.1	−0.5018	−0.5230	−0.5449	−0.5674	−0.5906
0.2	−0.4822	−0.4993	−0.5170	−0.5353	−0.5542
0.3	−0.4667	−0.4802	−0.4942	−0.5088	−0.5240
0.4	−0.4544	−0.4651	−0.4762	−0.4877	−0.4998
0.5	−0.4448	−0.4532	−0.4620	−0.4711	−0.4806
0.6	−0.4371	−0.4439	−0.4509	−0.4582	−0.4657
0.7	−0.4311	−0.4365	−0.4422	−0.4480	−0.4541
0.8	−0.4263	−0.4307	−0.4353	−0.4401	−0.4450

The errors of CAM-ωPBEa are still larger than those for CAM-B3LYP and LC-ωPBE for atomisation energy which is confusing considering that the functional was optimised to a small set of atomisation energies, however the larger set contains different types of molecules which could lead to an elevated error. The errors for some of the other properties are larger (ionisation potentials, diatomic bond lengths, vibrational frequencies, electron affinities). The remaining properties (reaction barriers and polarisabilities) have similar errors to the established range-separated hybrids. This suggests that the functional CAM-ωPBEa has some potential but requires more work to fully optimise the parameters of the functional.

2.2.2 Singlet Excitations using CAM-ωPBE and CAM-ωPBEa

The two new range-separated hybrids are tested on the set of singlet excitations and the errors are compared to each other and to CAM-B3LYP and LC-ωPBE. The mean and mean absolute errors can be seen in table 28.

Table 28: CAM-ωPBE and CAM-ωPBEa errors for singlet excitations, all values are in eV.

Excitation Type		CAM-B3LYP	LC-ωPBE	CAM-ωPBE	CAM-ωPBEa
Local	d	0·02	0·14	0·10	0·02
	$\lvert d\rvert$	0·21	0·23	0·23	0·20
	σ	0·29	0·27	0·29	0·27
Rydberg	d	−0·48	−0·02	−0·46	−0·73
	$\lvert d\rvert$	0·48	0·14	0·47	0·81
	σ	0·18	0·17	0·27	0·65
Charge Transfer	d	−0·18	0·11	−0·11	−0·26
	$\lvert d\rvert$	0·27	0·68	0·22	0·32
	σ	0·31	0·81	0·28	0·35

The local excitations show similar errors between the two new range-separated hybrids. The mean error is 0.10 eV for CAM-ωPBE and 0.02 eV for CAM-ωPBEa. The mean absolute error is 0.23 eV for CAM-ωPBE and 0.20 eV for CAM-ωPBEa. The errors are also comparable to those for CAM-B3YLP and LC-ωPBE. This is as expected as discussed previously due to the short-range nature of the local excitations.

The Rydberg excitations show an increased error going from CAM-ωPBE to CAM-ωPBEa. The mean error increases from -0.46 eV to -0.73 eV, the mean absolute error increases from 0.47 eV to 0.81 eV. This suggests that while the CAM-ωPBEa hybrid is optimised for some properties, as stated above, it does not perform as well for excitation energies. This suggests that the functional will need to be further optimised. The errors are comparable to those of CAM-B3LYP and ωB97XD. However the errors are much larger than those for LC-ωPBE, which is the functional from which CAM-ωPBE originates. This again suggests that more optimisation of the new functional is required.

The charge transfer excitations show a slightly increased error going from CAM-ωPBE to CAM-ωPBEa. The mean error increases from -0.11 eV to -0.26 eV and the mean absolute error increases from 0.22 eV to 0.32 eV. However comparing these errors to those from LC-ωPBE, they are a large improvement suggesting the new functionals have some potential for improving the accuracy of charge transfer excitation energies. The errors of the two new functionals are comparable to those from CAM-B3LYP and ωB97XD.

2.2.3 Triplet Excitations using CAM-ωPBE and CAM-ωPBEa

The two new range-separated hybrids are tested on the set of triplet excitations using both TDDFT and TDA-TDDFT and the errors are compared to each other and to CAM-B3LYP and LC-ωPBE. The mean and mean absolute errors can be seen in table 29.

Table 29: Errors for CAM-ωPBE and CAM-ωPBEa triplet excitations.

	CAM-B3LYP		LC-ωPBE		CAM-ωPBE		CAM-ωPBEa	
	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA	TDDFT	TDA
d/eV	−0·39	−0·16	−0·48	−0·14	−0·74	−0·32	−0·54	−0·26
$\lvert d\rvert$/eV	0·40	0·18	0·53	0·22	0·77	0·34	0·54	0·26

As with all previous methods the application of the TDA-TDDFT method leads to a decrease in the errors for the triplet excitations compared to the TDDFT method. The CAM-ωPBE mean and mean absolute errors drop from -0.74 eV and 0.77 eV with TDDFT to -0.32 eV and 0.34 eV respectively. The CAM-ωPBEa

mean and mean absolute errors drop from −0.54 eV and 0.54 eV with TDDFT to −0.26 eV and 0.26 eV respectively.

The errors from CAM-ωPBEa are decreased to those from CAM-ωPBE for both methods, albeit with the errors for TDA-TDDFT having a much smaller decrease (−0.32 eV to −0.26 eV and 0.34 eV to 0.26 eV). This shows that the optimised parameters of the CAM-ωPBE functional shows an improvement to the accuracy of triplet excitation energies calculated.

The errors from CAM-ωPBE and CAM-ωPBEa are higher than those from CAM-B3LYP and LC-ωPBE, especially for the TDA-TDDFT case, which is the more accurate method. This shows that while the CAM-ωPBEa is an improvement on the CAM-ωPBE functional, it is still not as accurate as present functionals for triplet excitations. This again shows that CAM-ωPBE functional still requires more optimisation.

2.3 Functional Tuning

The issues with the non-transferability of the CAM-B3LYP parameters to CAM-ωPBE and CAM-ωPBEa may be due to some fundamental difference in the underlying GGA functional (PBE[31] and BLYP[32,33]). The standard global hybrids based on PBE and BLYP (PBE0[40] and B3LYP[38]) both use different percentages of HFx (20% and 25% respectively). PBE0 mixes PBE exchange and HF exchange, with 25% HF exchange and PBE correlation. The form of E_{xc} for PBE0 is

$$E_{\text{xc}}^{\text{PBE0}} = 0.25E_{\text{x}}^{\text{HF}} + 0.75E_{\text{x}}^{\text{PBE}} + E_{\text{c}}^{\text{PBE}}. \tag{81}$$

This functional performs well over a range of properties as seen in section 2.1. B3LYP mixes Becke 88 exchange, LDA exchange and HF exchange along with mixing Lee Yang Parr correlation and Vosko Wilk Nusair correlation. The form of E_{xc} for B3LYP is

$$E_{\text{xc}}^{\text{B3LYP}} = 0.08E_{\text{x}}^{\text{LDA}} + 0.72E_{\text{x}}^{\text{B88X}} + 0.2E_{\text{x}}^{\text{HF}} + 0.19E_{\text{c}}^{\text{VWN}} + 0.81E_{\text{c}}^{\text{LYP}} \tag{82}$$

As the two functionals have differing percentages of HFx, this suggests that the percentage of HFx should be optimised for each different underlying functional and hence the influence of HFx varies for different underlying functionals. The influence of HFx is investigated by optimising the percentage of HFx over a range of atomisation energies, much as several DFAs have been previously optimised[38,40]. The set of molecules used are the G-1 and G-2 sets of atomisation energies[92–94]. There are 148 molecules across these 2 sets. However initial optimisation was undertaken using a reduced set with a range of different molecule types. The

molecules in this reduced set are shown in table 30. This is an expanded set of molecules from those that CAM-ωPBE and CAM-ωPBEa were tuned with as it was noticed that the error of the atomisation energies of the cyclic molecules in the complete set of atomisation energies was unusally large for both CAM-ωPBE and CAM-ωPBEa, thus these cyclic molecules are included in this training set.

Table 30: Molecules in reduced set

C_2H_2, C_2H_4, CH_3NH_2, CH_3OH, CH_4, CO, CO_2
H_2O, H_2O_2, LiH, N_2, N_2O, NH_3, NO_2, SiF_4
pyridine, pyrrole, benzene, spiropentane, bicyclobutane, furan, thiophene
methylene cyclopropane, cyclobutane, cyclobutene

2.3.1 PBE - Reduced Set

The tuning of the PBE based global hybrid has been done by changing the a_0 value in

$$E_{\mathrm{xc}} = a_0 E_{\mathrm{x}}^{\mathrm{HF}} + (1 - a_0) E_{\mathrm{x}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{PBE}}. \tag{83}$$

The a_0 value is varied between 0 (purely PBE exchange, 0% HFx) and 1 (purely HF exchange, 100% HFx). The performance of these hybrids is assessed by the mean error and mean absolute error over the atomisation energies of the reduced set. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 6 and 7 respectively.

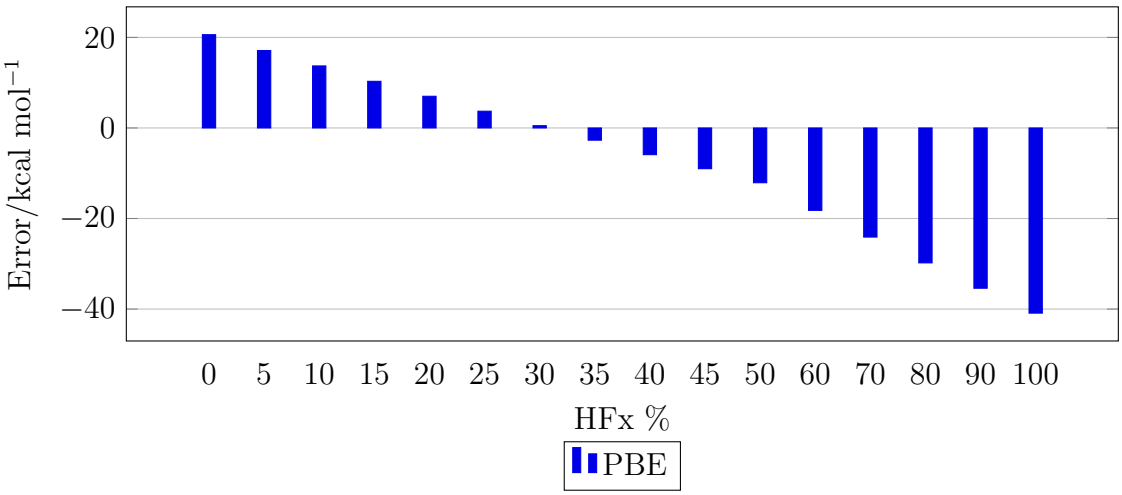


Figure 6: PBE reduced set mean error

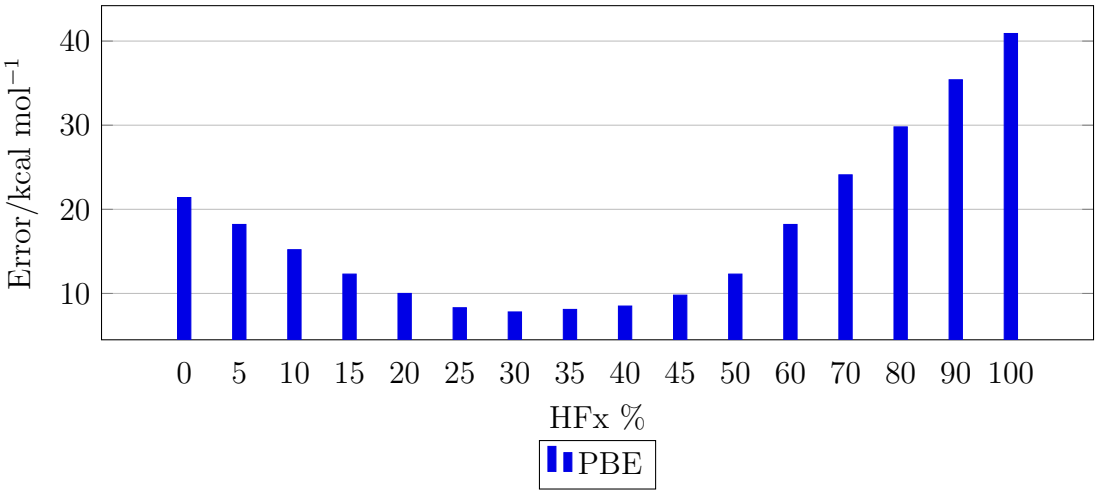


Figure 7: PBE reduced set mean absolute error

PBE0[40] is the standard global hybrid based on PBE, it has 25% HFx ($a_0 = 0.25$) and 75% PBE exchange. This is not the optimal percentage of HFx seen over the reduced set, the optimal percentage of HFx for the atomisation energies investigated here is 30% ($a_0 = 0.30$) due to having the lowest mean error (0.5 kcal mol^{-1}) and mean absolute error (7.8 kcal mol^{-1}). This suggests that the performance of PBE0 for atomisation energies may be improved upon including a higher percentage of HFx.

However it was noticeable from the data on an individual molecule basis that there were some cases where PBE required much more HFx to have optimal performance (e.g. pyridine with optimal percentage between 40–45%, compared with example CH_3OH with an optimal percentage of 20%). This suggests that it may be worth splitting the reduced set based on molecule type, i.e. cyclic molecules (like pyridine) vs. non-cyclic molecules (like CH_3OH). The split reduced set is shown in table 31.

Table 31: Split reduced set

Non-cyclic	C_2H_2, C_2H_4, CH_3NH_2, CH_3OH, CH_4, CO, CO_2, H_2O, H_2O_2, LiH, N_2, N_2O, NH_3, NO_2, SiF_4
Cyclic	pyridine, pyrrole, benzene, spiropentane, bicyclobutane, furan, thiophene, methylene cyclopropane, cyclobutane, cyclobutene

Due to the differing behaviour of PBE based global hybrids for cyclic and non-cyclic molecules, there may be a need to categorise the molecules and systems in benchmark sets. If this is not done and we look over all the benchmark data, there is a potential to miss important behaviour when developing new approximate E_{xc}s. There seems to be a need to balance the categories/subsets of molecules and systems in training sets in order to remove potential bias towards certain behaviours. The use of molecules that combine elements of both sets may be

a potential method to access these behaviours in a more unbiased way. When considering small molecules however it is not possible to get both cyclic and non-cyclic behaviour in one molecule.

Adding more systems into training sets will necessarily increase the diversity of the training set. However, this may not be in a balanced way, leading to biases towards certain behaviour if behaviour such as that seen for the cyclic and non-cyclic molecules is not taken into account and properly factored into selection of training sets.

The performance of the hybrids with varying the percentage of HFx are assessed again by the mean error and mean absolute error of the split reduced set of atomisation energies. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 8 and 9 respectively.

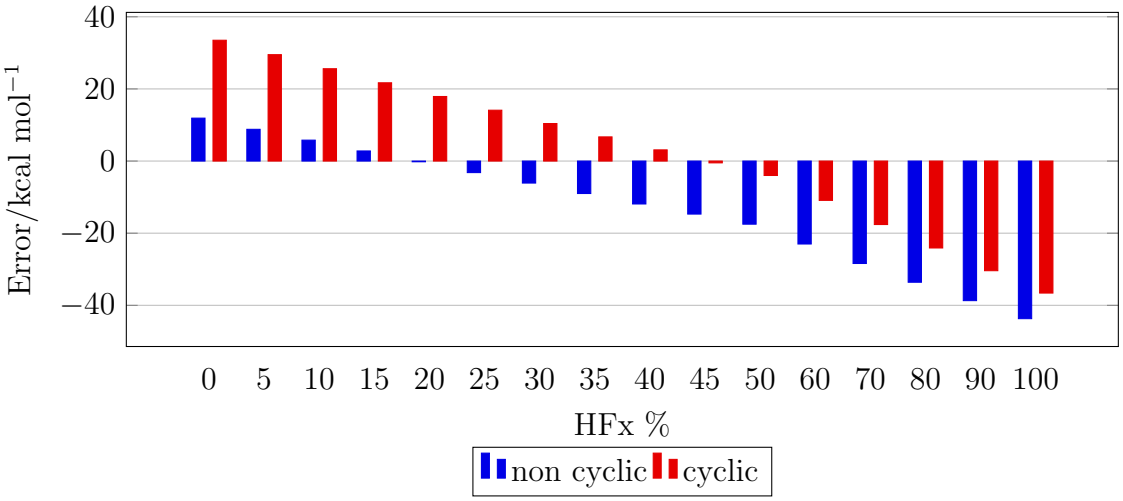


Figure 8: PBE split reduced set mean error

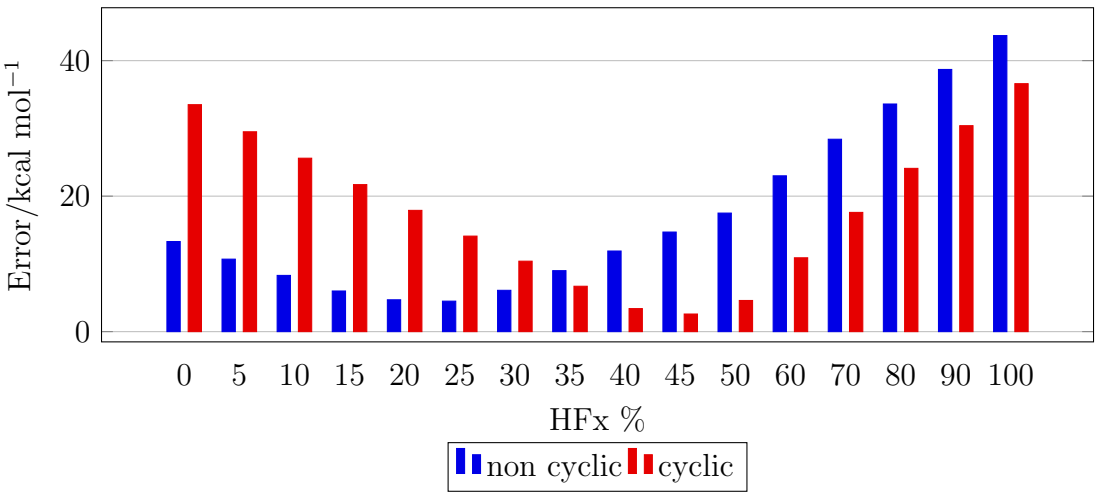


Figure 9: PBE split reduced set mean absolute error

There is a distinct difference between the optimal percentages of HFx required for optimal performance for the 2 parts of the split set. The non-cyclic molecules require a much smaller percentage of HFx for optimal performance compared to the cyclic molecules (20% HFx for non-cyclic molecules and 45% HFx for cyclic molecules). This is suggestive of the need for tweaking a global hybrid of PBE when calculating properties of cyclic molecules, due to the optimal percentage of HFx for cyclic molecules being 20% higher than the percentage used in PBE0.

2.3.2 BLYP - Reduced Set

The general expression of B3LYP based global hybrids is more complicated than that for PBE based global hybrids, due to the 3 parameter fitting of B3LYP. The tuning of the B3LYP based global hybrids can be achieved by changing the a_0 value in this general expression for E_{xc}

$$E_{\text{xc}} = 0.1(1-a_0)E_{\text{x}}^{\text{LDA}} + 0.9(1-a_0)E_{\text{x}}^{\text{B88X}} + a_0 E_{\text{x}}^{\text{HF}} + 0.19E_{\text{c}}^{\text{VWN}} + 0.81E_{\text{c}}^{\text{LYP}}. \tag{84}$$

The a_0 value is varied between 0 (10% LDA exchange and 90% Becke 88 exchange, 0% HFx) and 1 (purely HF exchange, 100% HFx). The performance of these hybrids are assessed by the mean error and mean absolute error over the atomisation energies of the reduced set. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 2.3.2 and 11 respectively.

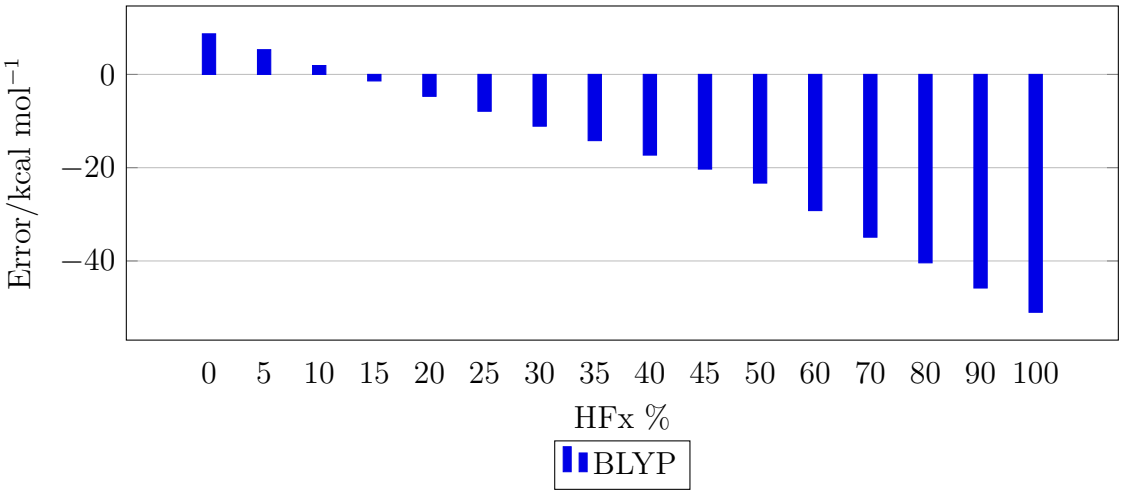


Figure 10: BLYP reduced set mean error

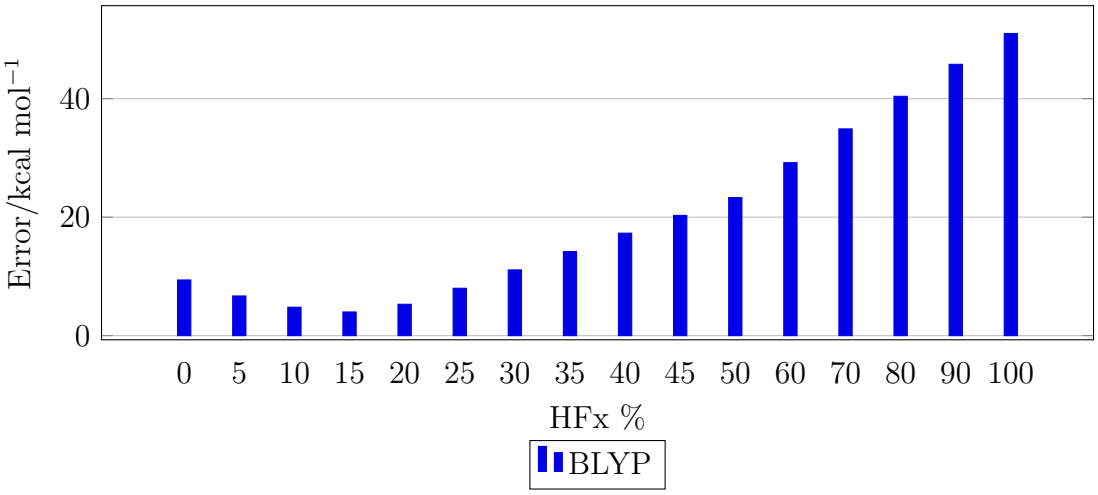


Figure 11: BLYP reduced set mean absolute error

B3LYP[38] is the standard global hybrid based on BLYP, it has 20% HFx ($a_0 = 0.20$), 8% LDA exchange and 72% Becke 88 exchange. As with PBE, the standard parameters are not optimal for the reduced set, the optimal percentage of HFx for the atomisation energies investigated here is 15% ($a_0 = 0.15$) due to having the lowest mean error (-1.4 kcal mol^{-1}) and mean absolute error (4.0 kcal mol^{-1}). This suggests that B3LYP may be improved for atomisation energies by including a lower percentage of HFx.

The split set (cyclic and non-cyclic molecules) has been investigated again, plots of the mean and mean absolute errors for the atomisation energies are shown in figures 12 and 13.

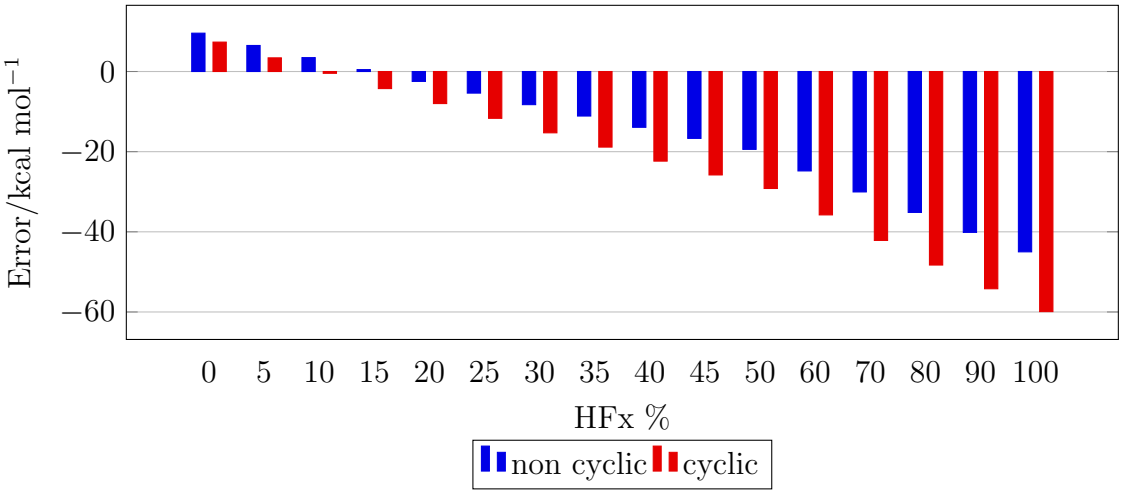


Figure 12: BLYP split reduced set mean error

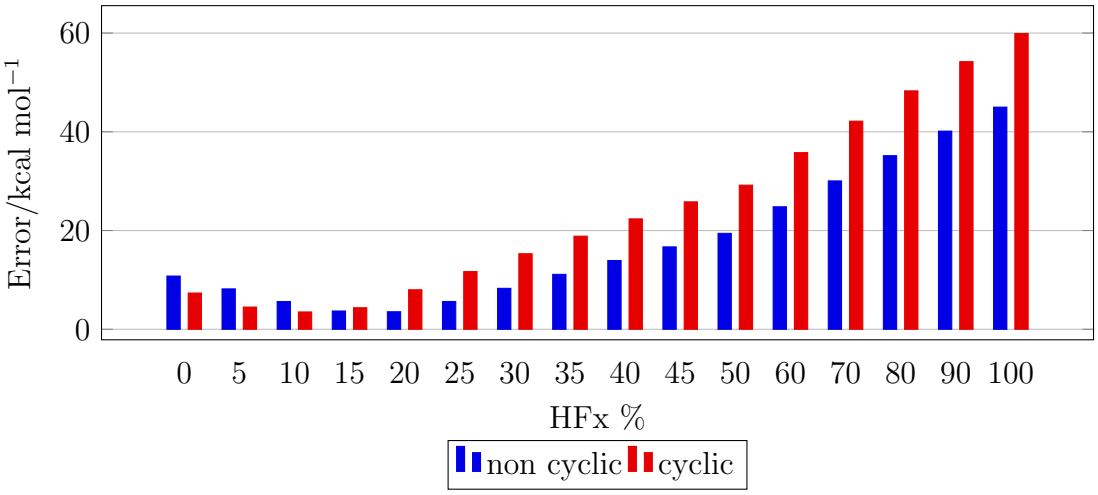


Figure 13: BLYP split reduced set mean absolute error

There is not as distinct a difference between the cyclic and non-cyclic molecules for the optimal percentage of HFx when using BLYP compared to PBE. The optimal HFx percentage for cyclic molecules is 10% and for non-cyclic molecules is 15%, this is only a 5% difference compared to the 25% difference seen in PBE. This suggests that there should not be the need for tweaking the global hybrid of BLYP for calculating cyclic molecules. However it should be noted that there is a larger error for cyclic molecules when using the HFx percentage of B3LYP (20%), with a mean absolute error of 8.0 kcal mol^{-1} compared to 3.5 kcal mol^{-1} for the optimal HFx percentage (10%).

On Perdew's Jacob's Ladder of density functional approximations[24], PBE and BLYP are both generalised gradient approximations (GGAs) and hence are on the second rung of the ladder. Hybrid functionals are on the fourth rung of the ladder[24]. It is possible to generate a global hybrid using functionals on the first and third rung to compare to the global hybrids from the two GGAs. The first rung is the local density approximations (LDA[20,21,95]) and the third rung are meta-GGA functionals, such as TPSS[35].

2.3.3 LDA - reduced set

The global hybrid of LDA was generated in a similar way to the PBE global hybrid. The global hybrids of LDA have the general form

$$E_{\mathrm{xc}} = a_0 E_{\mathrm{x}}^{\mathrm{HF}} + (1 - a_0) E_{\mathrm{x}}^{\mathrm{LDA}} + E_{\mathrm{c}}^{\mathrm{VWN5}}. \tag{85}$$

The value of a_0 is varied between 0 (purely LDA exchange, 0%HFx) and 1 (purely HF exchange, 100% HFx). The performance of these hybrids are assessed by the mean error and mean absolute error over the atomisation energies of the reduced set. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 14 and 15 respectively.

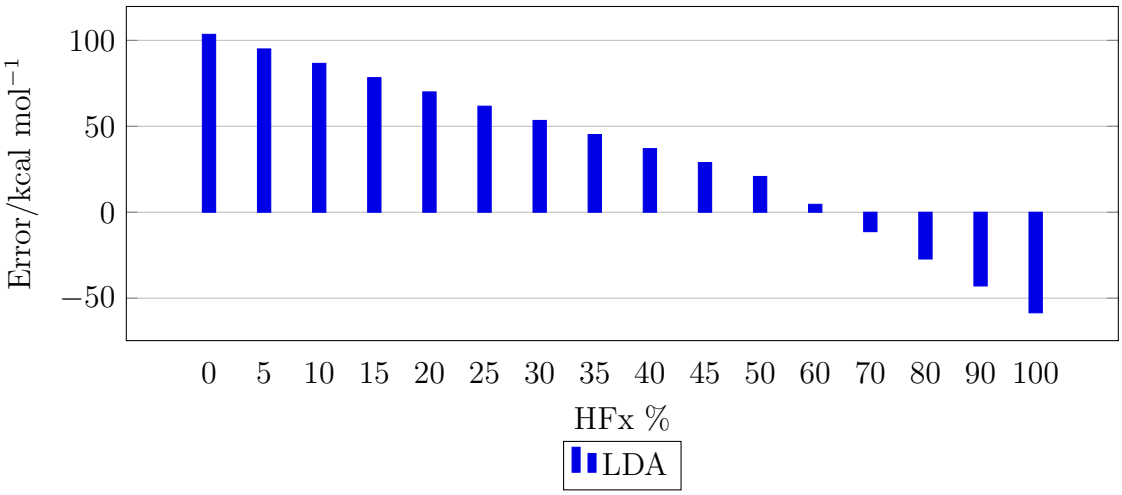


Figure 14: LDA reduced set mean error

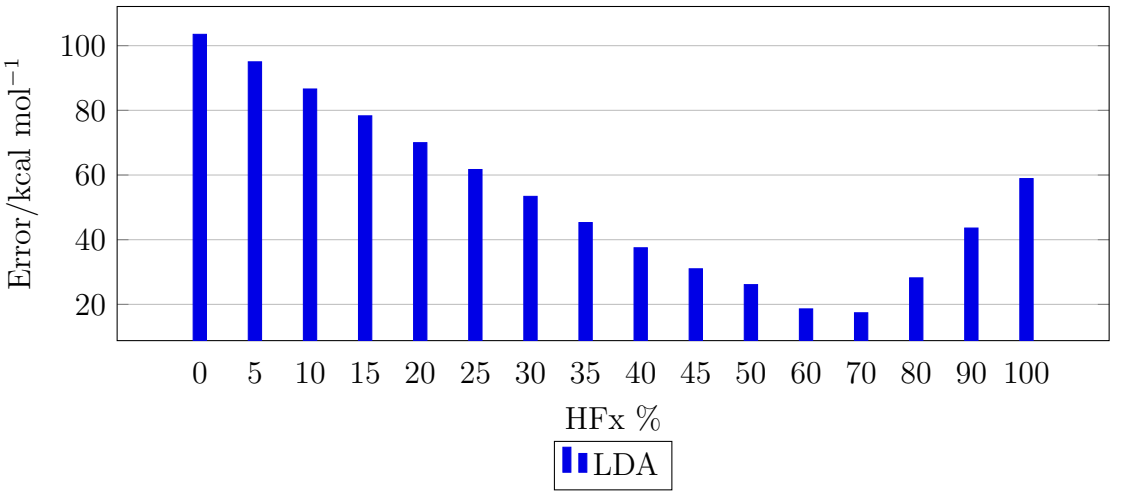


Figure 15: LDA reduced set mean absolute error

The optimal global hybrid of LDA contains 70% HFx. ($a_0 = 0.70$). This is

significantly higher than the optimal percentages for the two GGA based hybrids, PBE (30%, $a_0 = 0.30$) and BLYP (15%, $a_0 = 0.15$). It is also noticable that the error of the pure DFT exchange ($a_0 = 0$, 0% HFx) is much higher for LDA (mean error: 103.5 kcal mol^{-1} and mean absolute error: 103.5 kcal mol^{-1}) compared to PBE (mean error: 20.6 kcal mol^{-1} and mean absolute error: 21.4 kcal mol^{-1}) and BLYP (mean error: 8.7 kcal mol^{-1} and mean absolute error: 9.4 kcal mol^{-1}). This follows from Perdew's Jacob's Ladder as the GGAs are on the second rung so should perform better than LDA.

The split set (cyclic and non-cyclic molecules) has been investigated again, plots of the mean and mean absolute errors for the atomisation energies are shown in figures 16 and 17.

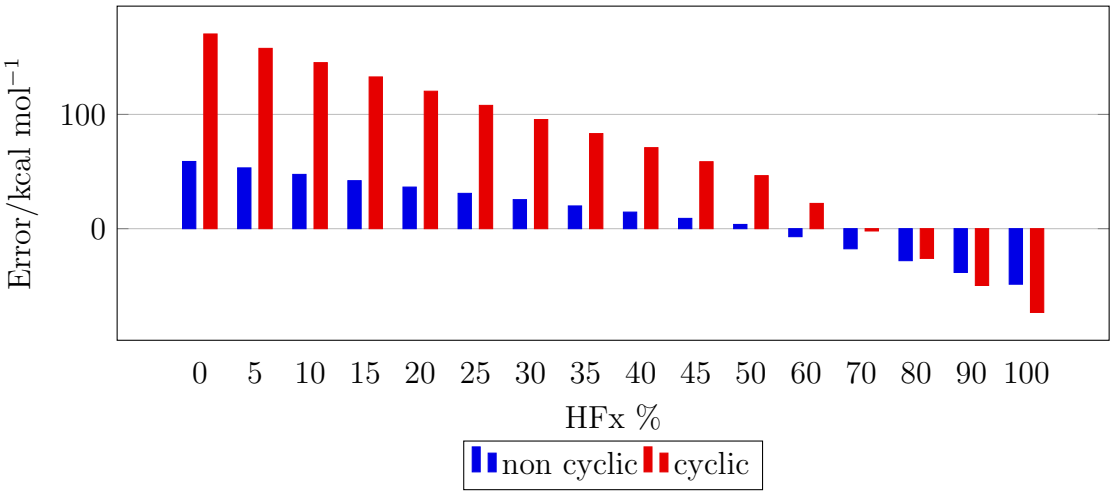


Figure 16: LDA split reduced set mean error

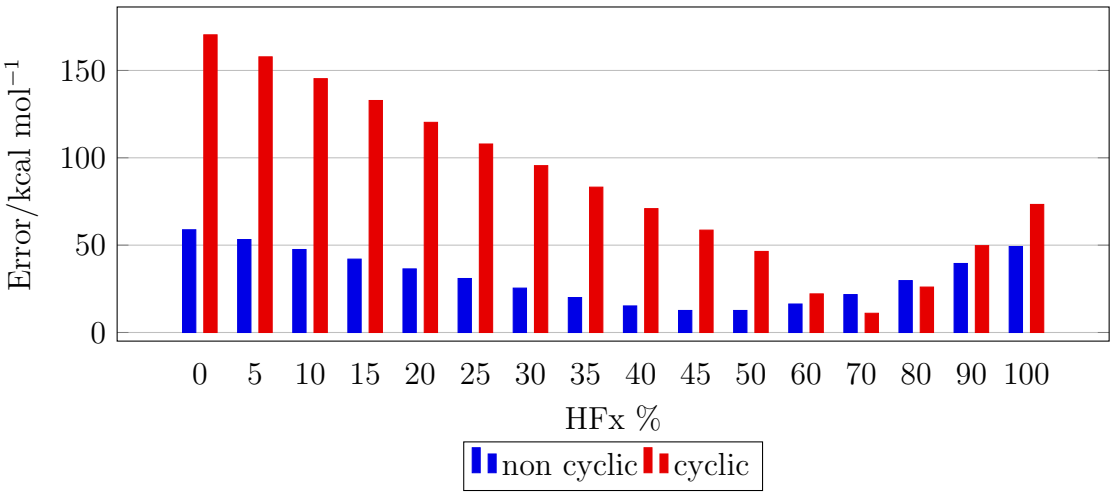


Figure 17: LDA split reduced set mean absolute error

There is a distinct difference between the optimal percentages of HFx required

for optimal performance for the 2 parts of the split set. The non-cyclic molecules require a much smaller percentage of HFx for optimal performance compared to the cyclic molecules (45% HFx for non-cyclic molecules and 70% HFx for cyclic molecules). As with PBE, this is suggestive of the need for tweaking a global hybrid of LDA when calculating properties of cyclic molecules, due to the optimal percentage of HFx for cyclic molecules being 25% higher than the optimal percentage of HFx for non-cyclic molecules. It is also noticable that the cyclic molecule errors dominate the mean errors of the set when the error of the LDA global hybrid is large e.g. at 0% HFx mean error for cyclic molecules is 170.4 kcal mol^{-1} and for non-cyclic molecules is 58.8 kcal mol^{-1}.

2.3.4 TPSS - reduced set

TPSS is a meta-GGA which is on the third rung of Perdew's Jacob's Ladder. This means it is expected to perform better than both LDA and GGAs. The tuning of the TPSS based global hybrid has been done by changing the a_0 value in

$$E_{\text{xc}} = a_0 E_{\text{x}}^{\text{HF}} + (1 - a_0) E_{\text{x}}^{\text{TPSS}} + E_{\text{c}}^{\text{TPSS}}. \tag{86}$$

The a_0 value is varied between 0 (purely PBE exchange, 0% HFx) and 1 (purely HF exchange, 100% HFx). The performance of these hybrids are assessed by the mean error and mean absolute error over the atomisation energies of the reduced set. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 18 and 19 respectively.

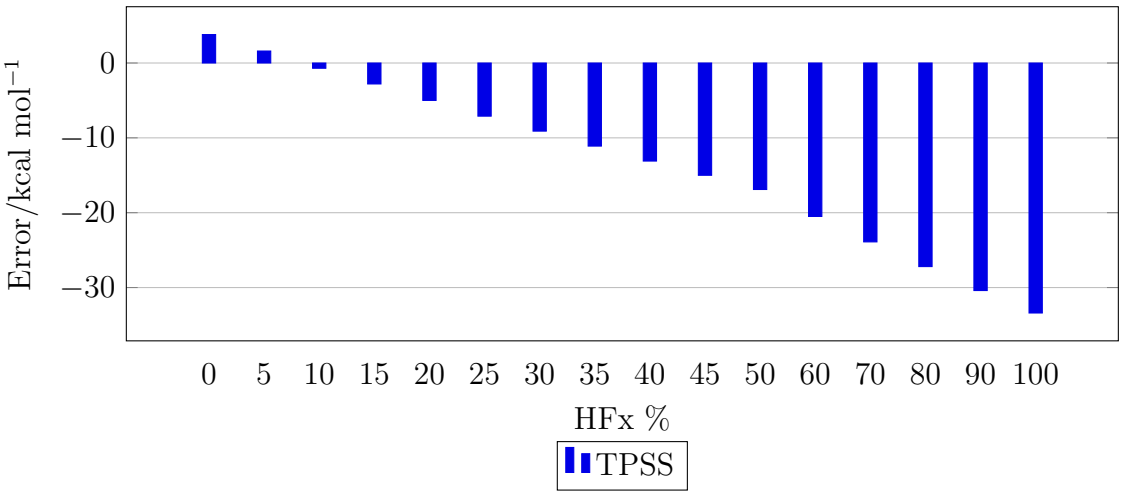


Figure 18: TPSS reduced set mean error

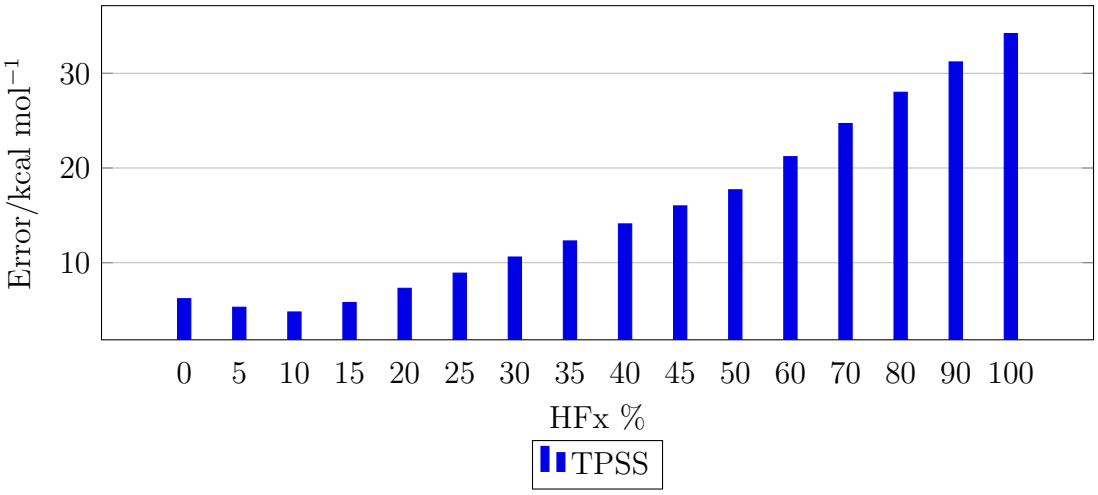


Figure 19: TPSS reduced set mean absolute error

TPSSh[35,96] is the standard global hybrid based on TPSS, it has 10% HFx ($a_0 = 0.10$), 90% TPSS exchange. This is the optimal percentage of HFx seen over the reduced set, having the lowest mean error (-0.7 kcal mol^{-1}) and mean absolute error (4.8 kcal mol^{-1}). This suggests that TPSSh contains the optimal percentage of HFx.

It is also noticable that the error of the pure DFT exchange case ($a_0 = 0$, 0% HFx) is lower for TPSS (mean error: 3.8 kcal mol^{-1} and mean absolute error: 6.2 kcal mol^{-1}) compared to LDA (mean error: 103.5 kcal mol^{-1} and mean absolute error: 103.5 kcal mol^{-1}), PBE (mean error: 20.6 kcal mol^{-1} and mean absolute error: 21.4 kcal mol^{-1}) and BLYP (mean error: 8.7 kcal mol^{-1} and mean absolute error: 9.4 kcal mol^{-1}). This follows Perdew's Jacob's Ladder as TPSS is a meta-GGA which is on the third rung whereas GGAs are on the second rung and LDA is on the first rung. The errors from the pure DFT follows Perdew's Jacob's Ladder as the error decreases as the rungs are ascended (LDA > GGA > meta-GGA). The split set (cyclic and non-cyclic molecules) has been investigated again; plots of the mean and mean absolute errors for the atomisation energies are shown in figures 20 and 21.

Figure 20: TPSS split reduced set mean error

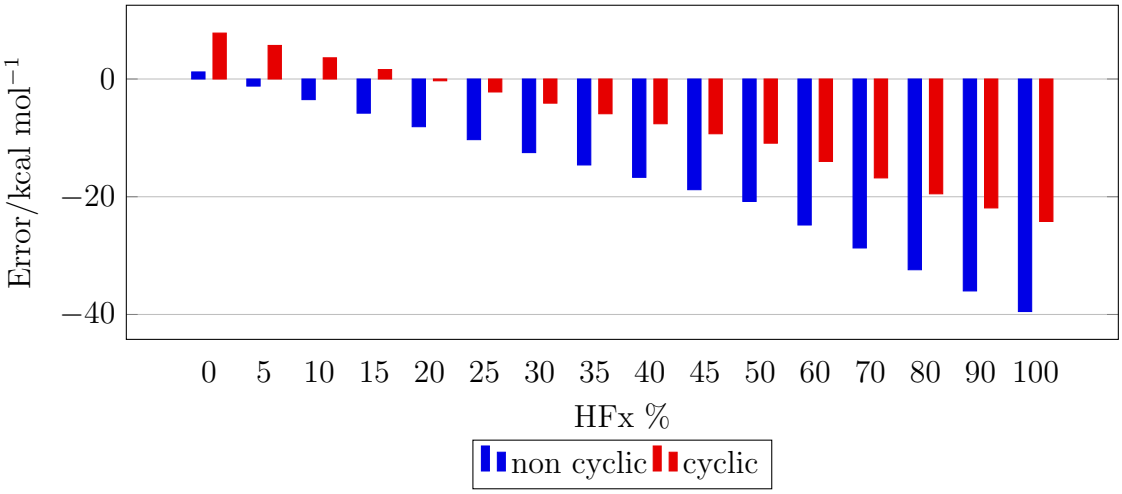
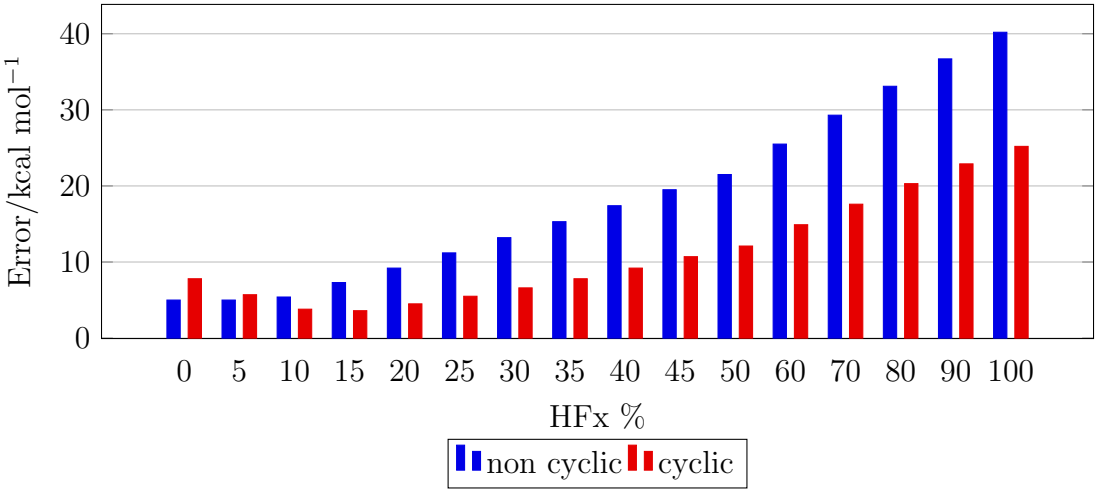


Figure 21: TPSS split reduced set mean absolute error



There is a difference between the cyclic and non-cyclic molecules for the optimal percentage of HFx when using TPSS. The optimal HFx percentage for cyclic molecules is 15% and for non-cyclic molecules is 5%, this is only a 10% difference. This suggests that there should not be the need for tweaking the global hybrid of TPSS for calculating cyclic molecules as the errors over both sets of molecules is reduced at 10% HFx, which is the optimal percentage of HFx for a TPSS global hybrid[35,96].

2.3.5 Functional Dependence on Hartree–Fock Exchange

The mean errors for all four of the functionals hybridised are shown in table 32. This shows a differing dependence on the percentage of HFx required to perform well/how quickly performance deteriorates away from the optimal global hybrid.

Table 32: Mean error of all four functionals hybridised for the atomisation energies of the reduced set, all values are in kcal mol^{-1}.

Hartree–Fock Exchange	LDA	PBE	BLYP	TPSS
0	103·5	20·6	8·7	3·8
5	95·0	17·1	5·3	1·6
10	86·6	13·7	1·9	−0·7
15	78·3	10·3	−1·4	−2·8
20	70·0	7·0	−4·7	−5·0
25	61·7	3·7	−7·9	−7·1
30	53·4	0·5	−11·1	−9·1
35	45·2	−2·7	−14·2	−11·1
40	37·0	−5·9	−17·3	−13·1
45	28·9	−9·0	−20·3	−15·0
50	20·8	−12·1	−23·3	−16·9
60	4·6	−18·2	−29·2	20·5
70	−11·3	−24·1	−34·9	23·9
80	−27·2	−29·9	−40·4	−27·2
90	−42·9	−35·4	−45·8	−30·4
100	−58·5	−40·9	−51·0	−33·4

The functionals based on those from higher rungs of the Jacob's Ladder perform better away from the optimal global than those on the lower rungs. The lower rungs have much more dependence on the percentage of HFx included. LDA has the greatest dependence as it performs poorly away from the optimal percentage and TPSS has the least dependence as it performs well (mean error between -7.5 and 7.5 kcal mol^{-1}) at the greatest range away from the optimal percentage. The range of HFx percentages that the global hybrids perform well with (mean error between -7.5 and 7.5 kcal mol^{-1}) are shown in table 33.

Table 33: Range of HFx percentages which the global hybrids perform well

Functional	HFx range
LDA	60%
PBE	20 – 40%
BLYP	5 – 20%
TPSS	0 – 25%

The LDA based global hybrid has the smallest range, only performing well at the optimal percentage and poorly at any other percentage investigated. The GGA based global hybrids, PBE and BLYP, have similar ranges with 20% and 15% respectively. The meta-GGA based global, TPSS, has the largest range (25%). This trend follows the Jacob's Ladder well as the dependence of the performance on the percentage of HFx is lowered as the rungs are ascended.

The ranges of the mean errors over the range of percentages investigated is

another trend that follows the Jacob's Ladder, with LDA based global hybrids having the largest range of mean errors; the two GGA based global hybrids having similar ranges and the meta-GGA based global hybrid having the lowest range. The range of mean errors for each functional's global hybrids is shown in table 34.

Table 34: Range of mean errors for each functional's global hybrids, all values are in kcal mol^{-1}.

Functional	Range/kcal mol^{-1}
LDA	161.9
PBE	61.5
BLYP	59.6
TPSS	37.2

2.3.6 Complete G-1 and G-2 Atomisation Energies Sets

To ensure that the optimal parameters are not biased by the choice of reduced set; the accuracy of the PBE and BLYP based global hybrids are tested over atomisation energies of the full set of molecules in the G-1 and G-2 sets of atomisation energies[92–94]. The performance of the hybrids over the complete G-1 and G-2 sets of atomisation energies are again assessed by the mean errors and mean absolute errors. Plots of the mean and mean absolute errors against the percentage of HFx are shown in figures 22 and 23 respectively.

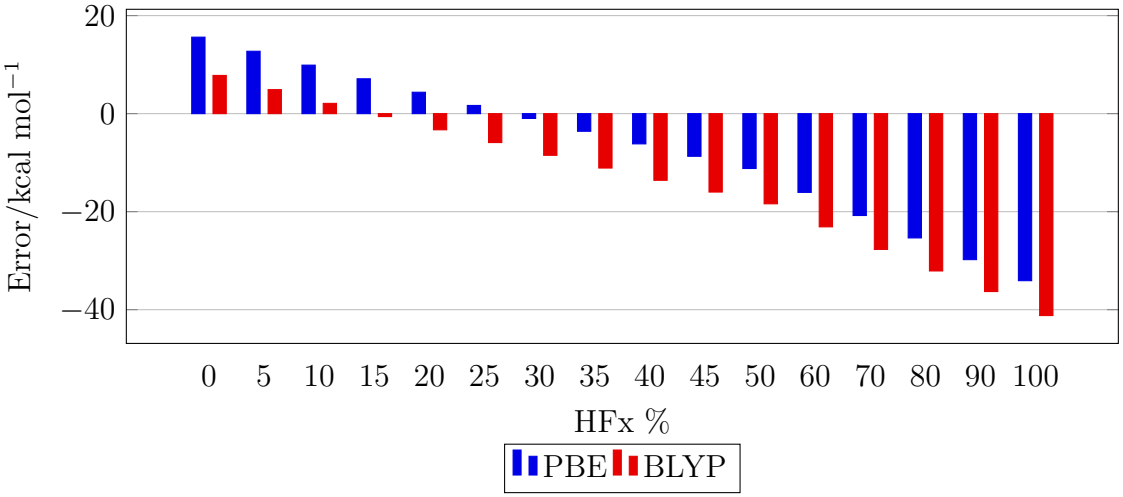


Figure 22: Mean error of the complete G-1 and G-2 sets of atomisation energies using PBE and BLYP

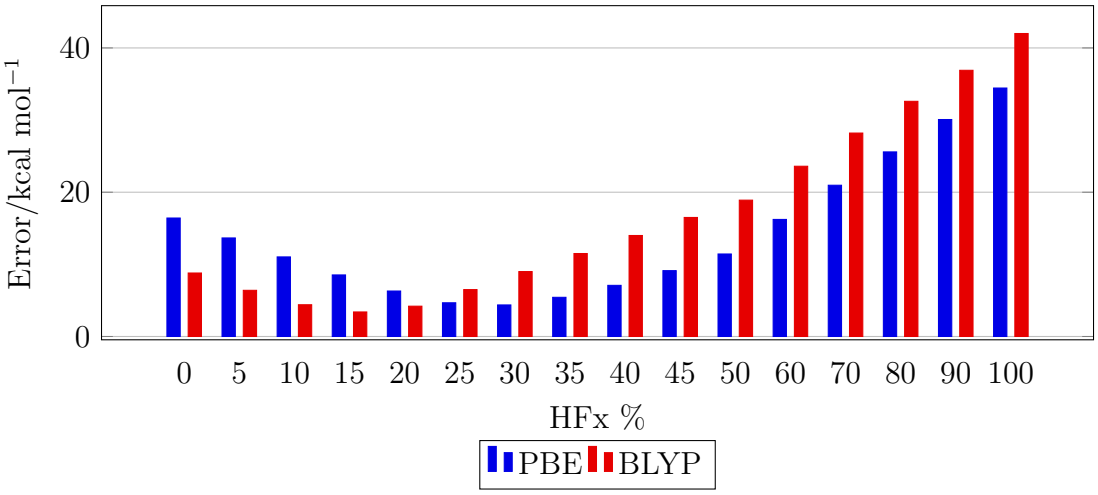


Figure 23: Mean absolute error of the complete G-1 and G-2 sets of atomisation energies using PBE and BLYP

The optimal global hybrid of PBE from the reduced set had 30% of HFx and this is still the optimal percentage of HFx for the complete set, with a mean error of -1.0 kca mol^{-1} and a mean absolute error of 4.4 kcal mol^{-1}. The optimal global hybrid of BLYP from the reduced set had 15% of HFx and this is still the optimal percentage of HFx for the complete set, with a mean error of -0.6 kcal mol^{-1} and a mean absolute error of 3.4 kcal mol^{-1}. As the optimal percentage of HFx remains constant for both the reduced set and the complete sets, the reduced set seems to give a representative picture of the influence of HFx on the accuracy of a global hybrid.

2.4 Revisiting CAM-ωPBE

The optimal global hybrids of BLYP and PBE have a 15% difference in the percentage of HFx included (15% for BLYP and 30% for PBE). The initial attempt in generating a CAM type range separated PBE was based around the parameters that were optimised for a BLYP based function ($a = 0.19$, $a + b = 0.65$ and $\omega = 0.33$ a$_0^{-1}$). The performance of both CAM-ωPBE and CAM-ωPBEa is poor over the range of properties as demonstrated in section 2.2. The issue with the performance of these functionals is possibly due to the fundamental difference between the influence of HFx on the performance on the underlying functional as seen in previous section (PBE requires more HFx to reduce the errors observed in atomisation energies), it seems that a different set of parameters should be used for a CAM type PBE based functional to reflect this.

2.4.1 Reoptimising CAM-ωPBE Parameters

As previously discussed, there are 3 parameters that can be optimised in a CAM type range separated functional, a (full range HFx), $a + b$ (long range HFx) and ω (scaling of HFx). Initially only the a and ω parameters were optimised and it seems that the $a + b$ parameter may be too low due to the requirement for more HFx in PBE based global hybrids. The parameters of a, $a + b$ and ω are optimised again over the atomisation energies of the set of molecules shown in table 35.

Table 35: Reduced set of atomisation energies that were studied.

C_2H_2, C_2H_4, CH_3NH_2, CH_3OH, CH_4, CO, CO_2, H_2O, H_2O_2, LiH, N_2, N_2O, NH_3, NO_2, pyridine

The mean and mean absolute errors of the parameters tested are shown in table 36 and are compared to the mean and mean absolute errors of CAM-ωPBE, CAM-ωPBEa and LC-ωPBE.

Table 36: Errors of parameters tested to optimise CAM-PBE

a	$a + b$	ω/a$_0^{-1}$	d/kcal mol^{-1}	$\lvert d\rvert$/kcal mol^{-1}
0·19	0·65	0·33	15·3	18·0
0·08	0·65	0·40	0·1	4·7
0·00	1·00	0·40	−1·5	3·5
0·00	0·70	0·35	−1·9	7·8
0·00	0·80	0·35	−0·6	6·3
0·00	0·90	0·35	0·6	4·9
0·05	0·90	0·20	15·5	16·1
0·05	0·90	0·30	9·1	9·7
0·05	0·90	0·40	1·3	4·4
0·05	0·80	0·40	−0·2	3·9
0·05	0·80	0·50	−9·0	9·0
0·05	0·80	0·60	−17·3	17·3
0·05	0·80	0·39	0·7	4·0
0·05	0·80	0·41	−1·1	4·2
0·04	0·80	0·40	0·7	4·2
0·06	0·80	0·40	−1·1	4·3

The parameters were initially optimised at differing $a + b$ values while using the a value from LC-ωPBE (0.00) and an ω value between the two (0.35 a$_0^{-1}$). The errors from these functionals showed that having a higher $a + b$ parameter gave a lower error. The changing of the a (0.05) and ω (0.20, 0.30 and 0.40) values using $a + b = 0.90$ gave the lowest error at $\omega = 0.40$, this suggests that a higher scaling parameter gives better performance. A set of functionals with $a + b = 0.80$ using $a = 0.05$ and varying ω values (0.40 a$_0^{-1}$, 0.50 a$_0^{-1}$, 0.60 a$_0^{-1}$) were then tested. The best performance out of the parameters tested so far ($a = 0.05$, $a + b = 0.80$ and $\omega = 0.40$ a$_0^{-1}$) gave a mean error of -0.2 kcal mol^{-1} and a mean absolute error

of 3.9 kcal mol^{-1}. Parameters were then tested around these values by changing a and ω by ± 0.01, however the performance of these paramters was worse than $a = 0.05$, $a + b = 0.80$ and $\omega = 0.40$ a$_0^{-1}$.

This set of parameters has a similar performance to LC-ωPBE over the reduced set of atomisations that were studied here. This set of parameters will be referred to CAM-PBE. CAM-PBE was tested over the full set of properties investigated in section _. The mean and mean absolute errors of CAM-PBE are compared to CAM-B3LYP, LC-ωPBE and CAM-ωPBEa. The ground state properties are shown in table 37 and the excited state properties are shown in table 38.

Table 37: Errors for CAM-PBE ground state properties.

Ionisation Potentials	CAM-B3LYP	LC-ωPBE	CAM-ωPBEa	CAM-PBE
d/eV	0·10	0·07	−1·17	−0·62
$\|d\|$/eV	0·18	0·20	1·16	0·62
σ/eV	0·21	0·25	0·24	0·23
Atomisation Energies				
d/kcal mol^{-1}	1·5	0·4	3·0	2·7
$\|d\|$/kcal mol^{-1}	3·2	3·2	5·4	4·9
σ/kcal mol^{-1}	3·7	4·1	6·5	5·7
Diatomic Bond Lengths				
d/Å	−0·006	−0·001	0·016	0·006
$\|d\|$/Å	0·013	0·027	0·046	0·037
σ/Å	0·018	0·039	0·082	0·060
Classical Reaction Barriers				
d/kcal mol^{-1}	−2·5	0·0	−1·7	−0·9
$\|d\|$/kcal mol^{-1}	2·8	1·2	2·1	1·5
σ/kcal mol^{-1}	2·1	1·7	1·8	1·7
Vibrational Wavenumbers				
d/cm^{-1}	33	51	58	58
$\|d\|$/cm^{-1}	37	56	63	60
σ/cm^{-1}	47	61	71	68
Isotropic Polarisabilities				
d/au	0·24	−0·25	0·27	−0·02
$\|d\|$/au	0·37	0·35	0·44	0·29
σ/au	0·36	0·41	0·51	0·41
Electron Affinities				
d/eV	0·04	0·00	−1·22	−0·70
$\|d\|$/eV	0·10	0·13	1·22	0·74
σ/eV	0·13	0·15	0·17	0·15

Table 38: Errors for CAM-PBE excited state properties.

Singlet–Local	CAM-B3LYP	LC-ωPBE	CAM-ωPBEa	CAM-PBE
d/eV	0·02	0·14	0·02	0·10
$\lvert d\rvert$/eV	0·21	0·23	0·20	0·21
σ/eV	0·29	0·27	0·27	0·27
Singlet–Rydberg				
d/eV	−0·48	−0·02	−0·73	−0·51
$\lvert d\rvert$/eV	0·48	0·14	0·81	0·57
σ/eV	0·18	0·17	0·65	0·38
Singlet–Charge Transfer				
d/eV	−0·18	0·11	−0·26	0·10
$\lvert d\rvert$/eV	0·27	0·68	0·32	0·28
σ/eV	0·31	0·81	0·35	0·34
Triplet–TDDFT				
d/eV	−0·39	−0·48	−0·54	−0·52
$\lvert d\rvert$/eV	0·40	0·53	0·54	0·53
Triplet–TDA-TDDFT				
d/eV	−0·16	−0·14	−0·26	−0·21
$\lvert d\rvert$/eV	0·18	0·22	0·26	0·24

CAM-PBE performs better than CAM-ωPBEa on all ground state properties and similarly to CAM-B3LYP and LC-ωPBE in most properties. It is noticable that there is a improved performance in isotropic polarisabilities, with the lowest mean absolute error seen of the functionals tested. As isotropic polarisabilities are related to changes in electron density it may follow that CAM-PBE performs well for excited state properties. For the singlet excited states studied, CAM-PBE performs similiary to CAM-B3LYP across the different types of excitations (Local, Rydberg and Charge Transfer). CAM-PBE shows a better performance than LC-ωPBE for charge transfer type singlet excitations but a worse performance for Rydberg type singlet excitations. The performance of CAM-PBE for triplet excitations is similar to the other functionals tested, using TDA-TDDFT has a much larger effect on the performance of the triplet excitations than the choice of functional. However the performance of CAM-PBE is still lacking in ionisation potentials and electron affinities, with a much worse performance than both LC-ωPBE and CAM-B3LYP on these two properties.

The parameters optimised (a, $a+b$ and ω) over atomisation energies will be optimised over the electron affinities shown in table 39, the set used for functional benchmarking, to check if correcting the performance of this property fixes the issue seen with CAM-PBE without breaking the performance of the other properties.

Table 39: Set of electron affinities that were studied.

C, O, F, Si, P, S, Cl, CH, CH_2, CH_3, NH, NH_2, OH, SiH, SiH_2
SiH_3, PH, PH_2, SH, O_2, NO, CN, PO, S_2, Cl_2

The mean and mean absolute errors of the parameters tested are shown in table 40 and are compared to the mean and mean absolute errors of CAM-PBE and LC-ωPBE.

Table 40: Errors of parameters tested to optimise CAM-PBE electron affinities

a	$a+b$	ω	d/eV	$\lvert d\rvert$/eV
0·05	0·80	0·40	−0·70	0·74
0·00	1·00	0·40	0·00	0·10
0·10	0·80	0·40	−0·67	0·71
0·10	0·80	0·30	−0·48	0·55
0·10	0·80	0·20	−0·24	0·33
0·05	0·80	0·20	−0·33	0·41
0·15	0·80	0·20	−0·13	0·31
0·15	0·80	0·15	0·05	0·27
0·15	0·80	0·10	0·25	0·35
0·10	0·80	0·15	−0·08	0·26

The performance of CAM-PBE on electron affinities can be improved by increasing the a parameter and decreasing the ω parameter. The optimal performance of the parameters tested came with $a = 0.10$, $a + b = 0.80$ and $\omega = 0.15$. This set of parameters will be refered to as CAM-PBE-ea. The performance of this new set of parameters was tested using the same set of properties as CAM-PBE. The mean and mean absolute errors of CAM-PBE-ea are compared to those for CAM-B3LYP, LC-ωPBE and CAM-PBE. The ground state properties are shown in table 41 and the excited state properties are shown in table 42.

Table 41: Errors for CAM-PBE ground state properties.

Ionisation Potentials	CAM-B3LYP	LC-ωPBE	CAM-PBE	CAM-PBE-ea
d/eV	0·10	0·07	−0·62	0·16
$\|d\|$/eV	0·18	0·20	0·62	0·24
σ/eV	0·21	0·25	0·23	0·30
Atomisation Energies				
d/kcal mol^{-1}	1·5	0·4	2·7	32·4
$\|d\|$/kcal mol^{-1}	3·2	3·2	4·9	32·6
σ/kcal mol^{-1}	3·7	4·1	5·7	22·0
Diatomic Bond Lengths				
d/Å	−0·006	−0·001	0·006	−0·024
$\|d\|$/Å	0·013	0·027	0·037	0·026
σ/Å	0·018	0·039	0·060	0·014
Classical Reaction Barriers				
d/kcal mol^{-1}	−2·5	0·0	−0·9	−5·2
$\|d\|$/kcal mol^{-1}	2·8	1·2	1·5	5·2
σ/kcal mol^{-1}	2·1	1·7	1·7	2·8
Vibrational Wavenumbers				
d/cm^{-1}	33	51	58	74
$\|d\|$/cm^{-1}	37	56	60	74
σ/cm^{-1}	47	61	68	66
Isotropic Polarisabilities				
d/au	0·24	−0·25	−0·02	−0·60
$\|d\|$/au	0·37	0·35	0·29	0·60
σ/au	0·36	0·41	0·41	0·48
Electron Affinities				
d/eV	0·04	0·00	−0·70	−0·08
$\|d\|$/eV	0·10	0·13	0·74	0·26
σ/eV	0·13	0·15	0·15	0·41

Table 42: Errors for CAM-PBE-ea excited state properties.

Singlet–Local	CAM-B3LYP	LC-ωPBE	CAM-PBE	CAM-PBE-ea
d/eV	0·02	0·14	0·10	−0·11
$\|d\|$/eV	0·21	0·23	0·21	0·19
σ/eV	0·29	0·27	0·27	0·26
Singlet–Rydberg				
d/eV	−0·48	−0·02	−0·51	−0·99
$\|d\|$/eV	0·48	0·14	0·57	0·99
σ/eV	0·18	0·17	0·38	0·29
Singlet–Charge Transfer				
d/eV	−0·18	0·11	0·10	−1·22
$\|d\|$/eV	0·27	0·68	0·28	1·24
σ/eV	0·31	0·81	0·34	0·71
Triplet–TDDFT				
d/eV	−0·39	−0·48	−0·52	−0·58
$\|d\|$/eV	0·40	0·53	0·53	0·58
Triplet–TDA-TDDFT				
d/eV	−0·16	−0·14	−0·21	−0·36
$\|d\|$/eV	0·18	0·22	0·24	0·37

From the good performance seen with CAM-PBE, excluding the errors seen with electron affinities and ionisation potentials, the performance of the electron affinity optimised functional (CAM-PBE-ea) is much worse across all the properties studied here (excluding the electron affinities and closely related ionisation potentials). There is a huge decrease in the accuracy of atomisation energies which is expected as CAM-PBE is tuned on atomisation energies. There is also a huge decrease in the accuracy of singlet charge transfer type excitations which is unexpected as CAM-B3LYP and CAM-PBE show an improved performance on this type of excitation. This suggests that it may be possible to generate a set of parameters for a 'CAM' type PBE based functional but it may need to be tuned to the particular property that is being investigated. For example CAM-PBE performs well for polarisabilities but is poor for electron affinities and ionisation potentials (these properties are both related to charged molecules). Therefore CAM-PBE should be used for calculation of polarisabilities and related properties (properties related to changes in electron density) but not for electron affinities and related properties (properties related to charged molecules).

2.5 Conclusions

The extensive assessment of the 3 global and 3 range separated hybrids (B3LYP, PBE0, B97-2, CAM-B3LYP, LC-ωPBE and ωB97-XD) over a range of ground state and excited state properties has shown some interesting results. It highlights that each functional tested has properties where it performs accurately and

those with less accurate performance. B3LYP performs well for ionisation energies, vibrational wavenumbers, electron affinities and local excitations but less well for isotropic polarisabilities, Rydberg excitations and charge transfer excitations. PBE0 performs well for isotropic polarisabilities and local excitations but less well for electron affinities and atomisation energies. B97-2 performs well for ionisation energies, atomisation energies, vibrational wavenumbers, isotropic polarisabilites and local excitations but less well for diatomic bond lengths, Rydberg excitations and charge transfer excitations. CAM-B3LYP performs well for diatomic bond lengths and all excitations but less well for isotropic polarisabilities. LC-ωPBE performs well for reaction barriers, local excitations and Rydberg excitations but less well for ionisation potentials, electron affinities, diatomic bond lengths, vibrational wavenumbers. ωB97-XD performs welll for most properties but less well ionisation potentials and vibrational wavenumbers.

The increased computational cost and complexity associated with using a range separated hybrid is important for some properties but less so for others. It is important to use range separated hybrids for the accurate calculation of excitations (other than local excitations, where all functionals tested performed well), atomisation energies and reaction barriers. The performance of LC-ωPBE is poorer than the other range separated hybrids in most excitations excluding Rydberg excitations.

However, the performance of LC-ωPBE may be improved by tuning the parameters of the range separation. Initially the parameters used where the same as those of CAM-B3LYP ($a = 0.19$,$a + b = 0.65$ and $\omega = 0.33$ a_0^{-1}), generating CAM-ωPBE. This new functional has poor performance across most properties with exception of classical reaction barriers which are calculated with comparative accuracy to LC-ωPBE. The parameters where optimised using a set of atomisation energies whilst applying the constraint $a + b = 0.65$ (keeping the $a + b$ value the same as CAM-B3LYP). This lead to an optimised version of CAM-ωPBE (CAM-ωPBEa). The performance of this optimised functional was improved over CAM-ωPBE. However, the performance was still poor when compared to CAM-B3LYP and LC-ωPBE.

In order to understand the underlying cause of the poor performance of CAM-ωPBE and CAM-ωPBEa. A study of the quantity of HFx required for optimally performing global hybrid of a series of underlying functionals (LDA, PBE, BLYP and TPSS) by assessing the performance of a global hybrid with increasing quantities of HFx over a set of atomisation energies. This lead to some interesting results, each underlying functional had a differing quantity of HFx in its optimal global, with LDA requiring the highest quantity and TPSS requiring the least.

It was also seen that the optimal quantity of HFx differed for some global

hybrids (especially for PBE and LDA) when considering different molecule types (cyclic and non cyclic). This shows that some molecule types have differing behaviour when considering the quantity of HFx required (cyclic requires more HFx for PBE based). This shows that what is meant by a diverse set may require redefinition as it could be possible to lose details when considering too large a set or under or over representing certain categories of molecules (such as cyclic and non cyclic molecules). This observation could lead to utilising more of the data that is present in current benchmark sets to enable more accurate calculation of properties with functionals tuned to the categories of molecules. Current benchmark data could be split into different subsets and this could provide information that is currently being missed due to the drive to tuning for average performance.

In order to properly assess the optimal global hybrid based on PBE and BLYP, a similar assessment was performed over the full G-1 and G-2 sets of atomisation energies. The optimal global hybrid for each underlying functional had a differing quantity of HFx (values of HFx in each optimal). This difference sheds some light on the issues seen with the initial CAM-ωPBE and CAM-ωPBEa functionals, due to insufficient HFx included when $a + b = 0.65$ was enforced.

With the observation about the differing quantity of HFx required for an optimally performing global hybrid of PBE and BLYP, another optimisation of the range separation parameters for CAM-ωPBE with the condition $a + b < 1$ was undertaken. The optimisation used a series of atomisation energies, including pyridine due to the differing behaviour of cyclic systems observed when considering global hybrids of PBE. The optimised parameters were $a = 0.05$, $a + b = 0.80$ and $\omega = 0.40\mathrm{a}_0^{-1}$, referred to as CAM-PBE. The performance of CAM-PBE was much improved over CAM-ωPBEa; especially for excited state properties where CAM-PBE has a comparable performance to CAM-B3LYP, which suggests that an attenuated range separated hybrid of PBE may have merits for the calculation of excited state properties. However, there is poor performance when calculating ionisation energies and electron affinities.

Thus a set of parameters were optimised on electron affinities ($a = 0.10$, $a + b = 0.80$ and $\omega = 0.15\ \mathrm{a}_0^{-1}$), referred to as CAM-PBEea. However, whilst these parameters improved the performance of electron affinities and ionisation potentials, they had a detrimental effect on the performance of other properties where CAM-PBE had good performance. Therefore whilst CAM-PBE has good performance generally it should be noted about the poor performance of electron affinities and ionisation potentials.

3 Excited State Geometries and Emission Energies

The excitations considered previously are all absorption energies—corresponding to the process whereby a photon is absorbed by a molecule in its ground state, causing in this case 1 electron to be excited. This is modelled as the energy difference between the minimum of the ground state potential energy surface to the corresponding point on the excited state surface. There is a second excitation process that can be studied: emission. Emission involves a molecule in an excited state losing energy via the emission of a photon, and is modelled as the energy difference between the minimum of the *excited state* potential surface and the corresponding point of the ground state surface. A schematic of these two types of excitations is shown in figure 24.

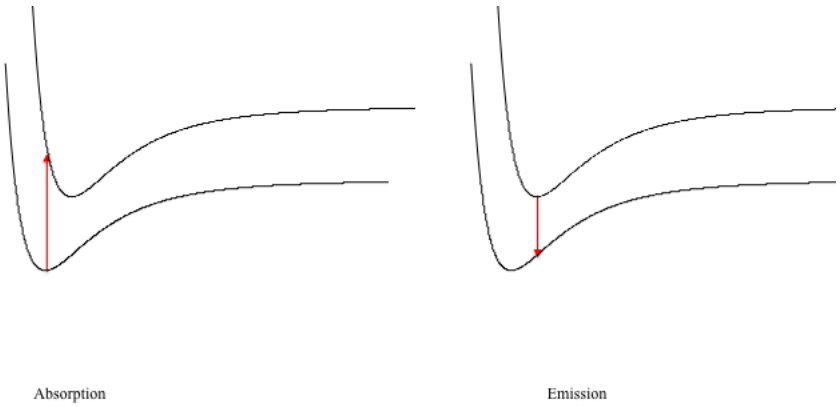


Figure 24: Schematic of absorption and emission excitations.

Although the same machinery (i.e., TDDFT) can be employed to describe both situations involving the interaction of light with a molecule, the subtle difference in the nature of the geometries involved between the absorption and emission processes results in additional complexity and consequently a significant hurdle when attempting to undertake accurate quantum chemical calculations.

The additional complexity that is implicit in the calculation of emission energies with quantum chemical methods comes from the increase in the steps required for the calculation. A schematic for the calculation of absorption and emission energies is shown in figure 25.

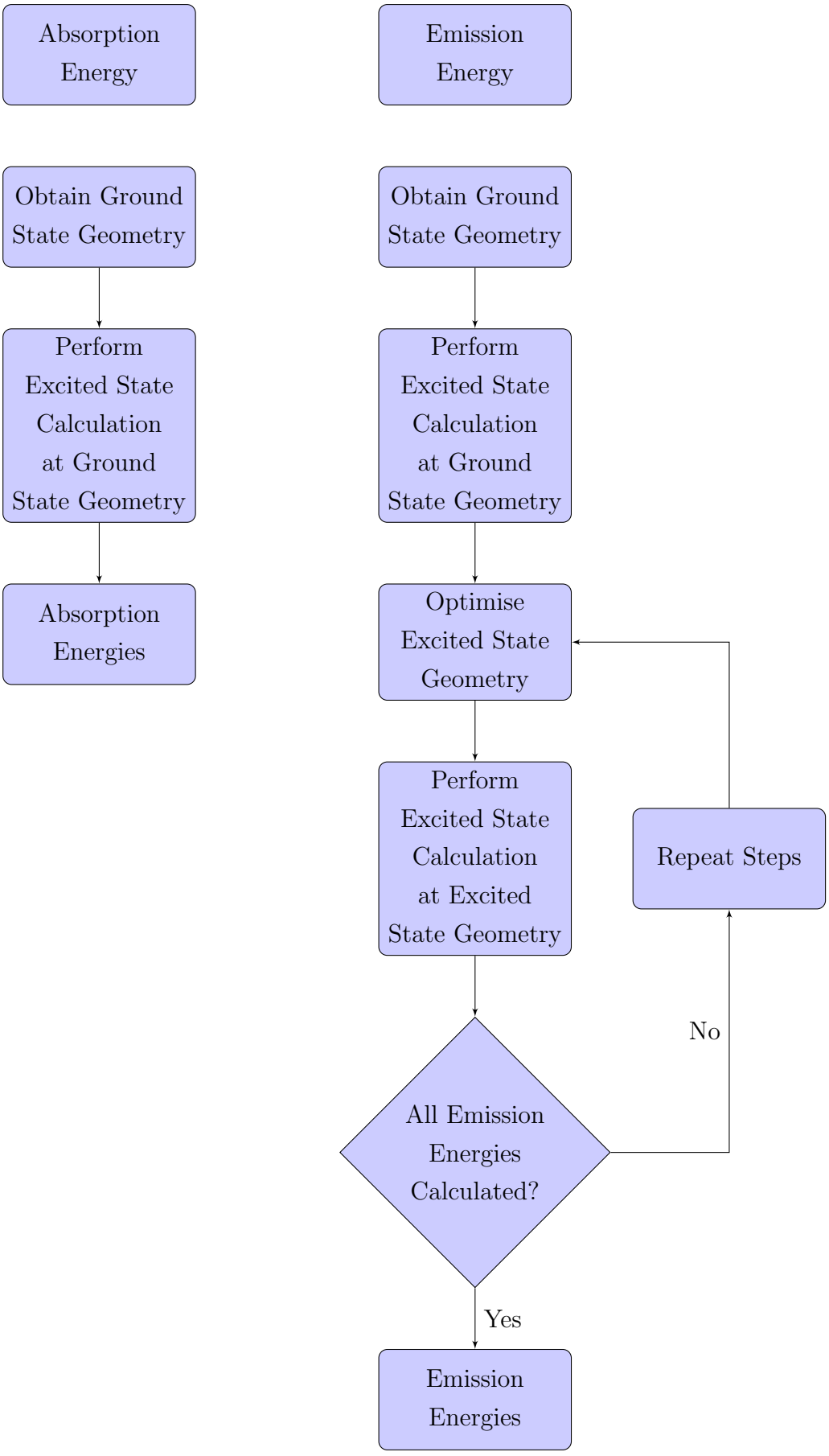


Figure 25: Schematic of the Steps required for the calculation of Absorption and Emission Energies

The additional complexity is obvious when looking at this schematic, there are a few causes for this that will be highlighted. The calculation of absorption energies required 2 steps, obtaining a ground state geometry (either from DFT or from a literature structure using a higher level method) and then an excited state calculation at this structure to obtain the absorption energies. These steps are also included in the calculation of emission energies along with several others. The calculation of emission energies requires an excited state geometry optimisation for each state of interest along with an excited state calculation at this geometry. These steps add considerable computational cost to the calculation of emission energies. Along with a compounding effect when there are multiple states of interest for a single molecule, as a set of absorption energies requires 2 steps and a set of emission energies requires $2 + 2n$ steps (where n is the number of states of interest). Further discussion on the calculation of excited state properties can be seen in references 97,98

This increased complexity leads to the significant hurdle mentioned above, the origin of which is two-fold. Firstly, the accuracy of quantum chemical methods may be diminished away from the ground state geometry. Secondly, we need to obtain an accurate excited state geometry to begin with. Currently, there is no notable benchmark set of excited state geometries, or emission energies, to quantify the accuracy of our results. However, there is experimental data describing both the geometry and emission energy of a multitude of excited states in diatomic molecules, but there is some data beyond diatomics but not the the same quantity of those present for ground state properties.

There is a lack of an extensive benchmark of excited state structures and emission energies in literature of a similar manner to those that are available for ground state properties and absorption energies which have been discussed in detail in the previous chapter. There are several reasons for this lack of an extensive benchmark, amongst which is an inability to experimentally access the results which would be required for the formation of a benchmark. However, there is an exception concerning diatomic molecules where extensive experimental data exists. Despite the existence of this experimental data, diatomic molecules are a poor choice to develop a benchmark set from due to the limited chemical interest of these molecules and the ability to brute force the correct structure due to these molecules having only a single coordinate (bond length) which can be altered. Another reason for the lack of an extensive benchmark of excited state structures and emission energy is due to the complexity and cost of obtaining high level theory results for these properties due to the nature of how the properties are modelled. When emission energies and excited state structures have been calculated/reported in previous literature using high level theory it is limited to a few specific states of interest

for a few specific molecules rather than an extensive set of states and molecules as is the case for absorptions, benchmark sets of which were used in the previous chapter.

The lack of an extensive benchmark set for these properties leads to questions about the accuracy of calculating the properties using various approximate E_{xc}s, as the assessment of approximate E_{xc}s is used to suggest which functionals should be used for the accurate calculation/the expected error in the results when using approximate E_{xc}s, thus the lack of a benchmark set for these properties is an important gap when considering using approximate E_{xc}s for the calculation of these properties. There is no way of telling how accurate approximate E_{xc}s are for these properties.

This chapter will initially discuss 2 diatomic molecules as a preliminary investigation using DFT to suggest where issues may exist when applying DFT to these properties. This is followed by the choice of molecules and states included and method that will be used in the development of a benchmark set of excited state properties and emission energies. Finally, there will be an assessment of several approximate E_{xc}s using the benchmark set, with suggestions of the type of functional/properties of approximations that should be used for the accurate calculation of emission energies/excited state structures.

3.1 CO Emissions and Excited State Geometries

The emission energies and excited state geometries of CO can be compared to extensive experimental data about the excited states from NIST webbook[99]. The energies of the CO states were followed using CAM-B3LYP and three basis sets (cc-pVDZ, aug-cc-pVDZ and daug-cc-pVDZ). The energy surfaces of the states (3 Σ, 3 Π and Δ states for cc-pVDZ and aug-cc-pVDZ and 4 Σ, 2 Π and Δ states for daug-cc-pVDZ) are modelled by incrementing the bond length of CO between 0.8Å and 1.9Å and calculating the energies of the states. The plots of the lowest three energy states (Σ^-, Π and Δ) are plotted with the ground state and compared to the plots seen in reference 100. The plots of these for the three basis sets can be seen in figures 26, 27 and 28.

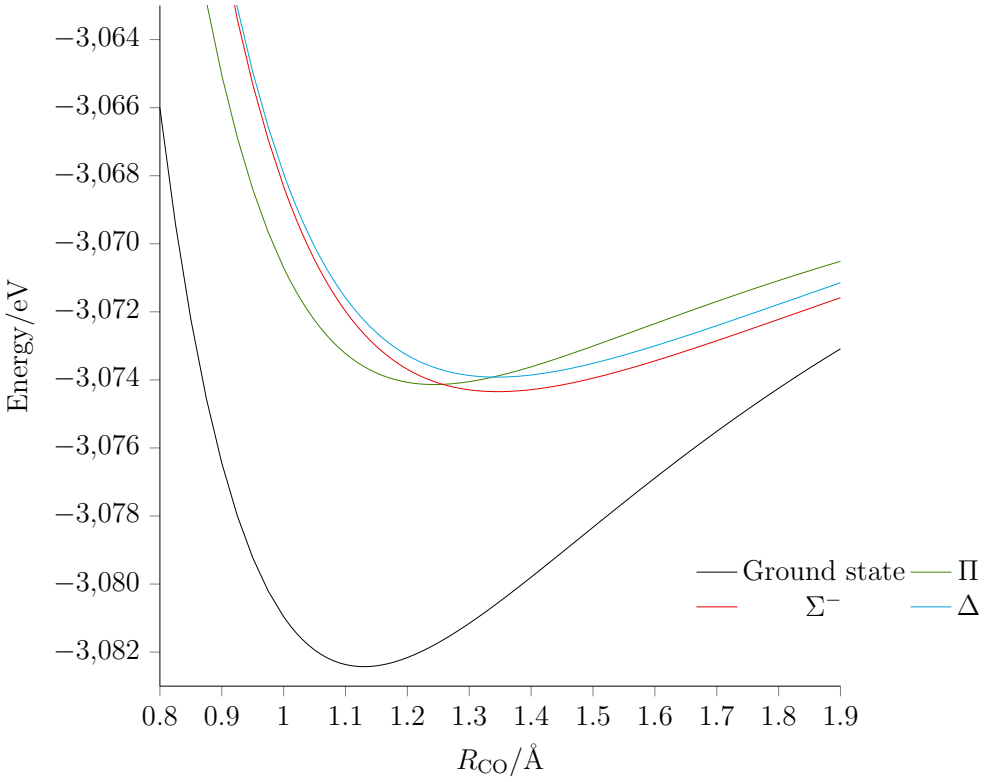


Figure 26: Plot of lowest three energy states for cc-pVDZ.

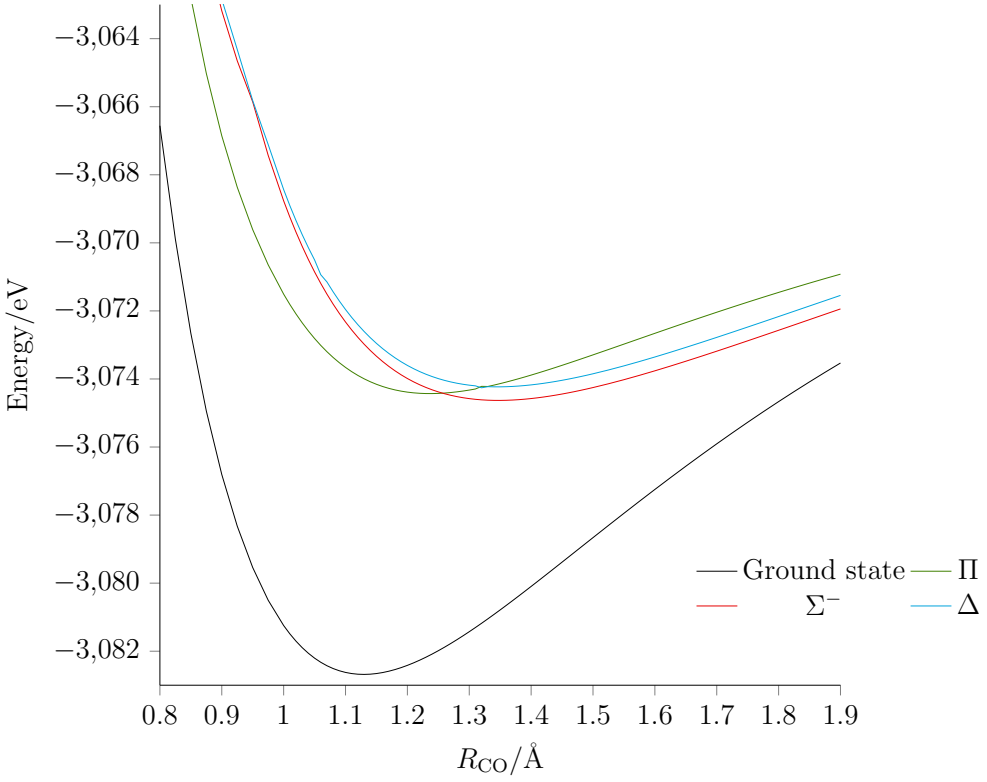


Figure 27: Plot of lowest three energy states for aug-cc-pVDZ.

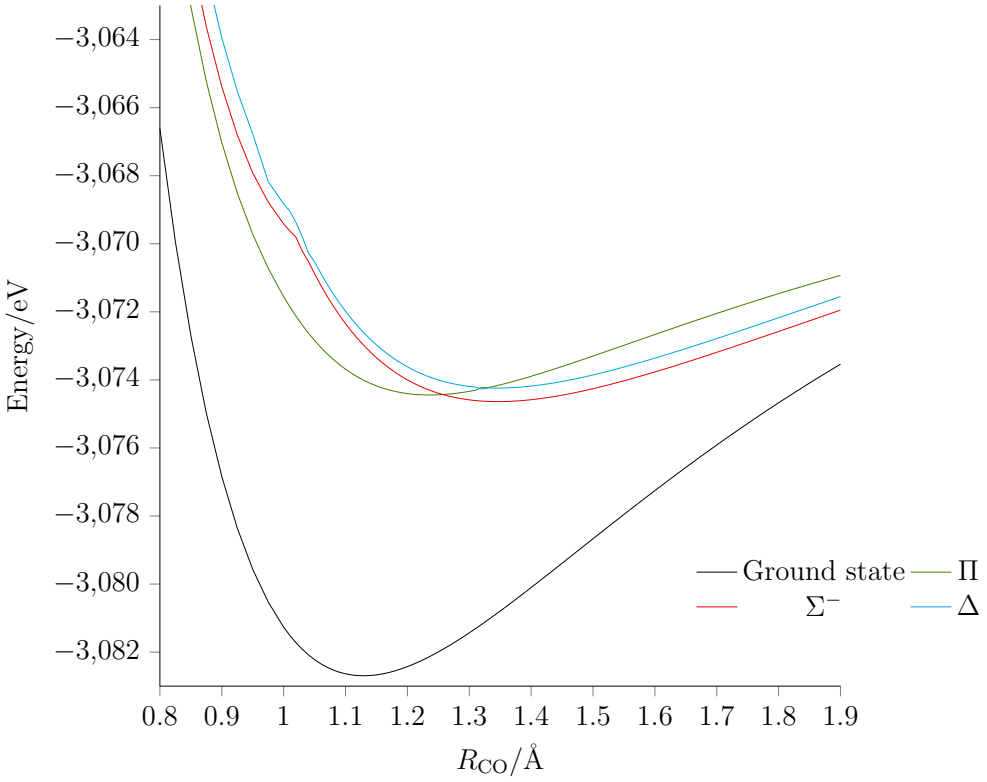


Figure 28: Plot of lowest three energy states for daug-cc-pVDZ.

The minima of the three states in all of these plots are close to the experimental bond lengths for the states (1.2353Å for Π, 1.3911Å for Σ^- and 1.399Å for Δ). With the energy surfaces plotted here, it can clearly be seen that the states cross over and may cause the incorrect state to be optimised during calculations. This means that great care will need to be taken to ensure the correct state is being followed during the excited state geometry optimisations.

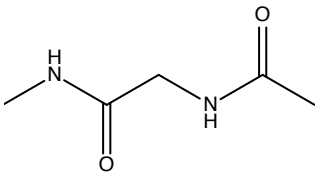
3.2 Selection of Benchmark Set

There exists no benchmark set of excited state geometries and thus emission energies in the same way there is for ground state geometries and absorption sets (e.g. Thiel set of triplet absorptions[86–88]). The selection of molecules and states in a benchmark set needs to be carefully considered as if there is a lack of variety the set will not be useful for investigating several types of excitation/molecules, and include molecules which are of chemical interest. Having a variety of molecules and excitations in the set will enable the assessment of functionals over a wider series of excited state geometries and thus emission energies.

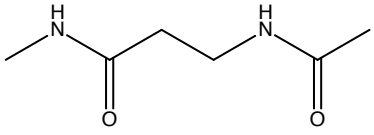
The diatomics investigated above are not of special chemical interest, have experimental data and can be solved using **brute force** (due to the single degree of freedom, bond length). Diatomic molecules are not a good type of molecule to

base a benchmark set on, they *could* be included but are of limited interest.

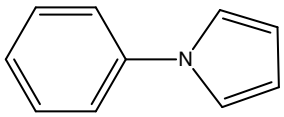
The benchmark set will be built from existing benchmark sets of molecules for absorptions, these sets will be the lambda set of singlet absorptions[47] and the Thiel set of triplet absorptions[86–88] (both used previously to assess density functionals). The lambda set has a variety of types of singlet excitations (local, charge-transfer and Rydberg) along with a selection of organic molecules. The Thiel set has a series of triplet excitations of a variety of organic molecules. The molecules chosen along from these two absorption sets with the states for each molecule are shown in figures 29 and 30. The set will be expanded by including the triplet equivalents of the lambda set and the singlet equivalents of the Thiel set.



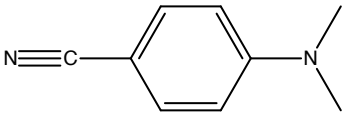
Dipeptide: $n_1 \rightarrow \pi_1^*$, $n_2 \rightarrow \pi_2^*$, $\pi_1 \rightarrow \pi_2^*$



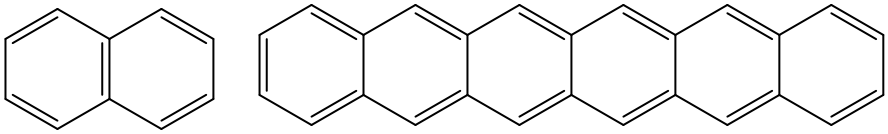
β-dipeptide: $n_1 \rightarrow \pi_1^*$, $n_2 \rightarrow \pi_2^*$, $\pi_1 \rightarrow \pi_2^*$



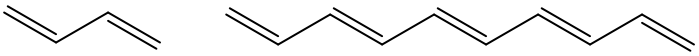
Phenylpyrrole: $1B_2$, $2A_1$, $2B_2$, $3A_1$



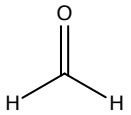
DMABN: B, A



Acene $n = 1 \rightarrow$ Acene $n = 5$: B_{2u}, B_{3u}



PA oligomer $n = 2 \rightarrow$ PA oligomer $n = 5$: B_u



H_2CO: $1A_2$, $2A_2$, $1B_2$, $1A_1$, $2B_2$, $3A_2$

Figure 29: Molecules and states from the Lambda set of absorptions that will form part of the excited state geometry and emission energy benchmark set

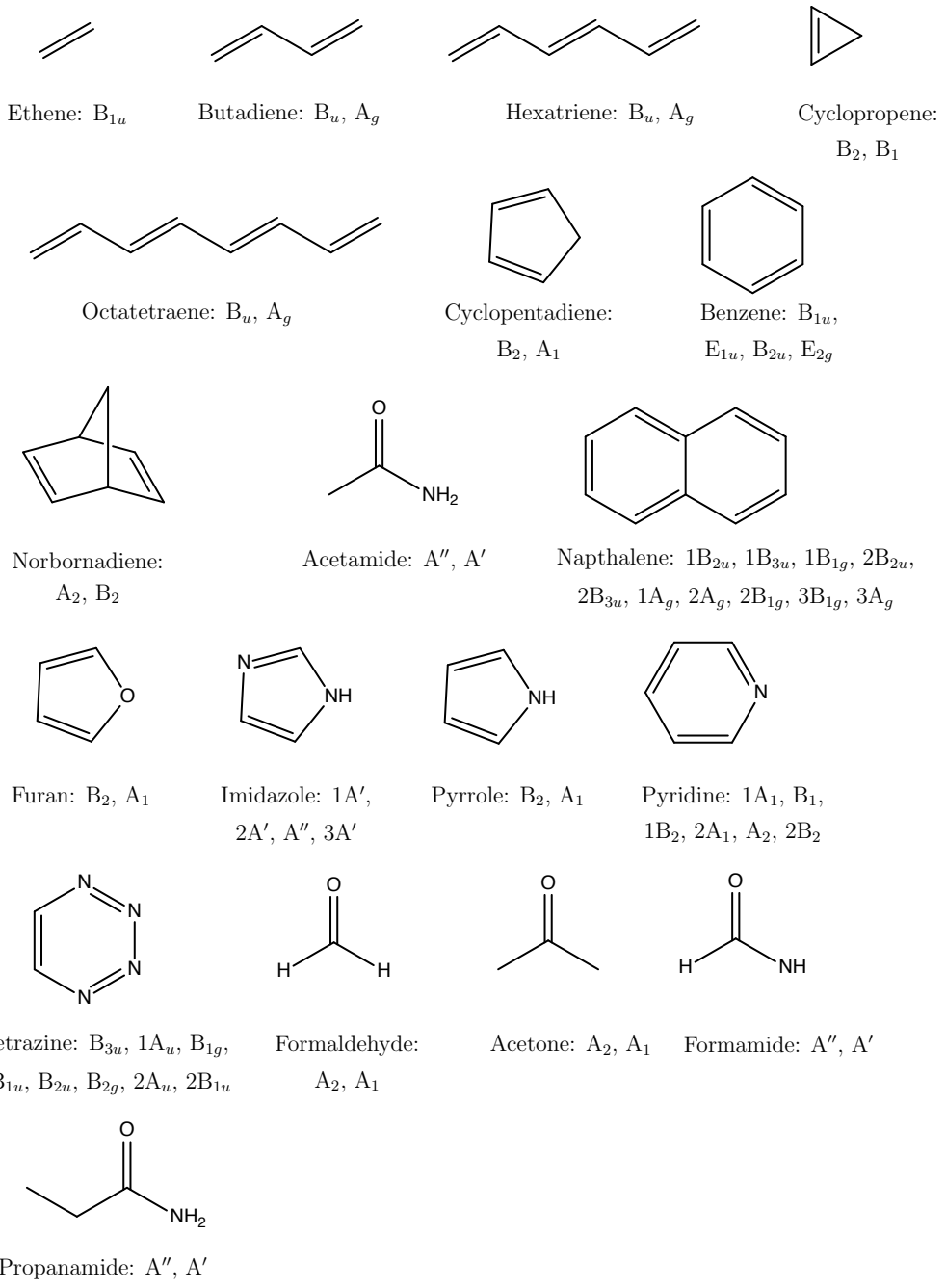


Figure 30: Molecules and states from the Thiel set of absorptions that will form part of the excited state geometry and emission energy benchmark set

The molecules chosen have an obvious chemical diversity and also contain series of closely related molecules (such as the polyacetylenes and acenes).

3.3 Validation of Method Selected for Calculation of Benchmark Set

Certain molecules and states are excluded due to issues with calculating them when using the method used for the benchmark set. The method selected for use in the calculation of the benchmark set is approximate second order coupled cluster, CC2. It has been previously used to benchmark DFT results for certain molecules, such as DMABN[101], and with the diversity and size of the set of molecules, it is a feasible way to obtain a full set of excited state structures and emission energies. In order to assess the suitablity of CC2 to calculate other molecules in the benchmark set, any existing values for emissions of the states in the set with higher level theory should be sought out. Literature values with higher levels of theory have been found for a few of the states in the benchmark set. The states found were all singlet states and are shown in table 43 along with the CC2 calculated values. The cc-pVDZ basis set was used to calculate these emission values. The method used for each literature emission is also shown in table 43.

Table 43: Literature emission energies found for the states in the benchmark set compared to the CC2 calculated values, all values in eV.

Molecule	State	CC2 emission /eV	Literature emission /eV
Phenylpyrrole	$1B_2$	4·75	4·45[a]
	$2B_2$	4·95	4·71
Butadiene	B_u	5·87	5·73[b]

[a] CASPT2 reference 102

[b] Expt. reference 103

The emission values for the states found in literature calculated with CC2 show good agreement with the literature values. There is some discrepancy between the results. This seems dependent on the method used in the previous literature; the phenylpyrrole and butadiene states are overestimated with CC2 compared to the CASPT2 and experimentally derived values in literature respectively. It is noted that CC2 has at worst a 0.3 eV error from the literature value obtained. This is a little high but this is only a single case, the errors in other cases is smaller. It shows that care may need to be used for some states but generally CC2 is a good enough method to use for the generation of the benchmark set of excited state geometries and emission energies, due to the fairly good agreement with literature values and the use of the method for previous studies into molecules in the benchmark set, along with the feasability of using the method to generate to volume of excited state structures and emission energies in the benchmark set at a reasonable computational cost.

As mentioned previously with the diatomics there may a benefit to using an augmented basis set for the calculation of higher energy singlet emissions. This is tested to check if the additional computation cost associated with undertaking this is necessary for the calculation of the full benchmark set of singlets. Table 44 shows the results for the CC2 emissions using cc-pVDZ and aug-cc-pVDZ.

Table 44: CC2 emission energies using cc-pVDZ and aug-cc-pVDZ basis sets for the states with literature values, all values in eV.

Molecule	State	cc-pVDZ/eV	aug-cc-pVDZ/eV	Literature/eV
Phenylpyrrole	$1B_2$	4·75	4·37	4·45
	$2B_2$	4·95	4·74	4·71
Butadiene	B_u	5·87	5·42	5·73

The augmented basis set gives much closer to literature values for the phenylpyrrole states, however it overcorrects the overestimation seen in the butadiene state. The augmented basis set gives a smaller mean absolute error for the states (0.23 eV for cc-pVDZ and 0.14 eV for aug-cc-pVDZ). This shows that aug-cc-pVDZ is the basis set that should be used for the calculation of the benchmark set.

3.4 CC2 Values for the Benchmark Set of Emission Energies

As the selection of states and molecules in the benchmark set has been completed and the method selected for the calculation of the benchmark excited state geometries and emission energies has been validated, it follows that the calculation of the full set of states needs to be undertaken. This will be split into four sections, the Lambda set singlet and triplet emissions and the Thiel set triplet and singlet emissions. The method used in each case will be discussed in the section. There will be an assessment of the influence of the basis set on some of the states in the benchmark set, especially including diffuse functions (cc-pVDZ vs aug-cc-pVDZ).

3.4.1 Lambda Singlet Emission Energies

The lambda singlet emission energies were calculated using CC2 and the aug-cc-pVDZ basis set for all states except the H_2CO states which used the daug-cc-pVTZ basis set. The much larger daug-cc-pVTZ basis set was used for the calculation of the H_2CO states due to the high energy of these states, some of the states have a similar energy to those which had issues when investigating the diatomics. Thus, the larger basis set was deemed necessary for the calculation of the H_2CO states.

The emission energies for the singlet states from the lambda set of states are shown in table 45.

Table 45: Lambda CC2 singlet emission energies, all values are in eV.

Molecule	State	Emission Energy/eV
Dipeptide	$\pi_1 \rightarrow \pi_2^*$	6·25
	$n_1 \rightarrow \pi_1^*$	3·24
	$n_2 \rightarrow \pi_2^*$	5·07
β-dipeptide	$\pi_1 \rightarrow \pi_2^*$	5·87
	$n_1 \rightarrow \pi_1^*$	3·10
	$n_2 \rightarrow \pi_2^*$	5·35
Acene 1	B_{2u}	4·07
	B_{3u}	4·19
Acene 2	B_{2u}	3·15
	B_{3u}	3·59
Acene 3	B_{2u}	2·47
	B_{3u}	3·27
Acene 4	B_{2u}	2·12[a]
	B_{3u}	3·12[a]
Acene 5	B_{2u}	1·78[a]
	B_{3u}	2·97[a]
PP	1 B_2	4·37
	1 A_1	4·43
	2 B_2	4·74
	2 A_1	5·24
DMABN	B	4·07[a]
	A	2·43[a]
PA2	B_u	5·42
PA3	B_u	4·57
PA4	B_u	3·97
PA5	B_u	3·53
H_2CO	3 A_2	8·59
	2 A_2	7·54
	1 B_1	8·45
	2 B_2	7·28
	1 A_1	7·23
	1 B_2	6·31
	1 A_2	3·10

[a] cc-pVDZ values due to issues isolating the excited state geometries with aug-cc-pVDZ

The emission values for the singlet states of DMABN and acene 4 and 5 have

calculated using cc-pVDZ (a smaller basis set) due to issues with isolating the excited states with aug-cc-pVDZ (similar to the diatomics). The excited state structure using cc-pVDZ was then used as an initial geometry in an attempt to start closer to the geometry for aug-cc-pVDZ and this did not enable isolation of the states.

3.4.2 Lambda Triplet Emission Energies

The lambda triplet emission energies were calculated using CC2 and the aug-cc-pVDZ basis set for all states except the H_2CO states which use the daug-cc-pVTZ basis set. Again the daug-cc-pVTZ basis set was deemed necessary for the calculation of the H_2CO states.

The emission energies for the singlet states from the lambda set of states are shown in table 46.

Table 46: Lambda CC2 triplet emission energies, all values are in eV.

Molecule	State	Emission Energy/eV
Dipeptide	$\pi_1 \rightarrow \pi_2^*$	4·30
	$n_1 \rightarrow \pi_1^*$	3·32
	$n_2 \rightarrow \pi_2^*$	4·69
β-dipeptide	$\pi_1 \rightarrow \pi_2^*$	5·04
	$n_1 \rightarrow \pi_1^*$	3·17
	$n_2 \rightarrow \pi_2^*$	4·52
Acene 1	B_{2u}	2·51
	B_{3u}	4·03
Acene 2	B_{2u}	1·69
	B_{3u}	3·60
Acene 3	B_{2u}	1·15
	B_{3u}	3·32
Acene 4	B_{2u}	0·75[a]
	B_{3u}	3·21[a]
Acene 5	B_{2u}	0·47[a]
	B_{3u}	3·08[a]
PP	1 B_2	3·13
	1 A_1	3·26
	2 B_2	4·08
	2 A_1	4·31
DMABN	A	2·80
PA2	B_u	2·08
PA3	B_u	1·59
PA4	B_u	1·26
PA5	B_u	1·02
H_2CO	3 A_2	8·58
	2 A_2	7·55
	1 B_1	5·84
	2 B_2	7·20
	1 A_1	3·52
	1 A_2	2·73

[a] cc-pVDZ values due to issues isolating the excited state geometries with aug-cc-pVDZ

The emission values for the triplet states of acene 4 and 5 have calculated using cc-pVDZ (a smaller basis set) due to issues with isolating the excited states with aug-cc-pVDZ (similar to the diatomics). The excited state structure using

cc-pVDZ was then used as an initial geometry in an attempt to start closer to the geometry for aug-cc-pVDZ and this still did not enable isolation of the states. It is also noted that there are a few triplet states that are absent from the calculated CC2 emission values that were present in the singlet states. These are the B state of DMABN and the $1B_2$ state of H_2CO. Both of these states had issues with optimising the excited state geometry, all avenues were exhausted in attempts to obtain structures for these two missing states and they remained absent.

3.4.3 Thiel Triplet Emission Energies

The Thiel triplet emission energies were calculated using CC2 and the aug-cc-pVDZ basis set for all states. The emission energies for the singlet states from the lambda set of states are shown in table 47.

Table 47: Thiel CC2 triplet emission energies, all values in eV.

Molecule	State	Emission Energy/eV
Ethene	B_{1u}	2·87
Butadiene	B_u	2·08
	A_g	4·41
Hexatriene	B_u	1·59
	A_g	3·70
Octatetraene	B_u	1·26
	A_g	3·15
Cyclopropene	B_2	2·60
	B_1	6·11
Cyclopentadiene	B_2	2·05
	A_1	4·40
Norbornadiene	B_2	3·37
	A_1	5·27
Benzene	B_{1u}	3·95
	E_{1u}	4·74
	B_{2u}	5·62
	E_{2g}	6·97
Naphthalene	1 B_{2u}	2·51
	1 B_{3u}	4·03
	1 B_{1g}	3·97
	2 B_{2u}	4·22
	2 B_{3u}	4·71
	1 A_g	5·35
	2 A_g	6·28
	2 B_{1g}	5·60
	3 B_{1g}	6·74
	3 A_g	6·47
Furan	B_2	3·15
	A_1	4·90
Pyrrole	B_2	3·60
	A_1	5·05
Imidazole	1 A’	3·60
	2 A’	4·84
	1 A”	4·85
	3 A’	5·55
Pyridine	1 A_1	3·58
	B_1	3·54
	1 B_2	4·49
	2 A_1	4·86
	A_2	3·63
	2 B_2	5·92
Tetrazine	B_{3u}	1·60
	1 A_u	1·48
	B_{1g}	4·26
	1 B_{1u}	3·18
	B_{2u}	4·21
	B_{2g}	3·31
	2 A_u	4·33
	2 B_{1u}	5·09
Formaldehyde	A_2	2·61
	A_1	3·42
Acetone	A_2	2·63
	A_1	3·51
Formamide	A”	3·21
	A’	3·87
Acetamide	A”	3·20
	A’	4·31
Propanamide	A”	3·21
	A’	4·34

3.4.4 Thiel Singlet Emission Energies

The Thiel triplet emission energies were calculated using CC2 and the aug-cc-pVDZ basis set for all states. The emission energies for the singlet states from the lambda set of states are shown in table 48.

Table 48: Thiel CC2 singlet emission energies, all values in eV.

Molecule	State	Emission Energy/eV
Ethene	B_{1u}	6·89
Butadiene	B_u	5·42
	A_g	5·99
Hexatriene	B_u	4·57
	A_g	5·10
Octatetraene	B_u	3·97
	A_g	4·31
Cyclopropene	B_2	5·21
	B_1	6·23
Cyclopentadiene	B_2	4·48
	A_1	5·56
Norbornadiene	B_2	3·37
	A_1	5·32
Benzene	B_{1u}	4·83
	E_{1u}	6·81
	B_{2u}	6·21
	E_{2g}	7·62
Naphthalene	1 B_{2u}	4·07
	1 B_{3u}	4·19
	1 B_{1g}	5·29
	2 B_{2u}	5·82
	2 B_{3u}	5·75
	1 A_g	5·77
	2 A_g	6·48
	2 B_{1g}	6·04
	3 B_{1g}	6·98
	3 A_g	7·14
Furan	B_2	5·64
	A_1	5·80
Pyrrole	B_2	5·78
	A_1	5·77
Imidazole	1 A’	5·24
	2 A’	6·00
	1 A”	4·87
	3 A’	6·21
Pyridine	1 A_1	5·14
	B_1	4·04
	1 B_2	4·78
	2 A_1	6·12
	A_2	3·66
	2 B_2	5·63
Tetrazine	B_{3u}	2·18
	1 A_u	1·64
	B_{1g}	5·03
	1 B_{1u}	6·36
	B_{2u}	4·53
	B_{2g}	3·63
	2 A_u	4·65
	2 B_{1u}	6·27
Formaldehyde	A_2	3·00
	A_1	7·04
Acetone	A_2	2·69
	A_1	6·25
Formamide	A”	3·15
	A’	4·63
Acetamide	A”	3·12
	A’	4·30
Propanamide	A”	3·12
	A’	4·33

3.4.5 Internal Consistency

There are some molecules and states which repeat in the two sets, these can be used to check the internal consistency of the method (the emission for each state should be equivalent). These molecules, and states in brackets, are: butadiene, hexatriene and octatetraene (B_u) and naphthalene ($1B_{2u}$ and $1B_{3u}$). This will check that the method is getting to the same excited state structure and thus emission energy regardless of the initial geometry. Table 49 shows the results for each of these states for each set.

Table 49: Emission energies for the common states across the two sets with each basis set used, all values in eV.

Molecule	State	Lambda Singlet/eV	Thiel Singlet/eV	Lambda Triplet/eV	Thiel Triplet/eV
Butadiene	B_u	5·42	5·42	2·08	2·08
Hexatriene	B_u	4·57	4·57	1·59	1·59
Octatetraene	B_u	3·97	3·97	1·26	1·26
Naphthalene	B_{2u}	4·07	4·07	2·51	2·51
	B_{3u}	4·19	4·19	4·03	4·03

The results for each of these states is identical between the two sets, which is expected. This means that only one copy of each state is required to be used in the benchmark set.

3.5 Complete CC2 Emission Energy Benchmark

The two sets will be combined to generate a complete benchmark of emission energies. The emission energies calculated with CC2 are shown in table 50.

Table 50: Complete CC2 emission energy benchmark, all values in eV.

Molecule	State	Singlet Emission/eV	Triplet Emission/eV
Dipeptide	$\pi_1 \rightarrow \pi_2^*$	6·25	4·30
	$n_1 \rightarrow \pi_1^*$	3·24	3·32
	$n_2 \rightarrow \pi_2^*$	5·07	4·69
β-dipeptide	$\pi_1 \rightarrow \pi_2^*$	5·87	5·04
	$n_1 \rightarrow \pi_1^*$	3·10	3·17
	$n_2 \rightarrow \pi_2^*$	5·35	4·52
Naphthalene/	1 B_{2u}	4·07	2·51
Acene 1	1 B_{3u}	4·19	4·03
	1 B_{1g}	5·29	3·97
	2 B_{2u}	5·82	4·22
	2 B_{3u}	5·75	4·71
	1 A_g	5·77	5·35
	2 A_g	6·48	6·28
	2 B_{1g}	6·04	5·60
	3 B_{1g}	6·98	6·74
	3 A_g	7·14	6·47
Acene 2	B_{2u}	3·15	1·69
	B_{3u}	3·59	3·60
Acene 3	B_{2u}	2·47	1·15
	B_{3u}	3·27	3·32
Acene 4	B_{2u}	2·12	0·75
	B_{3u}	1·78	3·21
Acene 5	B_{2u}	1·78	0·47
	B_{3u}	2·97	3·08
Phenylpyrrole	$1B_2$	4·37	3·13
	$1A_1$	4·43	3·26
	$2B_2$	4·74	4·08
	$2A_1$	5·24	4·31
DMABN	B	4·07	
	A	2·43	2·80
Butadiene/	B_u	5·42	2·08
PA 2	A_g	5·99	4·41
Hexatriene/	B_u	4·57	1·59
PA 3	A_g	5·10	3·70
Octatetraene/	B_u	3·97	1·26
PA 4	A_g	4·31	3·15
PA 5	B_u	3·53	1·02
H_2CO	3 A_2	8·59	8·58
	2 A_2	7·54	7·55
	1 B_1	8·45	5·84
	2 B_2	7·28	7·20
	1 A_1	7·23	3·52
	1 B_2	6·31	
	1 A_2	3·10	2·73

Molecule	State	Singlet Emission/eV	Triplet Emission/eV
Ethene	B_{1u}	6·89	2·87
Cyclopropene	B_2	5·21	2·60
	B_1	6·23	6·11
Cyclopentadiene	B_2	4·48	2·05
	A_1	5·56	4·40
Norbornadiene	B_2	3·37	3·37
	A_1	5·32	5·27
Benzene	B_{1u}	4·83	3·95
	E_{1u}	6·81	4·74
	B_{2u}	6·21	5·62
	E_{2g}	7·62	6·97
Furan	B_2	5·64	3·15
	A_1	5·80	4·90
Pyrrole	B_2	5·78	3·60
	A_1	5·77	5·05
Imidazole	1 A'	5·24	3·60
	2 A'	6·00	4·84
	1 A"	4·87	4·85
	3 A'	6·21	5·55
Pyridine	1 A_1	5·14	3·58
	B_1	4·04	3·54
	1 B_2	4·78	4·49
	2 A_1	6·12	4·86
	A_2	3·66	3·63
	2 B_2	5·63	5·92
Tetrazine	B_{3u}	2·18	1·60
	1 A_u	1·64	1·48
	B_{1g}	5·03	4·26
	1 B_{1u}	6·36	3·18
	B_{2u}	4·53	4·21
	B_{2g}	3·63	3·31
	2 A_u	4·65	4·33
	2 B_{1u}	6·27	5·09
Formaldehyde	A_2	3·00	2·61
	A_1	7·04	3·42
Acetone	A_2	2·69	2·63
	A_1	6·25	3·42
Formamide	A"	3·15	3·21
	A'	4·63	3·87
Acetamide	A"	3·12	3·20
	A'	4·30	4·31
Propanamide	A"	3·12	3·21
	A'	4·33	4·34

This complete set of emission energies and their related excited state structures will enable us to perform an assessment of the accuracy of DFT functionals for calculating these properties.

3.6 Application of Benchmark Set to DFT

In the previous section, a benchmark set of singlet and triplet excited state geometries were obtained using the RI-CC2 method. This benchmark set is now used to assess the performance of several DFT approximations, similar to those used in section 2.1.

The assessment of the DFT approximations will be in two parts. The first part will be in a similar way to the use of the ground state/absorption benchmark, using a fixed set of structures (those from the benchmark set, RI-CC2 structures), calculating the excitation energy at each of these structures and comparison of these values to those obtained for the emission energies in the benchmark set. The second part will be an assessment of the DFT approximations ability to predict emission energies, namely an excited state geometry will be calculated for each state using each DFT approximation. The emission energies at each of these structures will be compared to RI-CC2 emission energies.

3.6.1 Influence of TDA-TDDFT on Triplet Instabilities

It has been shown that stability issues can lead to issues in the accurate calculation of absorption energies for triplet states especially when using TDDFT. This issue in the absorptions can be corrected with the use of TDA-TDDFT. Therefore an initial investigation of the influence of the use of TDA-TDDFT on the calculation of a selection of triplet emission energies was undertaken. This will allow us to see if TDA-TDDFT should be used over TDDFT for the calculation of the rest of the assessment of the functionals.

Using the triplet states of a selection of molecules in the benchmark set will allow us to investigate the influence of TDA-TDDFT on the calculation of the triplet emission energies of the full set. The molecules and states chosen for this investigation are shown in table 51 along with whether they are low or high stability states; the low stability states have a Hartree—Fock stability <2 eV and the high stability states have a Hartree—Fock stability >2 eV following references 85, 89 and 90. Hartree–Fock instabilities are seen when the energy of the Hartree–Fock wavefunction is not the absolute minimum. Further information on stability can be seen in reference 104.

Table 51: Molecules and states included in the investigation of the influence of TDA-TDDFT along with their stability as defined above.

Molecule	State	Stability
Ethene	B_{1u}	Low
Butadiene	B_u	Low
	A_g	Low
Hexatriene	B_u	Low
	A_g	Low
Octatetraene	B_u	Low
	A_g	Low
Cyclopropene	B_2	Low
	B_1	High
Cyclopentadiene	B_2	Low
	A_1	Low
Benzene	B_{1u}	Low
	E_{1u}	High
	B_{2u}	High
	E_{2g}	High
Napthalene	1 B_{2u}	Low
	1 B_{3u}	High
	1 B_{1g}	Low
	2 B_{2u}	High
	2 B_{3u}	High
	1 A_g	High
	2 A_g	High
	2 B_{1g}	High
	3 B_{1g}	High
	3 A_g	High
Furan	B_2	Low
	A_1	High
Pyrrole	B_2	Low
	A_1	High
Imidazole	1A'	Low
	2 A'	High
	A"	High
	3 A'	High
Pyridine	1 A_1	Low
	B_1	High
	1 B_2	High
	2 A_1	High
	A_2	High
	2 B_2	High
Tetrazine	B_{3u}	Low
	1 A_u	High
	B_{1g}	High
	1 B_{1u}	Low
	B_{2u}	Low
	B_{2g}	High
	2 A_u	High
	2 B_{1u}	High
Formaldehyde	A_2	High
	A_1	Low
Acetone	A_2	High
	A_1	Low
Formamide	A"	High
	A'	Low
Acetamide	A"	High
	A'	High
Propanamide	A"	High
	A'	High

The preliminary investigation was undertaken using PBE0, B3LYP, LC-ωPBE and CAM-B3LYP. The results of the DFT emissions at the RI-CC2 structures was compared to the CC2 emission energies when using both TDDFT and TDA-TDDFT. The results for these are visualised in figures 31 and 32 respectively.

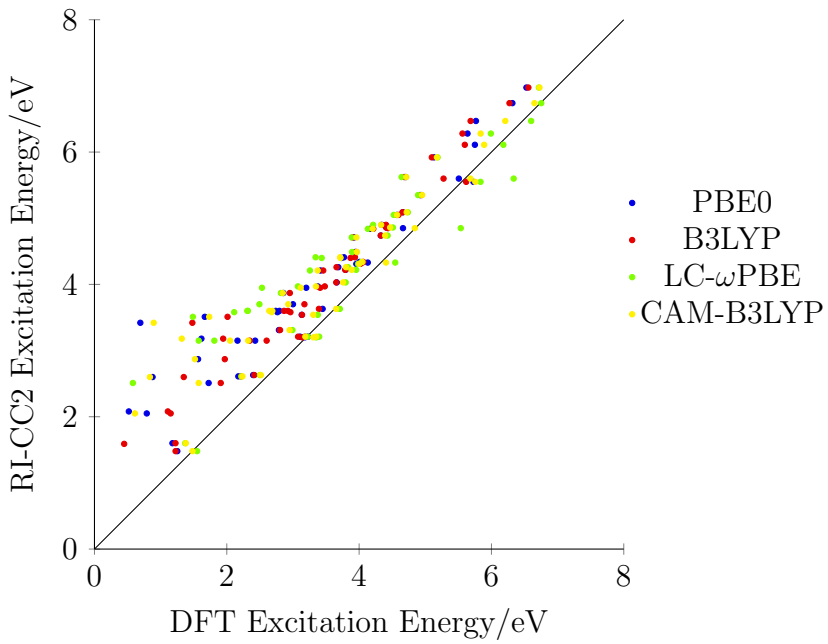


Figure 31: Plot of RI-CC2 emission energy against DFT emission energy for the preliminary investigation using TDDFT.

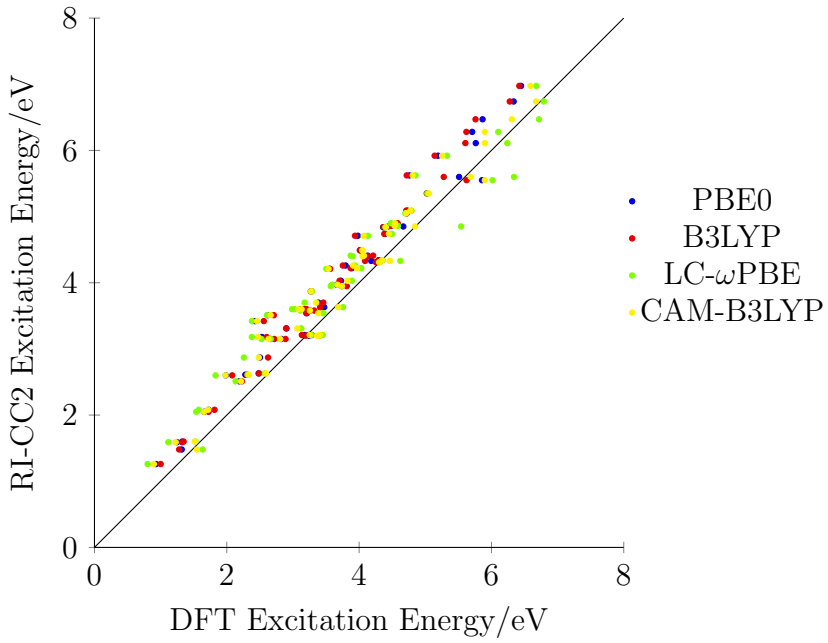


Figure 32: Plot of RI-CC2 emission energy against DFT emission energy for the preliminary investigation using TDA-TDDFT.

When comparing the two plots it is plain to see that the emissions which are lower in energy with RI-CC2 move closer to the diagonal when using TDA-TDDFT compared to TDDFT. In order to get a clearer view of the influence of using TDA-TDDFT on the low stability states, the set will be split following the stabilities in

table 51. The results for the low stability states are shown in figures 33 and 34 for TDDFT and TDA-TDDFT respectively. The results for the high stability states are shown in figures 35 and 36 for TDDFT and TDA-TDDFT respectively.

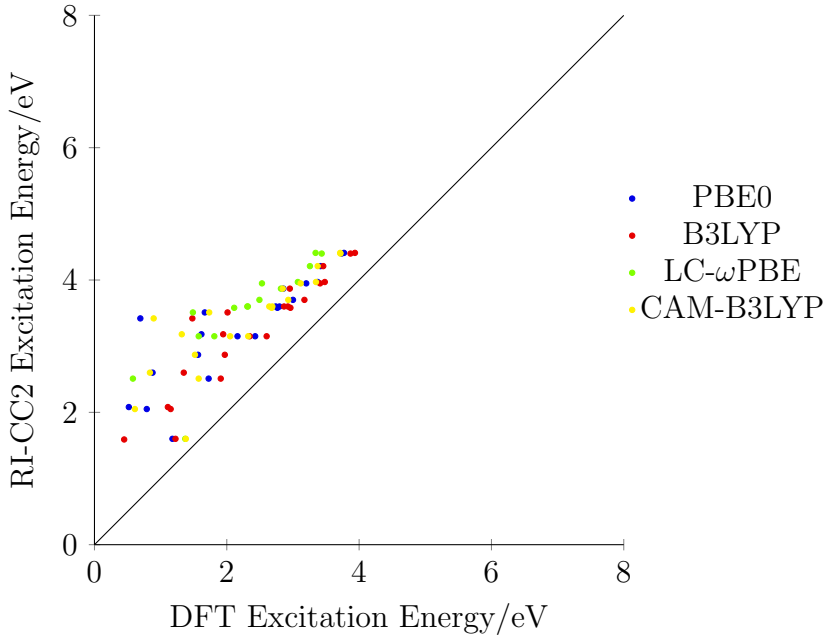


Figure 33: Plot of RI-CC2 emission energy against DFT emission energy for the low stability states using TDDFT.

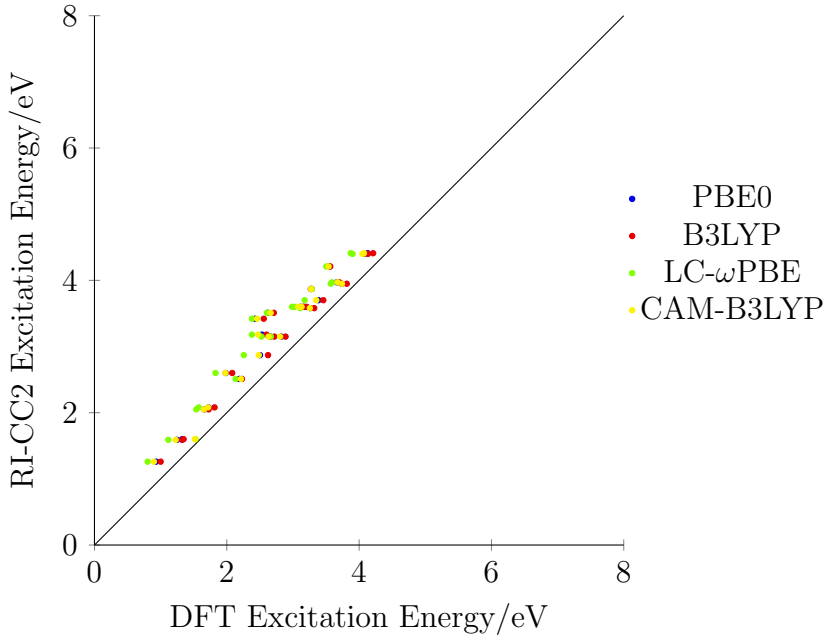


Figure 34: Plot of RI-CC2 emission energy against DFT emission energy for the low stability states using TDA-TDDFT.

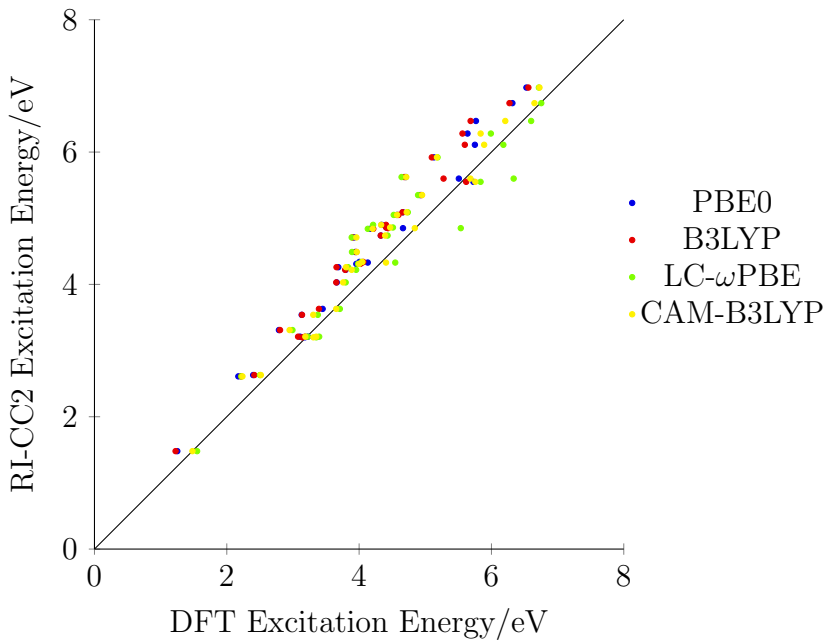


Figure 35: Plot of RI-CC2 emission energy against DFT emission energy for the high stability states using TDDFT.

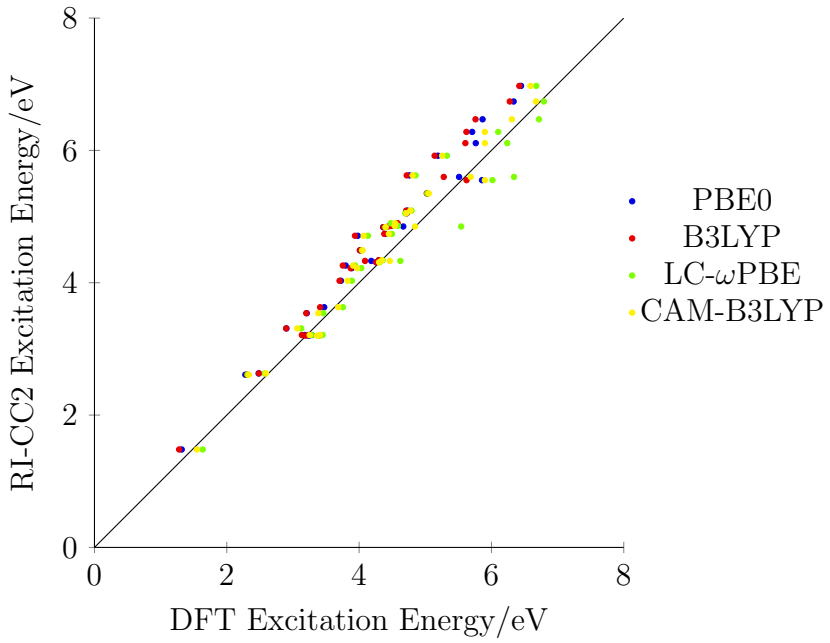


Figure 36: Plot of RI-CC2 emission energy against DFT emission energy for the high stability states using TDA-TDDFT.

The influence of TDA-TDDFT is clear to see on the low stability states where there is a clear shift towards the diagonal between figures 33 and 34. There is little to no change seen in the plots for the high stability states. This follows what has been previously seen in literature for absorptions[85,89,90]. Therefore TDA-TDDFT

will be used to calculate all DFT excitations in the subsequent assessment due to having a large correcting influence on the low stability states whilst having no effect on the high stability states, meaning it will not cause large errors in these states.

3.6.2 Emission energies using DFT at RI-CC2 Structures

The excited state structures obtained for the benchmark set were used to calculate DFT emission energies using the same basis set as those used for the benchmark calculations. TDA-TDDFT will be employed in order to counteract any stability issues as discussed above. The benchmark is split into two parts, singlet and triplet excitations. The DFT approximations used were PBE, PBE0, PBE50 (a global hybrid based on PBE with 50% Hartree–Fock exchange (HFx)), LC-ωPBE, BLYP, B3LYP, BHHLYP and CAM-B3LYP. These DFT approximations were chosen to investigate the influence of both the inclusion of HFx (with the change from pure DFT to hybrids, both global and range-separated) and the influence of two different base approximations (PBE vs BLYP).

3.6.2.1 Singlet Excitations

There are a total of 87 singlet states in the benchmark set. The RI-CC2 structures for each state were used to calculate the DFT emission energy for each state. The emission values for each state using each DFT approximation are then subtracted from the RI-CC2 reference values to give errors. This enables the generation of mean errors, mean absolute errors and standard deviations for the set of singlet excitations when using DFT approximations at the RI-CC2 structures. The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the 8 functionals are shown in table 52.

Table 52: The mean error, mean absolute error and standard deviations for each DFT approximations at the RI-CC2 singlet state structures, all values in eV

	PBE	PBE0	PBE50	LC-ωPBE	BLYP	B3LYP	BHHLYP	CAM-B3LYP
d	−0·36	0·13	0·61	0·64	−0·41	−0·01	0·50	0·31
$\|d\|$	0·48	0·28	0·62	0·65	0·52	0·28	0·58	0·35
σ	0·47	0·38	0·56	0·54	0·50	0·41	0·56	0·41

PBE Based Approximations

It can be seen that for singlet emissions using PBE based functionals at the RI-CC2 structures that the best performing functional is PBE0, with a mean absolute error of 0.28 eV. This is different to the performance of the functionals for absorption excitations where LC-ωPBE performs best. LC-ωPBE is the worst performing

functional for the emission energies at the CC2 structures, with a mean absolute error of 0.65 eV. However this may not be the case once the relevant DFT structures are used. As the reuslts here are showing how accurate the DFT approximation is at reproducing the RI-CC2 emission at the RI-CC2 excited state structure which may not be close the DFT excited state structure.

PBE underestimates the singlet emissions on average and the inclusion of HFx with the hybrid functionals leads an overestimation of the singlet emissions on average.

BLYP Based Approximations

It can be seen that for the singlet emissions using BLYP based functionals at the RI-CC2 structures that the best performing functional is B3LYP, with a mean absolute error of 0.28 eV. This is again follows the trend seen with PBE based functionals, however the range-separated functional, CAM-B3LYP, is the second best performing functional of the BLYP based functionals, with a mean absolute error of 0.35 eV.

BLYP underestimates the singlet emission energies on average, the inclusion of HFx increases the calculated emission energies on average, with more positive mean errors between BLYP and the hyrbid functionals.

Inclusion of Hartree–Fock Exchange

The inclusion of HFx leads to a increase in the singlet emissions calculated on average, with the pure DFT functionals having an underestimation of the singlet emissions on average (negative mean error) and the hybrids increase the predicted values of the emission energies (with more positive mean errors).

The inclusion of a limited percentage of HFx is shown as important as PBE0 (25%), B3LYP (20%) and CAM-B3LYP (19-65% scaling) are the best performing functionals, with mean absolute errors of 0.28 eV, 0.28 eV and 0.35 eV respectively. However if too much HFx is included the error will increase as seen with the two global hybrid functionals with 50% HFx (PBE50 and BHHLYP) included along with LC-ωPBE having the largest errors, with mean absolute errors of 0.62 eV, 0.65 eV and 0.58 eV respectively.

The pure DFT functionals (PBE and BLYP) perform somewhat better than functionals with large percentages of HFx but they are outperformed by functionals with a lower percentage of HFx included. This highlights the importance for a careful consideration of the percentage of HFx included when investigating singlet emissions, as some is necessary but too much is detrimental.

3.6.2.2 Triplet Excitations

There are a total of 85 triplet states in the benchmark set. The RI-CC2 structures for each state were used to calculate the DFT emission energy for each state. The emission values for each state using each DFT approximation are then subtracted from the RI-CC2 reference values to give errors. This enables the generation of mean errors, mean absolute errors and standard deviations for the set of triplet excitations when using DFT approximations at the RI-CC2 structures. The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the 8 functionals are shown in table 53.

Table 53: The mean error, mean absolute error and standard deviations for each DFT approximations at the RI-CC2 triplet state structures, all values in eV

	PBE	PBE0	PBE50	LC-ωPBE	BLYP	B3LYP	BHHLYP	CAM-B3LYP
d	−0·45	−0·31	−0·21	−0·21	−0·46	−0·33	−0·21	−0·23
$\|d\|$	0·47	0·36	0·43	0·38	0·48	0·35	0·43	0·32
σ	0·31	0·28	0·48	0·41	0·32	0·24	0·49	0·31

PBE Based Approximations

It can be seen that for triplet emissions using PBE based functionals at the RI-CC2 structures that the best performing functional is PBE0, with a mean absolute error of 0.36 eV. This is closely followed by LC-ωPBE, with a mean absolute error of 0.38 eV. This is a different trend to that seen in the singlet excitations where LC-ωPBE is the worst performing functional. All the PBE based functionals underestimate the triplet emissions on average, with all having negative mean errors.

BLYP Based Approximations

It can be seen that for the triplet emissions using BLYP based functionals at the RI-CC2 structures that the best performing functional is CAM-B3LYP, with a mean error of 0.32 eV. Again B3LYP is close in accuracy to CAM-B3LYP, with a mean absolute error of 0.35 eV, as is seen in the PBE based functionals with both the global and range-separated hybrid functionals performing similarly for triplet emissions at the RI-CC2 structure. All the BLYP based functionals underestimate the triplet emissions on average, with all having negative mean errors.

Inclusion of Hartree–Fock Exchange

The inclusion of HFx is shown as important as PBE0 (25%), B3LYP (20%), LC-ωPBE (0-100% scaling) and CAM-B3LYP (19-65% scaling) are the best performing functionals, with mean absolute errors of 0.36 eV, 0.35 eV, 0.38 eV and 0.32 eV

respectively. However if too much HFx is included the error will increase as seen with the two global hybrid functionals with 50% HFx (PBE50 and BHHLYP) having large errors, with mean absolute errors of 0.43 eV and 0.43 eV respectively. The pure DFT functionals are the worst performing functionals for the triplet emissions at the RI-CC2 structures, highlighting the importance for the inclusion of HFx.

3.6.3 Emission Energies using DFT at DFT Structures

The values calculated in the previous section were all at the RI-CC2 structures, this gives an idea of the accuracy of the DFT approximations at a fixed structure however the RI-CC2 structures may not be the excited state structures for the DFT approximations used. Therefore an excited state structure optimisation and subsequent emission energy calculation gives a more realistic idea of the accuracy of the DFT approximations for the calculation of emission energies as there may not be a fixed/benchmark structure to start from in the calculation of emissions for systems not present in the benchmark set.

The same 8 functionals are investigated, again along with the differences between the singlet and triplet states. The basis set used in each case is the same as that used in the benchmark set. TDA-TDDFT will be employed in order to counteract any triplet instabilities that are present in the triplet states. The emission values calculated at the DFT structures will be compared to the emission energies in the benchmark set using RI-CC2.

LC-ωPBE structures are unfeasable to obtain due to exceptional computational cost. Thus a smaller investigation will be conducted using the CAM-B3LYP structures and the PBE0 structures.

3.6.3.1 Singlet Excitations

There are a total of 87 singlet states in the benchmark set. The emission values for each state using each DFT approximation are then subtracted from the RI-CC2 reference values to give errors. This enables the generation of mean errors, mean absolute errors and standard deviations for the set of singlet emissions when using DFT approximations. The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the 8 functionals are shown in table 54.

Table 54: The mean error, mean absolute error and standard deviations for each DFT approximations at DFT singlet state structures, all values in eV

	PBE	PBE0	PBE50	BLYP	B3LYP	BHHLYP	CAM-B3LYP
d	−0·43	0·18	0·74	−0·47	0·07	0·57	0·40
$\|d\|$	0·69	0·51	0·81	0·67	0·48	0·78	0·54
σ	0·98	0·80	0·71	0·78	0·81	0·86	0·62

PBE Based Approximations

It can be seen that for singlet emissions using PBE based functionals at the DFT structures that the most accurate functional is PBE0, with a mean absolute error of 0.51 eV. This is similar to the trend shown at the RI-CC2 structures for singlet emissions, with PBE0 performing best followed by PBE and then PBE50.

PBE underestimates the singlet emissions on average and the inclusion of HFx with the hybrid functionals leads an overestimation of the singlet emissions on average, following the trend observed at the RI-CC2 structures.

BLYP Based Approximations

It can be seen that for the singlet emissions using BLYP based functionals at the DFT structures that the best performing functional is B3LYP, with a mean absolute error of 0.48 eV. This is again follows the trend seen with PBE based functionals, however the range-separated functional, CAM-B3LYP, is the second best performing functional of the BLYP based functionals, with a mean absolute error of 0.54 eV, again following the trend seen at the RI-CC2 structures. The pure DFT functional, BLYP, underestimates the singlet emissions on average and the hybrid functionals over estimate the singlet emissions on average. BLYP shows underestimated singlet emission energies on average, with a negative mean error. The hybrid functionals lead to an overestimation of the singlet emission energies on average, with positive mean errors. This follows the trend observed at the RI-CC2 structures.

Inclusion of Hartree–Fock Exchange

The inclusion of HFx leads to a increase in the singlet emissions calculated on average, with the pure DFT functionals having an underestimation of the singlet emissions on average (negative mean error) and the hybrids increase the predicted values of the emission energies (with increasing mean errors).

The inclusion of a limited percentage of HFx is shown as important as PBE0 (25%), B3LYP (20%) and CAM-B3LYP (19-65% scaling) are the best performing

functionals, with mean absolute errors of 0.51 eV, 0.48 eV and 0.54 eV respectively. However if too much HFx is included the error will increase as seen with the two global hybrid functionals with 50% HFx (PBE50 and BHHLYP) included having the largest errors, with mean absolute errors of 0.81 eV and 0.78 eV respectively.

The pure DFT functionals (PBE and BLYP) perform somewhat better than functionals with large percentages of HFx but they are outperformed by functionals with a lower percentage of HFx included. This highlights the importance for a careful consideration of the percentage of HFx included when investigating singlet emissions, as some is necessary but too much is detrimental. This is the same observation seen in the singlet emissions at the RI-CC2 structures.

3.6.3.2 Triplet Excitations

There are a total of 85 triplet states in the benchmark set. The emission values for each state using each DFT approximation are then subtracted from the RI-CC2 reference values to give errors. This enables the generation of mean errors, mean absolute errors and standard deviations for the set of triplet emissions when using DFT approximations. The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the 8 functionals are shown in table 55.

Table 55: The mean error, mean absolute error and standard deviations for each DFT approximations at DFT triplet state structures, all values in eV

	PBE	PBE0	PBE50	BLYP	B3LYP	BHHLYP	CAM-B3LYP
d	−0·41	−0·20	−0·04	−0·53	−0·25	−0·12	−0·13
$\|d\|$	0·44	0·31	0·33	0·57	0·34	0·38	0·29
σ	0·45	0·34	0·47	0·67	0·32	0·54	0·37

PBE Based Approximations

It can be seen that for triplet emissions using PBE based functionals at the DFT structures that the most accurate functional is PBE0, with a mean absolute error of 0.31 eV. This is similar to the trend shown at the RI-CC2 structures for triplet emissions, with PBE0 performing best followed by PBE50 and then PBE. All the PBE based functionals underestimate the triplet emissions on average, with all having negative mean errors.

BLYP Based Approximations

It can be seen that for the triplet emissions using BLYP based functionals at the DFT structures that the best performing functional is CAM-B3LYP, with a mean

absolute error of 0.29 eV. This again follows the trend seen at the RI-CC2 structures. The pure DFT functional, BLYP, is the worst performing functional, with a mean absolute error of 0.57 eV. All the BLYP based functionals underestimate the triplet emissions on average, with all having negative mean errors.

Inclusion of Hartree–Fock Exchange

The inclusion of HFx is shown as important as all the hybrid functionals having similar mean absolute errors and lower mean absolute errors than the pure DFT functionals. The pure DFT functionals are again the worst performing functionals for the triplet emissions at the RI-CC2 structures, highlighting the importance for the inclusion of HFx.

3.6.4 Overall Functional Performance

In order to get a clearer view of the overall performances of each functional tested here over the two assessments with the benchmark set, a ranking of the performance with a range of 0-7 for the RI-CC2 structures and a range of 0-6 for the DFT structures due to issues with LC-ωPBE is shown in table 56. The issues with LC-ωPBE is discussed in section 3.6.5, the computational cost of performing the excited state geometry optimisations using LC-ωPBE was prohibitively high.

Table 56: Ranking of the performance of the functionals tested across the series of emission energy tests.

Property	PBE	PBE0	PBE50	LC-ωPBE	BLYP	B3LYP	BHHLYP	CAM-B3LYP
Singlet @ CC2	3	0	6	7	4	0	5	2
Triplet @ CC2	6	2	4	3	7	1	4	0
Singlet @ DFT	4	1	6		3	0	5	2
Triplet @ DFT	5	1	2		6	3	4	0

This makes it clear that the inclusion of HFx into functionals in necessary for accurate performance when calculating emission energies, with the best performing 3 functionals are PBE0, B3LYP and CAM-B3LYP. However it also shows that it is important to be careful with the percentage of HFx included in the functional as the two global hybrids with 50% HFx included (PBE50 and BHHLYP) perform poorly generally, with ranking similar to the two pure functionals which are shown to be poorly performing.

3.6.5 LC-ωPBE at CAM-B3LYP and PBE0 Structures

Due to technical issues in obtaining LC-ωPBE excited state structures leading to an exceptionally high computational cost. It was decided to use the CAM-B3LYP and PBE0 structures of a small part of the benchmark and calculate the LC-ωPBE emission energies at these structures. These emission energies are then compared to the RI-CC2 emission energies along with the emission energies for the two density functionals which the excited state structures are from.

The states that are used in this smaller investigation are show in table 57. They will have both the singlet and triplet states investigated.

Table 57: The states that will be included in the smaller investigation of LC-ωPBE at CAM-B3LYP and PBE0 structures

Molecule	State
Napthalene	$1B_{2u}$, $1B_{3u}$
Acene 2	B_{2u}, B_{3u}
Butadiene	B_u, A_g
Hexatriene	B_u, A_g
Pyrrole	B_2, A_1
Acetone	A_2, A_1
Formaldehyde	A_2, A_1

The mean error (d), mean absolute error ($|d|$) and standard deviations (σ) for the smaller set of states are shown in table 58 for the LC-ωPBE emissions at the PBE0 and CAM-B3LYP structures along with the error values for the PBE0 and CAM-B3LYP emissions at their respective structures.

Table 58: The mean error, mean absolute error and standard deviations for the investigation of LC-ωPBE emission energies at DFT structures, all values in eV

	LC-ωPBE at PBE0	LC-ωPBE at CAM-B3LYP	PBE0	CAM-B3LYP
Singlet d	0·75	0·72	0·29	0·44
Singlet $\|d\|$	0·75	0·72	0·30	0·44
Singlet σ	0·45	0·45	0·29	0·33
Triplet d	−0·19	−0·20	−0·17	−0·14
Triplet $\|d\|$	0·28	0·30	0·24	0·24
Triplet σ	0·29	0·30	0·24	0·27

Singlet Emissions

The singlet emissions with LC-ωPBE are overstimated as is seen at the RI-CC2 structures. The overestimation is slightly larger at the PBE0 structures but is similar across both the sets of structures from the DFT functionals. The performance of PBE0 and CAM-B3LYP is much better than that for LC-ωPBE, which again follows the trend seen at the RI-CC2 structures. This again highlights the importance for the careful selection on the percentage of HFx included in a functional when calculating accurate singlet emission energies.

Triplet Emissions

The triplet emissions with LC-ωPBE are underestimated as is seen at the RI-CC2 structures. The performance of PBE0 and CAM-B3LYP is again better than that for LC-ωPBE, however the performance of LC-ωPBE for triplet emission energies is much closer to that of PBE0 and CAM-B3LYP. This again follows the trend seen at the RI-CC2 structures.

3.7 Further Work with and beyond the Benchmark Set

The benchmark set obtained here can be taken further. The further use/development of the benchmark set can be split into two obvious areas, namely using the set and going beyond the set.

In the use of the benchmark set it would be possible to use higher methods, for example CCSD(T), at the RI-CC2 structures obtained here. This would enable another verification of the quality of the structure obtained here. It would also be possible to investigate the energetics of the excited states under the lens of the dependence of structures. This could be done via calculated relative energies of the states (emission energy + energy of excited state structure compared to the ground state energy). Another method of doing this would be investigating each state of the molecule in the benchmark set at each structure to investigate how the structural changes seen between the excited state structures of each state affects the other states in the molecules. Other excited state properties can be investigated other than the energies and structures which have been the focus here such as the dipoles. In order to go beyond the set, the set currently is limited to lighter atoms so an obvious avenue would be the inclusion of some heavier atoms such as sulphur or phosphorous, in order to investigate the influence of these heavier atoms on excited state structures and energies. Also it is a possible avenue to go beyond the structures/molecules currently present in the set to structures/molecules that are more chemically useful/relevant such as various common ligands/building blocks for ligands.

3.8 Conclusions

The benchmark set of excited state structures and emission energies developed here is attempting to fill noticeable gap in literature. As nothing similar to benchmark sets seen in literature for ground state properties (those included in the GMTNK55 database) exists for these properties. The need for this benchmark set is due to it being difficult to experimentally obtain data (short lived etc), with the exception of diatomics.

Diatomics are used as a preliminary study to highlight issues that may be faced when looking at more complicated systems. However, they are of limited use as diatomics can be brute forced (due to one possible coordinate) to get the correct behaviour of each state. The diatomics studied showed that it is important to use basis sets with diffuse functions when considering higher energy excitations (>9 eV). Another issue it highlighted was that there can be a potential issue when states cross during optimisation which will require careful monitoring to ensure the correct state is followed during the excited state geometry optimisation when developing the benchmark set and when using the benchmark set to assess approximate E_{xc}s.

Selection of molecules based on existing benchmark sets for absorption, this includes a variety of small organic molecules, including 2 groups of closely related molecules (polyacetylene and acene groups).

The method used (RI-CC2) has been validated against higher level theory or experimentally derived results for some molecules and states from literature. RICC2 shows good performance for these states along with existing use of RI-CC2 as reference results for a molecule in the set[101]. Once the method was validated, excited state geometries and emission values were calculated for each state in the benchmark set, combined to produce the benchmark set. The benchmark set developed here can and should be expanded upon. This could happen via higher level theory being used to calculate the emission energies at the structures obtained here or by expanding the set to include a greater diversity of molecules (include molecules which contain heavier atoms such as sulphur or phosphorous).

An assessment of the performance of a series of approximate E_{xc}s for emission energies/excited state geometries was undertaken. Singlet emissions are much more dependent of the quantity of HFx included in the functional that the triplet emissions. The trends observed are preserved between at the benchmark structures and at the structures obtained via excited state geometry optimisations using the related approximate E_{xc}. However there seems to be no approximate E_{xc} of those tested which has a great performance for the emission energies from the benchmark set either at the structures from the benchmark set or from structures

optimised using the approximate E_{xc}. One suggestion to take from this is that when approximate E_{xc}s are developed using ground state properties they are not accurate for excited state properties. This idea extends further as range separated hybrids perform better for absorptions than global hybrids. However, this is not the case for emission energies and thus excited state geometries. This means that the methods of improving the accuracy of excited state energies in the ground state do not necessarily improve the accuracy of excited state properties away from the ground state structure. Therefore a fresh view of developing approximate E_{xc}s for the calculation of excited state properties is needed. This could potentially come in different ways, one obvious suggestion is to use the newly generated benchmark set of excited state geomteries and emission energies when developing approximate E_{xc} as an additional exact data for approximate E_{xc} to reproduce.

4 Porous Materials

Porous materials are a growing field in material science. They are noted for their large surface areas. They have several applications such as gas storage, catalysis and separation[105]. However these materials are amorphous leading to issues investigating the properties of them theoretically with DFT, due to a lack of crystal structures to use as a starting point. This issue need to be bypassed via the use of model systems which need to include as much information as possible without becoming so computationally expensive as to be unfeasible to use. Two such materials will be discussed and will have model systems generated for use in theoretical investigations of their excited state properties. Further information about the challenges in investigating porous materials can be found in reference 105.

4.1 PAF-1

Porous aromatic frameworks (PAF) are a series of frameworks that are noted for their high surface areas along with good physicochemical stability[106]. PAFs are characterised by a rigid aromatic structure constructed via covalent bonds[106]. The pores contained in the structure are wormlike with a uniform pore size[106]. The physicochemical stability is due to the covalent bonds which make up the framework[106].

PAF-1 is the first PAF that was synthesised in 2009[107]. PAF-1 shows long range order along with, to a certain extent, amorphous nature and a high surface area ($S_{BET} = 5640\ m^2g^{-1}$). A schematic of the building block for PAF-1 is shown in figure 37.

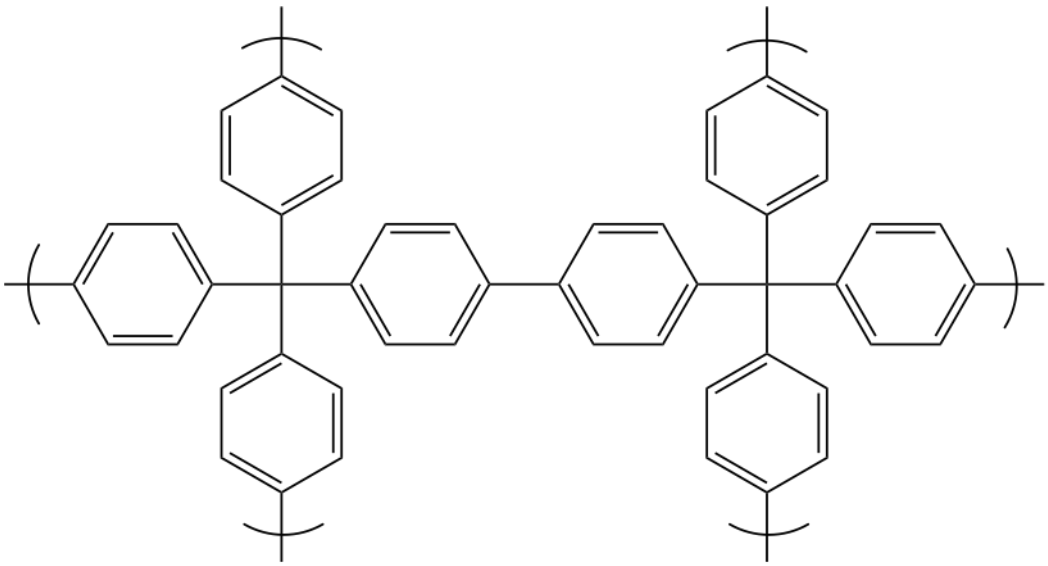


Figure 37: Building Block of PAF-1

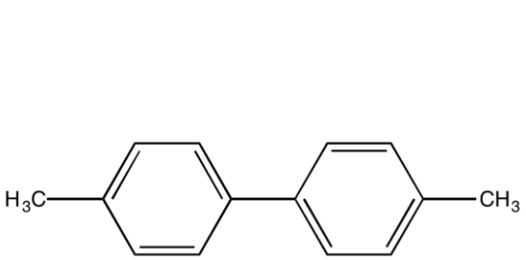
The addition of high concentrations of H_2SO_4 to a suspension of PAF-1 and subsequent drying led to a colour change from white (PAF-1) to blue violet (H_2SO_4@PAF-1), corresponding to a UV-Vis peak appearing at 608 nm. The colour change remains consistent with the addition of H_3PO_4 and HCl. The underlying cause of this colour change must be due to protonation as it is consistent between the different acids.

The investigation that has been undertaken using DFT was to explain the underlying cause of this colour change. DFT was used to identify the protonation of the PAF-1 system and subsequently TDDFT was used to investigate and explain the colour change that is caused by the protonation.

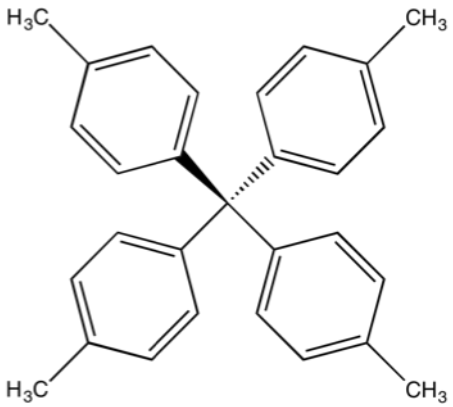
4.1.1 Model Choice

The PAF-1 framework is amorphous thus there is not a way to obtain an experimental structure thus a model molecule must be used to represent the framework. This model must be chosen to explore the particular property that is being investigated, in this case the colour change on protonation.

From the structure of the repeat unit of PAF-1, there are two obvious model molecules. These are shown in figure 38, one is a tetraphenyl based model and the other is biphenyl based.



Biphenyl



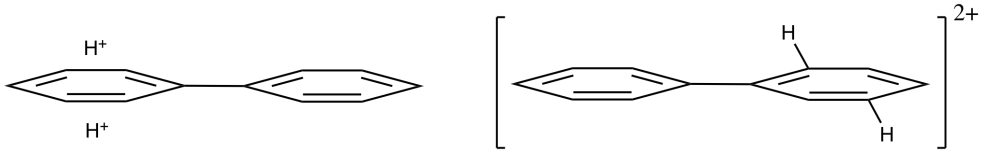
Tetraphenyl

Figure 38: Proposed models of PAF-1

The property being looked at is likely due to a conjugated π system. The π system in the tetraphenyl model stops at the central carbon so is therefore isolated on each individual benzene ring (which is actually a biphenyl in the PAF framework), whereas for the biphenyl model it continues over both benzene rings. Thus the natural choice is the biphenyl model. This is the model molecule that will be used for further investigation.

4.1.2 Protonation of model system

There are two obvious methods of protonation for the biphenyl model system. They are a π bonded system and a σ bonded system. These two methods are illustrated in figure 39.



π-bonded σ-bonded

Figure 39: Protonation models of biphenyl model

These two structures were optimized using B3LYP and 6-31G(d) basis set. The result of this optimisation was initially unexpected. The π system was not stable and optimized to the σ system. Literature shows that the protonation of benzene follows a similar pattern with the σ bonded system being more stable than the π bonded system[108]. Therefore the σ bonded systems will be investigated further.

4.1.3 Energetics of Protonation of model system

A systematic study of the protonation of PAF-1 via a model system based on a biphenyl molecule has been undertaken. This was undertaken to investigate whether it is possible to protonate PAF-1, whether it is possible to protonate multiple times and where in the molecule it is preferential to protonate. The description of the carbons with protons added will follow the naming convention shown in figure 40.

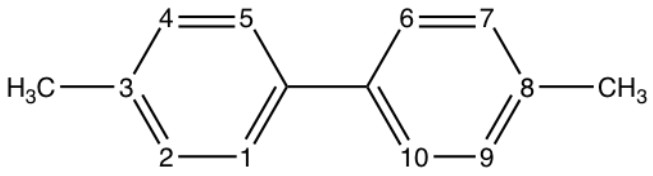


Figure 40: Schematic of proton position naming

The influence of the solvent present in the experimental data was considered using the polarizable continuum model. This solvent model creates a solvent cavity via a set of overlapping spheres. More detail can be found in reference 109

4.1.3.1 One Proton

The addition of a single proton onto the model system is possible and it is stabilised compared to the non-protonated system. For one proton there are three distinct positions where protonation can occur. However there seems to be little preference for which position it will bind to, the largest difference between the positions is ∼0.3 eV. A plot of the normalised energy differences between these 3 positions is shown in figure 41.

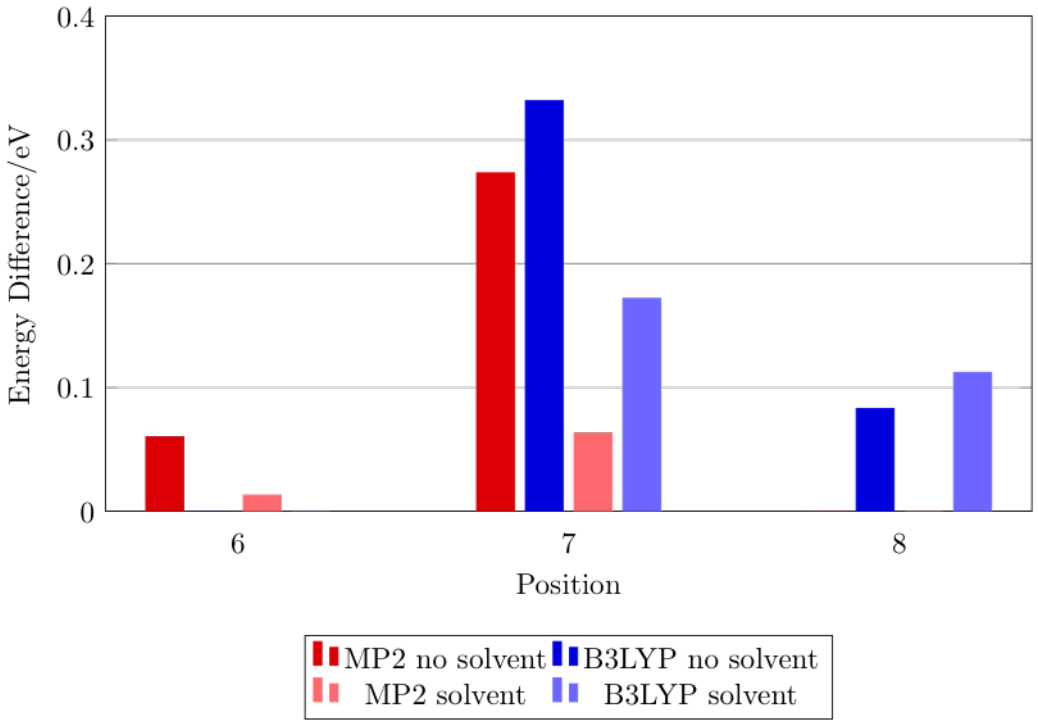


Figure 41: Plot of normalised energy differences for 1 proton

4.1.3.2 Two Protons

There are two possible mechanisms for protonating the model system with two protons; adding both protons to the same ring (2 on 1 ring) and adding the protons to different rings (2 on 2 rings).

Two on One Ring

The addition of two protons remains feasible. There is a much more pronounced preference for the positions (of which there are 6 distinct positions) on which the protons can be added. The difference between the most and least stable positions is greater than 2 eV and the differences in energy are generally larger than for one proton – this is easily accounted for by the charges on the protons being in

proximity. A plot of the normalised energy differences between these 6 positions is shown in figure 42.

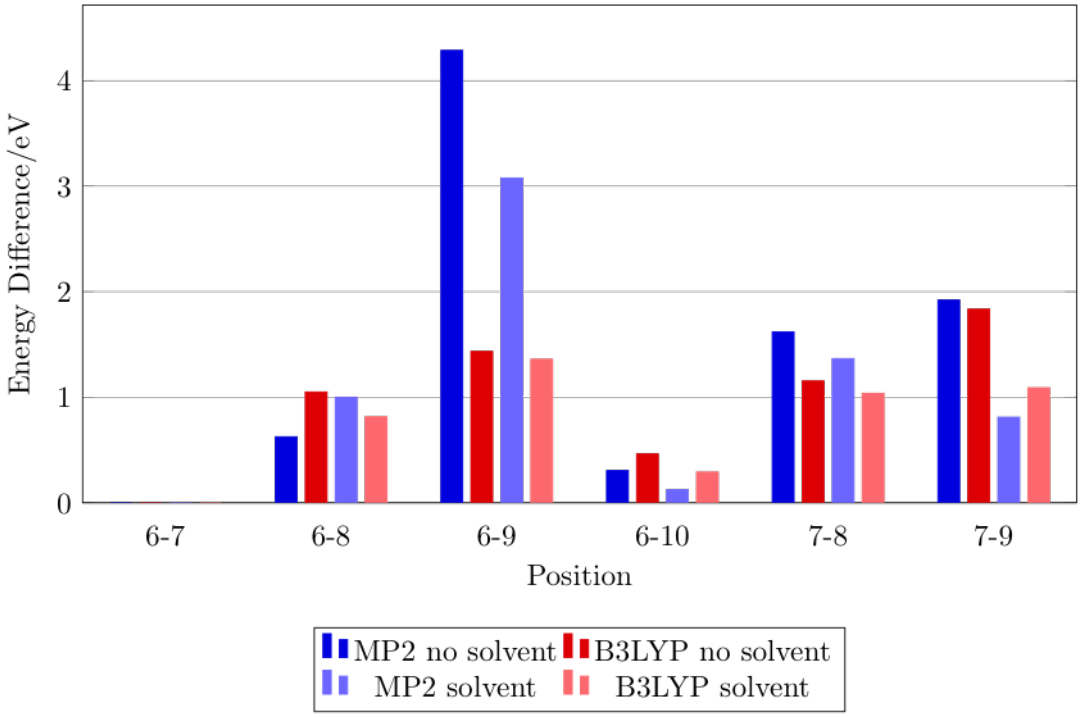


Figure 42: Plot of normalised energy differences for 2 protons on 1 ring

Two on Two Rings

The differences between the possible positions (of which there are 9 distinct positions) are much smaller for two protons on two rings compared to two protons on one ring. This demonstrates that there is a high energetic cost to concentrating the charge added compared to separating the charge that is added. This is as expected as separating the charge should be more favourable. The largest difference between the structures for two on two rings is 0.8 eV. A plot of the normalised energy differences between these 9 positions is shown in figure 43.

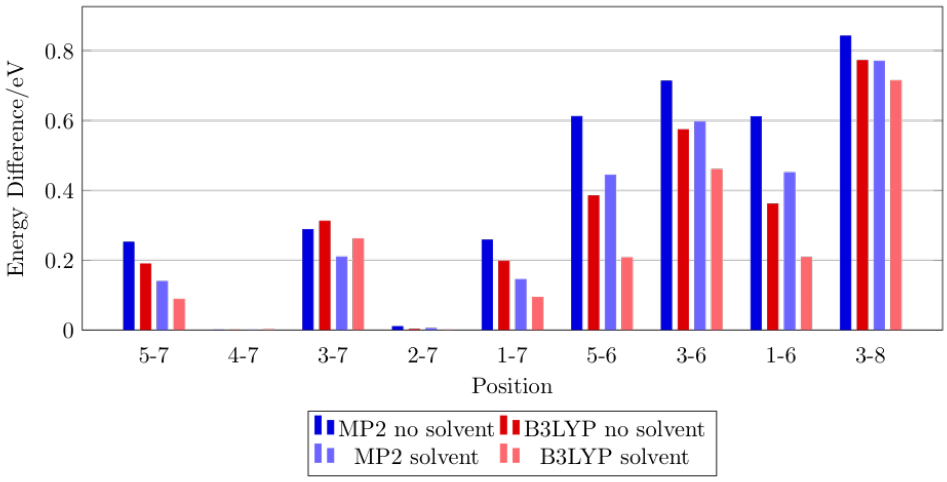


Figure 43: Plot of normalised energy differences for 2 protons on 2 rings

Of the possible positions it seems there is a slight preference for maximising the number of carbons between the protons. The two structures with 5 carbons between the protons are the most stable, then the two with 4 carbons between the protons and then the two structures with 3 carbons between the protons. There are other cases of these which have the proton attached to the carbon which is methylated in the model (to represent the tetrahedral carbon in the PAF-1 framework), these type of structures are generally higher in energy than the ones without when there are the same number of carbons between the protons (4-7/2-7 vs 3-6).

4.1.3.3 Three Protons

The three proton positions are made from two protons on one ring and one proton on the other. There are a much larger number of possible positions with three protons compared to two protons on 2 rings (26 versus 9 respectively). There are large observable differences between the different possible positions. The largest difference is 3.5 eV, this is the largest difference observed between the proton positions. A plot of the normalised energy differences between these 26 positions is shown in figure 44.

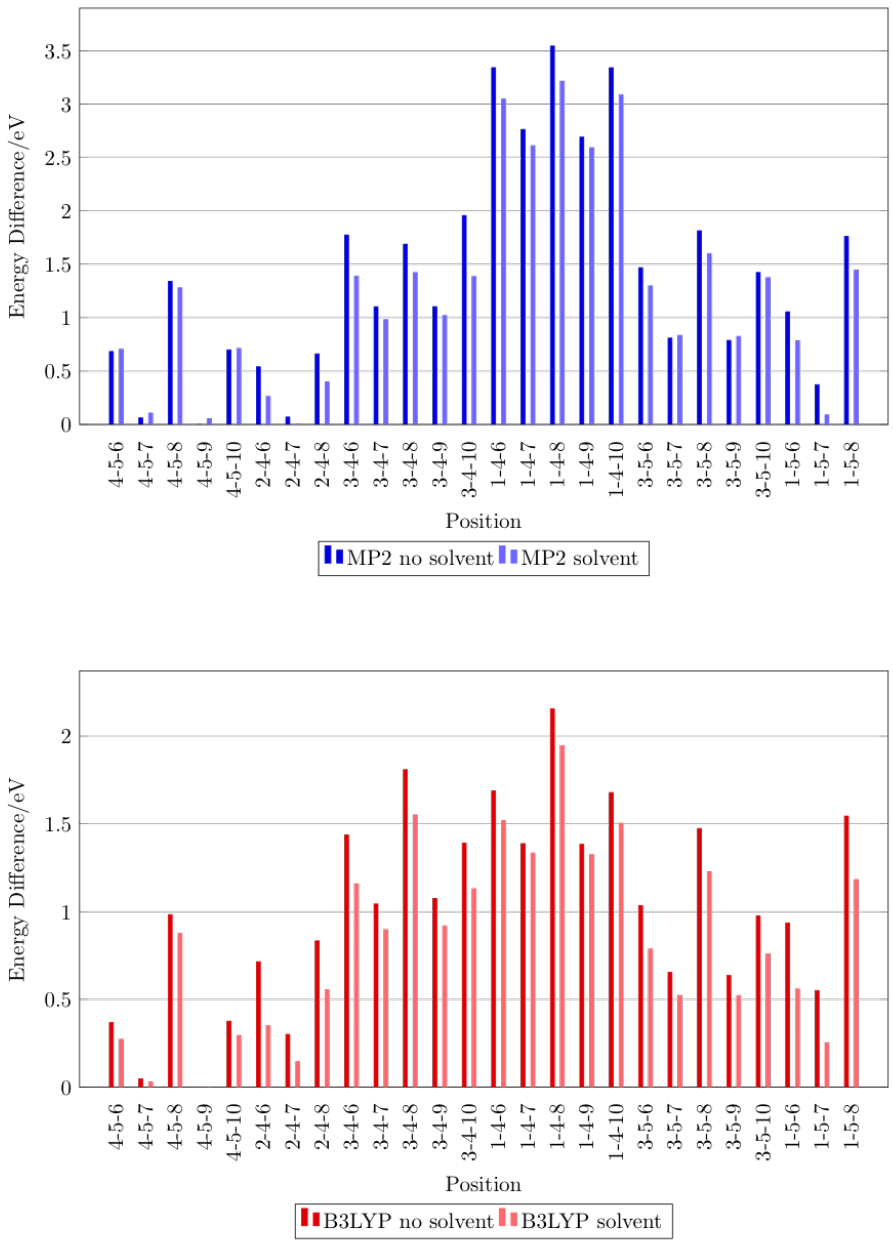


Figure 44: Plot of normalised energy differences for 3 protons

Again there can be seen several different related 'mini-groups' of structures in the full set of structures. These structures have similar energetics when they are related. For example, 4-5-7 and 4-5-9 have almost identical energetics and are obviously related, as there are 4 carbons between the protons closest to each other on the different rings.

4.1.3.4 Four Protons

The model was tested with four protons (two protons on each ring), however there were problems with performing a similar type of systematic study as for the

other number of protons. The model began to fall apart due to there being too much charge for the model to sustain. Therefore the four proton structures were not investigated in more detail and the structures obtained gave a much smaller energy decrease compared to the addition of the one, two and three protons.

4.1.3.5 Baseline Stability of Proton Addition

The most stable positions of those tested for each subsequent proton addition were compared to the neutral biphenyl model molecule to give a stability baseline (figure 45). This baseline shows that the energy decrease for each subsequent proton addition remains similar (~ 0.35 Ha) from 1 to 3 protons (with a slight tailing off with the addition of the first proton having an energy decrease of $\sim$0.4 Ha and the third an energy decrease of ~ 0.32 Ha), however the addition of a fourth proton leads to a significant drop in the energy decrease to $\sim$0.25 Ha.

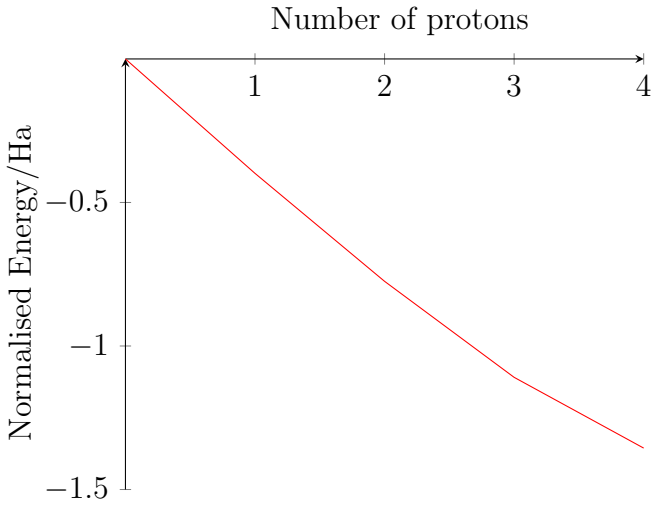


Figure 45: Plot of normalised energy of proton addition

The energetics of the systematic protonation show that it is possible to protonate the framework and it is possible in several positions in each number of protons added to the biphenyl model. The study also shows that there is a limit to the number of protons it is possible to add to the model system.

4.1.4 Excitations of Protonated model system

Continuing on from the systematic investigation of the energetics of the protonation of the model system, an investigation of the electronic excitations of the model system has been undertaken. The motivation for this study was to investigate and explain the colour change seen with the protonation of the PAF-1 system, with a change from white to blue violet and a corresponding UV-Vis peak at 608nm.

The proton positions shown for the energetics are used and calculated the lowest singlet for each position. This has only been done using the solvent structures for each method, as this is closer to the conditions of experiment. Attempting to correlate to the experimental spectra means that excitations of around 1.90 to 2.25 eV (550-650nm) are required to correspond to the experimental spectra. However, the method used (CAM-B3LYP) to calculate the excitations overestimates the excitations of benzene by $\sim$0.4 eV[47], therefore a corrected target range must be used. Therefore the range of excitations that correspond to the experimental values are 2.30 to 2.65 eV.

4.1.4.1 Neutral System

Initially the excitations of the unprotonated model system so a starting point to investigate the influence the protonation of this system has on the electronic excitations can be obtained. The lowest singlet of the model system for B3LYP and MP2 solvated structures were in good agreement with excitations of 5.24 and 5.29 eV respectively. This value would mean that the PAF-1 polymer doesn't absorb in the visible spectrum even including the CAM-B3LYP correction to the excitation energy of benzene. Thus the unprotonated polymer should have no colour/white and this is observed experimentally.

4.1.4.2 One Proton

The addition of the single proton as described above in section 4.1.3.1, showed a drop in the excitation energy as shown in table 59.

Table 59: Excitation of 1 proton structures

Position	B3LYP/eV	MP2/eV
6	3·12	3·23
7	3·12	3·14
8	3·63	3·59

As with the unprotonated structures there is good agreement between the excitations of the structures from the different methods. The addition of protons to the model system has led to a significant drop in the excitation energy of around 2 eV. This suggests that the addition of protons to the model system leads to a drop in the excitation energies. However the addition of a singlet proton does not lead to a large enough decrease to bring the excitation into the target range, it follows that the addition of an increasing number of protons should lead to further drops in the excitation energy.

4.1.4.3 Two Protons

Again as with the energetics there are two ways to protonate the model system with 2 protons; 2 on 1 ring and 2 on 2 rings.

2 on 1 Ring

The addition of a second proton leads to a further decrease in the excitation energies as show in table 60.

Table 60: Excitation of 2 protons on 1 ring structures

Position	B3LYP/eV	MP2/eV
6–7	2·74	2·70
6–8	3·16	2·94
6–9	1·43	1·63
6–10	3·49	3·51
7–8	2·04	2·03
7–9	2·06	2·02

Once again there is good agreement between the excitations of the structures from the different methods. The addition of another proton lowers the excitation energy further. However there are still no excitations in the target region (2.30 – 2.65 eV). The excitations are either too high or too low, the excitations also seem to follow on the energies of the structures, with the lowest excitation energies coming from the least stable structures (6-9 is the least stable structure and it has the lowest excitation energies).

2 on 2 Ring

The excitation energies from the structures with two protons on two rings are shown in table 61.

Table 61: Excitation of 2 protons on 2 rings structures

Position	B3LYP/eV	MP2/eV
5–7	3·66	3·76
4–7	3·96	3·98
3–7	4·10	4·12
2–7	3·97	3·97
1–7	3·60	3·75
5–6	3·13	3·15
3–6	3·58	3·62
1–6	3·04	3·12
3–8	3·89	4·00

Again there is good agreement between the excitations of the structures from the different methods. The addition of the second proton to the other ring to which the first was on, leads to a decrease compared to the unprotonated system but not as large a decrease as placing both protons on the same. This means that the excitations are still not in the target region.

The 2 proton positions give some hints on the way to obtain excitations that are potentially low enough in energy to obtain the target region. It appears as though uneven protonation lowers the excitation energy more than even protonation.

4.1.4.4 Three Protons

The excitaition energies from the structures with three protons added using the B3LYP structures are shown in table 62. Only the B3LYP structures are used due to the number of structures with three protons and the previous excitations show that there is a good agreement between the excitations of the B3LYP and MP2 structures.

Table 62: Excitation of 3 proton structures

Position	B3LYP/eV
4–5–6	3·04
4–5–7	3·29
4–5–8	2·54
4–5–9	3·33
4–5–10	3·04
2–4–6	3·24
2–4–7	3·26
2–4–8	3·51
3–4–6	2·42
3–4–7	2·52
3–4–8	2·35
3–4–9	2·46
3–4–10	2·47
1–4–6	2·10
1–4–7	2·24
1–4–8	3·51
1–4–9	2·26
1–4–10	2·10
3–5–6	3·26
3–5–7	3·77
3–5–8	3·23
3–5–9	3·77
3–5–10	3·23
1–5–6	3·16
1–5–7	3·79
1–5–8	3·01

The addition of the third proton leads to a further decrease in the excitation energies compared to the unprotonated model. With this protonation scheme we get some excitations in the target region ($\sim$ 2.30 to 2.65 eV), there are eight excitations in this region. These eight excitations and their oscillator strengths, in brackets, are shown in table 63.

Table 63: Excitations in target region

Position	B3LYP/eV oscillator strengths in brackets
4–5–8	2.54 (0.04)
3–4–6	2.42 (0.56)
3–4–7	2.52 (0.46)
3–4–8	2.35 (0.01)
3–4–9	2.46 (0.35)
3–4–10	2.47 (0.62)
1–4–7	2.24 (0.72)
1–4–9	2.26 (0.64)

The excitations with the low oscillator strengths (positions 4–5–8 and 3–4–8) are probably not going to be seen due to low probability of the transition occurring. With the excitations in the target region, we have visualized the excitation by generating density different plots to view how the density is altered during the excitation. The plots for the eight excitations in the target region are shown in figure 46. The green lobes are from the ground state and the purple lobes are from the excited state. With all the excitations in the target region the density difference plots look clearly charge transfer (with the electron density moving from one ring to the other), the ground state is on the ring with a single proton on and the excited state is on the ring with two protons on. This makes sense due to the electron density moving to the more positively charged ring. These plots also suggest that there may be a potential for proton transfer/conductance going in the opposite direction to the movement of the electron density. This will be investigated by looking at a potential proton transfer pathway through the model system.

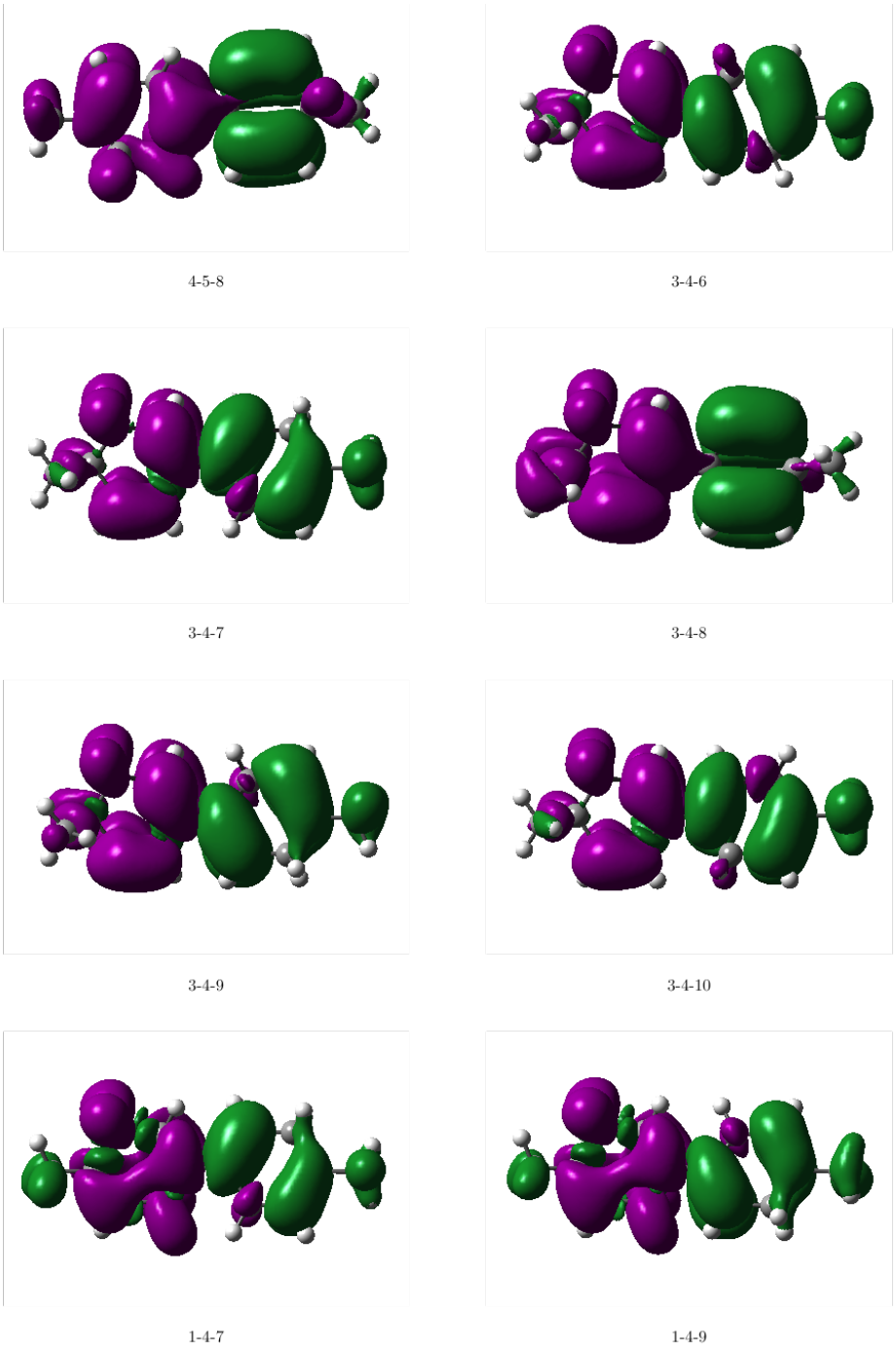


Figure 46: Density difference plots of excitations in target region

4.2 Pyrene-based Conjugated Microporous Polymers

Conjugated microporous polymers (CMPs) are a group of porous polymers. They are synthesised from building blocks in a similar vein to the closely related PAF discussed previously. Pyrene-based CMPs exhibit both microporosity and luminescent properties[110]. Pyrene-based CMPs are suited to applications such as pho-

tocatalysis, the important property for governing this use is the band-gap of the CMP. Cooper *et al.*[2] showed that is it possible to change the band-gap of pyrene-based CMPs via the inclusion of linker molecules. The building blocks of CMPs studied by Cooper[2] are shown in figure 47.

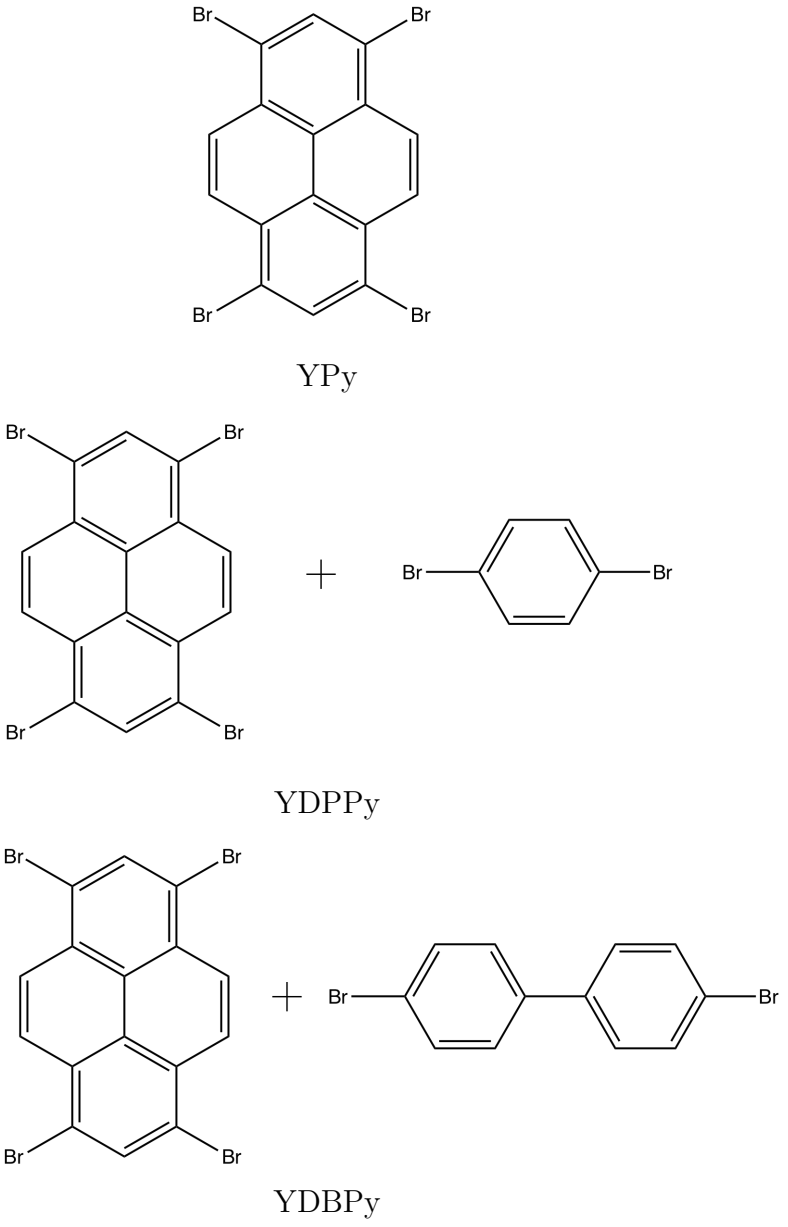


Figure 47: Building blocks of the CMPs studied by Cooper *et al.*[2]

YPy has the lowest band gap of these three CMPs (1.84 eV) and YDBPY has the largest (2.05 eV). The inclusion of these linker molecules has lead to a change in the band gap. One possible cause of this change is band gap is the inclusion of the linkers changing the smaller molecular structures present in the CMPs, such as smaller molecular rings.

These molecular rings where suggested as a possible cause for the band gap

differences[2]. The size and quantity of these molecular rings present in the CMP may lead to changes in the band gap of the CMPs. An example molecular ring is shown in figure 48.

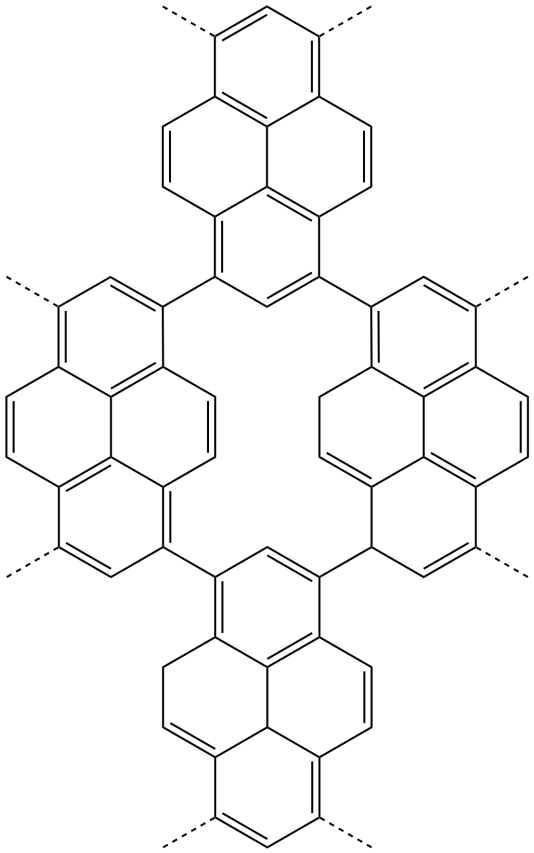


Figure 48: A representation of molecular rings formed by 4 pyrene building blocks.

4.2.1 Invesitgation of Structural Diversity of Pyrene-based CMPs

In order to understand the influence of these molecular rings and the influence of additional linker molecules along with their diversity/likelihood of appearance in extended CMPs, a statistical representation of the structural diversity was obtained. This was performed using AmBuild41 along with synthesis of the CMPs studied computationally by collaborators Patrick Heasman and Abbie Trewin. Clusters of the YPy were formed initially using an increasing number of pyrene building blocks (3-10).These building blocks (figure 49) have 4 possible binding sites at the bromine atoms and thus the blocks are linked at these sites. The clusters were formed by forming bonds when the building blocks had reached a critical angle and distance away from each other. The molecules that were at this critical point where bonded through the bromine groups and the simulation was continued. This simulation method was run 100 times for each number of building blocks. Further information about the procedure to generate these clusters can be

found in reference 111.

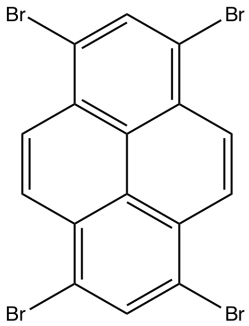


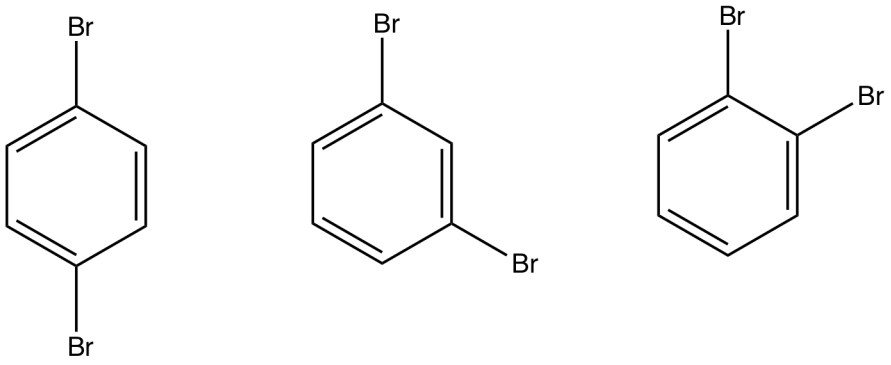
Figure 49: Pyrene building block for the diversity investigation

100 simulations of cluster formation were performed for each number of building blocks and thus give a representation of the structural diversity. The results of these simulations for YPy are shown in table 64. It is clear that the number of molecular rings formed increases as the number of building blocks increases.

Table 64: Structural Diversity of YPy clusters

	Building Blocks							
Structure	3	4	5	6	7	8	9	10
Linear	100	66	34	10	5	1	0	0
Branched	0	28	49	52	39	26	12	4
Double branched	0	0	1	7	15	23	26	19
Multibranched	0	0	0	0	0	2	7	13
Pure single ring	0	6	1	0	0	0	0	0
Branched single ring	0	0	15	29	37	39	40	44
Pure double ring	0	0	0	1	0	0	0	0
Branched double ring	0	0	0	1	3	7	13	17
Branched triple ring	0	0	0	0	1	2	2	3
Straight chain	100	94	84	69	59	52	45	36
Ring structure	0	6	16	31	41	48	55	64

The influence of linker molecules was investigated, the linker molecules are shown in figure 50 along with the names the resulting CMPs will be given. The linker molecules are all based on the linker used for YDPPy. S0 is the same linker molecule as the previously studied YDPPy[2]. The linker molecules are closely related with the position of the bromines moved around the benzene ring. The inclusion of these linker molecules will lead to more densely packed materials when these are included, with S0 being the least densely packed and S2 being the most densely packed.



S0 S1 S2

Figure 50: Linker molecules added to the cluster formation simulations

Cluster formation simulations, with 3-10 pyrene building blocks and twice the number of linkers were undertaken for each linker molecule, were undertaken in a similar fashion to those for YPy discussed above. The results for each set of simulations are shown in tables 65, 66 and 67 for S0, S1 and S2 respectively.

Table 65: Structural Diversity of S0 clusters

	Building Blocks							
Structure	3	4	5	6	7	8	9	10
Linear	12	0	0	0	0	0	0	0
Branched	47	32	9	3	1	0	0	0
Double branched	33	33	32	18	9	1	0	0
Multibranched	8	35	57	76	85	93	91	91
Branched single ring	0	0	2	3	5	6	9	9
Straight chain	100	100	98	97	95	94	91	91
Ring structure	0	0	2	3	5	6	9	9

Table 66: Structural Diversity of S1 clusters

	Building Blocks							
Structure	3	4	5	6	7	8	9	10
Linear	12	3	2	1	0	0	0	0
Branched	42	22	12	4	1	0	0	0
Double branched	20	25	11	8	2	0	0	0
Multibranched	8	25	41	47	51	45	34	25
Branched single ring	17	23	30	29	35	36	37	32
Branched double ring	1	1	2	9	8	15	24	30
Branched triple ring	0	1	2	2	3	3	4	10
Branched quadruple ring	0	0	0	0	0	1	0	2
Branched quintuple ring	0	0	0	0	0	0	1	1
Straight chain	82	75	66	60	54	45	34	25
Ring structure	18	25	34	40	46	55	66	75

Table 67: Structural Diversity of S1 clusters

	Building Blocks							
Structure	3	4	5	6	7	8	9	10
Linear	6	1	0	0	0	0	0	0
Branched	29	13	7	1	0	0	0	0
Double branched	15	18	4	3	0	0	0	0
Multibranched	7	17	29	27	16	8	7	4
Branched single ring	37	38	38	37	45	38	28	20
Branched double ring	6	13	19	22	20	27	34	34
Branched triple ring	0	0	3	10	18	21	20	20
Branched quadruple ring	0	0	0	0	1	6	7	13
Branched quintuple ring	0	0	0	0	0	0	3	6
Branched sextuple ring	0	0	0	0	0	0	1	3
Straight chain	57	49	40	31	16	8	7	4
Ring structure	43	51	60	69	84	92	93	96

In order to enable a simpler comparison between the structural diversity of these 4 structures, the straight chain and ring structure totals for each number of pyrene building blocks are shown in table 68.

Table 68: Structural diversity of all four structures studied via cluster formation

	YPy		S0		S1		S2	
Building Blocks	Straight	Ring	Straight	Ring	Straight	Ring	Straight	Ring
3	100	0	100	0	82	18	57	43
4	94	6	100	0	75	25	49	51
5	84	16	98	2	66	34	40	60
6	69	31	97	3	60	40	31	69
7	59	41	95	5	54	46	16	84
8	52	48	94	6	45	55	8	92
9	45	55	91	9	34	66	7	93
10	36	64	91	9	25	75	4	96

The influence of increasing the number of pyrene building blocks is consistent, a higher number of ring structures with a higher number of building blocks. The influence of the linker is dependent on the orientation of the two bromines. S0 has the least number of ring structures and it has the bromines in a para- orientation. S2 has the highest number of ring structures and it has the bromines in an ortho- orientation. The orientation of the bromines in the linker controls the structural diversity of the CMPs with linkers.

4.2.2 Substitution studies of Structures

In order to investigate the excited state properties and band gaps of the structures obtained from the cluster formation, DFT calculations were undertaken. All calculations undertaken were using CAM-B3LYP and the def2-SVP basis set, all excited state calculations used TDA-TDDFT. Due to the size of the structures obtained from the cluster formation, a preliminary investigation was undertaken to evaluate if computational cost can be reduced with substitution of the bromine atoms present in the clusters with less expensive atoms (e.g hydrogen, fluorine, chlorine).

This initial investigation used clusters of YPy with 3 pyrene building blocks. The bromine atoms were replaced by hydrogen, fluorine and chlorine and the first ten singlet and triplet excitations of each of these substituted structures were compared to those for the bromine containing YPy cluster. The singlet and triplet excitations for these substituted structures are shown in figures 51 and 52 respectively.

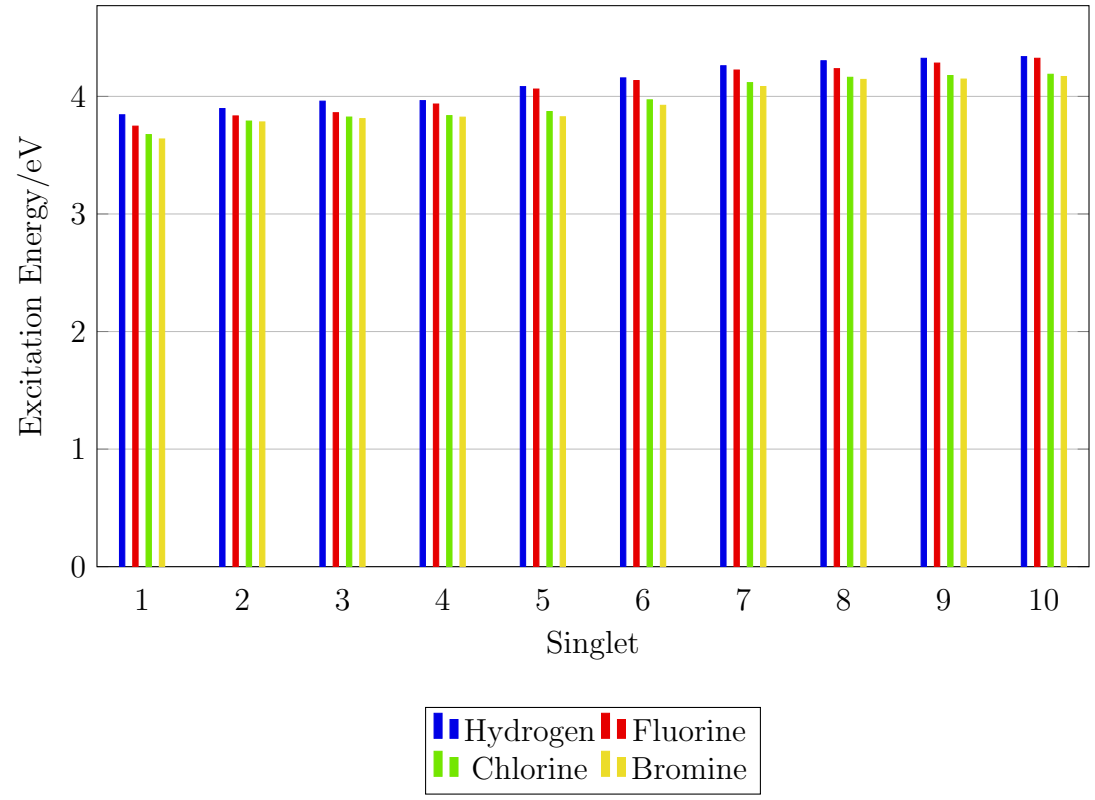


Figure 51: First ten singlet excitations of substituted 3 pyrene YPy clusters

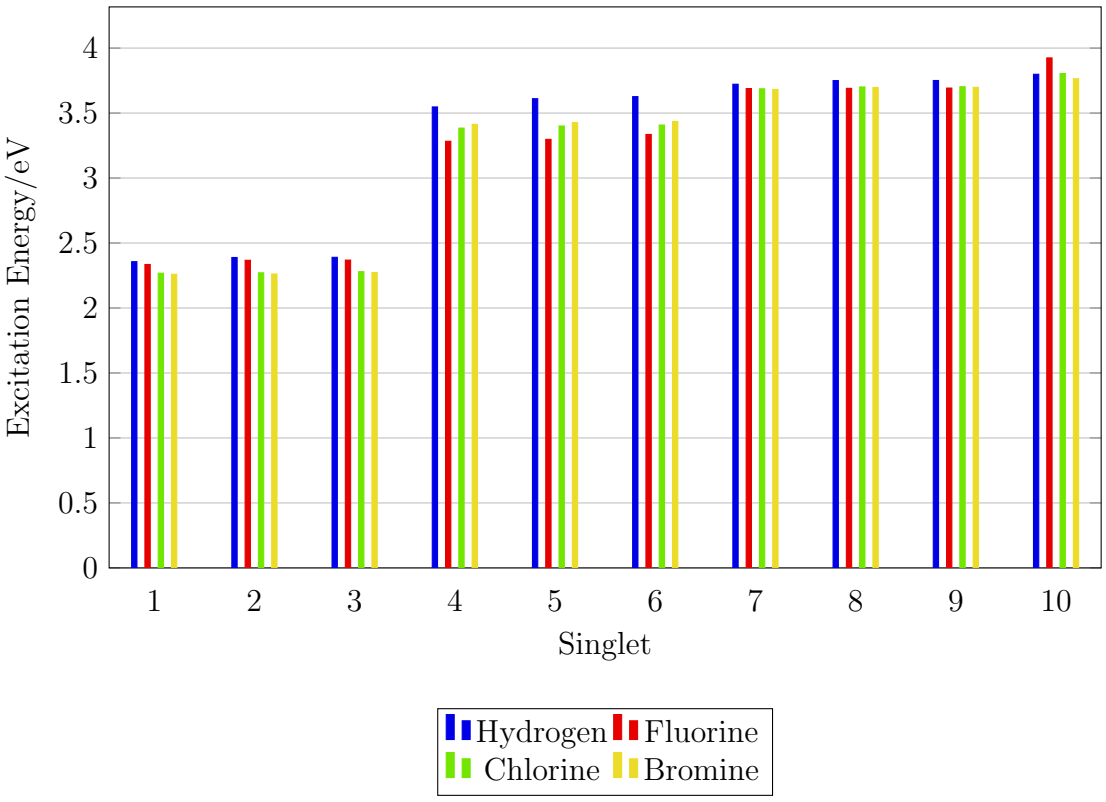


Figure 52: First ten triplet excitations of substituted 3 pyrene YPy clusters

It can be seen that the difference between the Br YPy excitations and the substituted YPy excitations remains fairly constant. The hydrogen substituted structures have higher excitation energies but these remain a fairly constant value above the bromine structure excitations. The influence of the substitution of bromine on the excitation energy across these first 10 excitations is consistent. Thus, the substitution of the bromine atoms with hydrogen atoms is a good approximation in order to reduce the computational cost of the calculations undertaken on the larger clusters of CMPs studied subsequently.

4.2.3 Excited State Study – 7 Pyrene units

An investigation of the excited states of clusters with 7 pyrene units of YPy and the three linker molecules were used to calculate the first 5 singlet and triplet states for each cluster. There is a variety of structures seen across each of the CMPs. The least diverse CMP is S0 and the most is S2. The structures are in two major groups: linear/branched and branched ring. The linear/branched structures are: linear (YPy, S0, S1, S2), branched (YPy, S2), double branched (YPy, S2) and multibranched (YPy, S0, S1, S2). The branched ring structures are: branched single ring (YPy, S0, S1, S2), branched double ring (YPy, S1, S2), branched triple ring (S1, S2) and branched quadruple ring (S1, S2).

4.2.3.1 Singlet Excitations

The results for the first 5 singlet states for YPy, S0, S1 and S2 are shown in tables 69, 70, 71 and 72.

Table 69: Excitation energies of the first 5 singlet states of YPy clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	3·64	3·79	3·82	3·83	3·83
Branched	3·50	3·64	3·68	3·70	3·77
Double branched	3·53	3·60	3·68	3·77	3·78
Branched single ring	3·29	3·43	3·63	3·67	3·69
Branched double ring	3·37	3·61	3·62	3·69	3·72

Table 70: Excitation energies of the first 5 singlet states of S0 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	3·64	3·79	3·82	3·83	3·83
Multibranched	3·32	3·47	3·52	3·56	3·63
Branched single ring	3·08	3·28	3·40	3·50	3·57

Table 71: Excitation energies of the first 5 singlet states of S1 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	3·38	3·55	3·64	3·71	3·72
Multibranched	3·08	3·26	3·31	3·44	3·48
Branched single ring	2·91	3·07	3·18	3·24	3·33
Branched double ring	2·51	2·76	2·83	2·92	3·00
Branched triple ring	2·68	2·75	2·94	2·99	3·07
Branched quadruple ring	2·68	2·72	2·82	2·94	3·10

Table 72: Excitation energies of the first 5 singlet states of S2 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	3·50	3·56	3·62	3·75	3·75
Branched	3·27	3·36	3·48	3·54	3·56
Double branched	3·31	3·32	3·41	3·61	3·66
Multibranched	3·00	3·13	3·38	3·51	3·58
Branched single ring	2·97	3·11	3·19	3·30	3·35
Branched double ring	2·96	3·02	3·33	3·47	3·50
Branched triple ring	2·73	2·96	3·08	3·19	3·23
Branched quadruple ring	3·18	3·24	3·35	3·40	3·52

A clear trend can be seen in the singlet excitations between the linear/branched structures and the ring structures. The linear/branched structures have higher excitation energies than the ring structures (difference roughly between 0.3 eV and 0.8 eV). This trend holds true across all CMPs studied here.

4.2.3.2 Triplet Excitations

The results for the first 5 triplet states for YPy, S0, S1 and S2 are shown in tables 73, 74, 75 and 76.

Table 73: Excitation energies of the first 5 triplet states of YPy clusters with 7 pyrene units, all values in eV.

	Triplet State				
Structure	1	2	3	4	5
Linear	2·26	2·27	2·28	3·42	3·43
Branched	2·26	2·26	2·26	2·27	2·27
Double branched	2·26	2·26	2·26	2·26	2·27
Branched single ring	2·19	2·21	2·26	2·26	2·26
Branched double ring	2·23	2·24	2·24	2·26	2·26

Table 74: Excitation energies of the first 5 triplet states of S0 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	2·25	2·26	2·27	2·40	2·56
Multibranched	2·16	2·21	2·22	2·23	2·23
Branched single ring	2·08	2·15	2·18	2·22	2·23

Table 75: Excitation energies of the first 5 triplet states of S1 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	2·18	2·19	2·21	3·29	3·35
Multibranched	2·07	2·12	2·14	2·17	2·18
Branched single ring	1·97	2·02	2·07	2·13	2·17
Branched double ring	1·39	1·56	1·79	2·01	2·08
Branched triple ring	1·54	1·55	1·89	1·91	1·96
Branched quadruple ring	1·58	1·62	2·06	2·11	2·19

Table 76: Excitation energies of the first 5 triplet states of S2 clusters with 7 pyrene units, all values in eV.

	Singlet State				
Structure	1	2	3	4	5
Linear	2·18	2·19	2·19	3·27	3·30
Branched	2·12	2·12	2·17	2·18	2·19
Double branched	2·12	2·15	2·17	2·22	2·25
Multibranched	2·00	2·04	2·15	2·19	2·20
Branched single ring	1·83	2·05	2·07	2·08	2·13
Branched double ring	1·80	1·97	2·1	2·17	2·19
Branched triple ring	1·63	1·94	2·00	2·08	2·15
Branched quadruple ring	2·02	2·11	2·15	2·16	2·20

The trend observed in the singlet excitations is still present in the triplet excitations for S0, S1 and S2 clusters but is not obvious in the YPy clusters with all the triplet excitations being similar across all the structures, with a small decrease for the ring structures (< 0.1 eV).

Density difference (DD) plots can again be used to enable a visualisation of the excitation processes in the CMP clusters, which enables clearer view of the excitation process. DD plots were generated for clusters of YPy and S2 can be seen in figure 53.

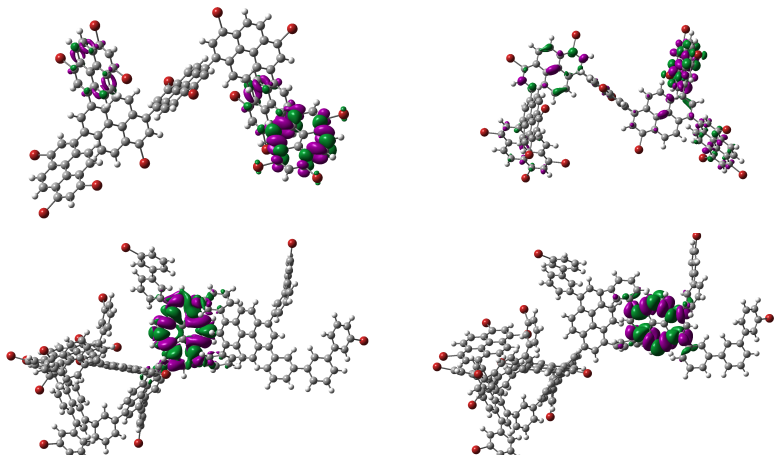


Figure 53: Density difference plots of YPy and S2 clusters

It can clearly be seen from these DD plots that the excitations are localised onto the pyrene units of the CMP clusters. It follows that the local environment of these pyrene units is important when attempting to control or change the excitations.

Therefore, functionalising the linkers included in these CMPs may enable a method to alter the excitations and band gaps of the CMP polymers which is important for photocatalysis.

4.2.4 Statistically Representative Set of Clusters

In order to enable a more realistic comparison to experiment, namely to UV-Vis spectra produced in experiment, a more representative set of clusters should be used. As the set used in the previous excited state study were not representative of the structural diversity of the present in the full set of clusters formed in the structural diversity study. Thus a set of clusters using 10 pyrene units was selected for the 4 CMPs, a total of 11 structures were included in each of these representative sets. Table 77 shows the number of each structure present in the representative set for each CMP.

Table 77: Statistically representative set of clusters for each CMP from the structural diversity study, 11 total for each CMP

Structure	YPy	S0	S1	S2
Branched	1	0	0	0
Double branched	1	0	0	0
Multibranched	2	9	2	1
Branched single ring	4	2	3	2
Branched double ring	3	0	3	3
Branched triple ring	0	0	1	2
Branched quadruple ring	0	0	1	1
Branched quintuple ring	0	0	1	1
Branched sextuple ring	0	0	0	1

Each of the clusters in the representative set had the first ten singlet states calculated, this is plotted against the oscillator strength of each excitation (gives a comparison to the intensity of the excitation). These are plotted against each other and thus generates a representative plot of these which is comparable to the experimental UV-Vis spectra for each CMP. Plots of exication energies against oscillator strengths for YPy, S0, S1 and S2 are shown in figures 54, 55, 56 and 57 respectively.

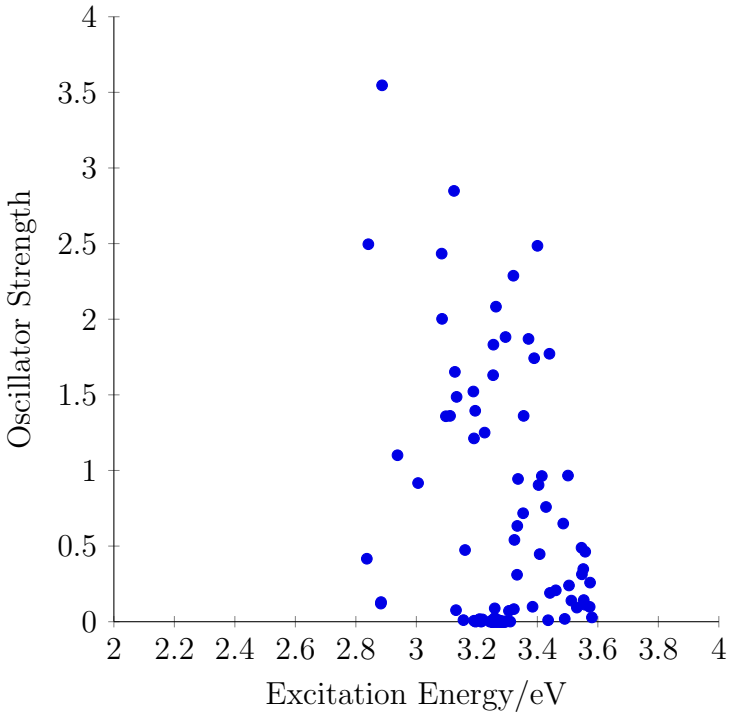


Figure 54: Plot of excitation energy against oscillator strength for the YPy representative set

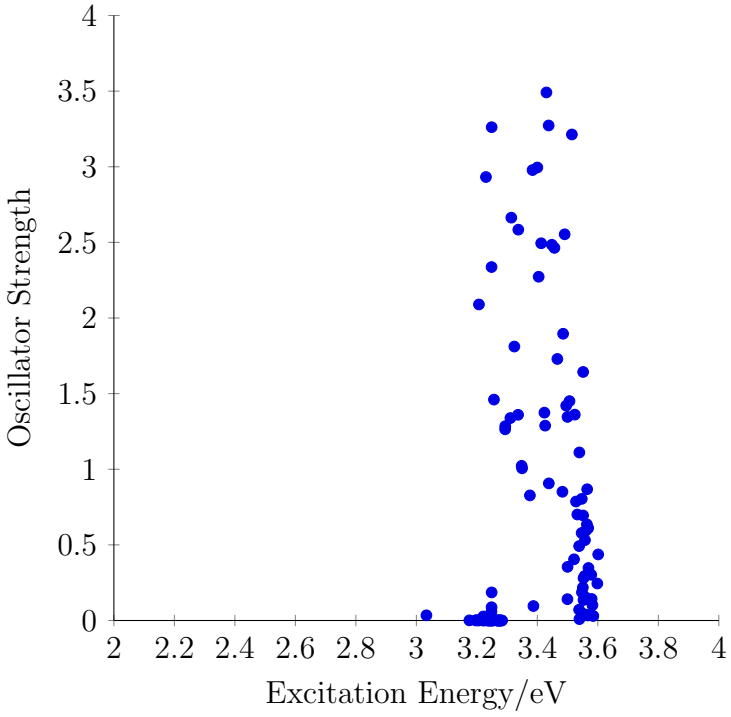


Figure 55: Plot of excitation energy against oscillator strength for the S0 representative set

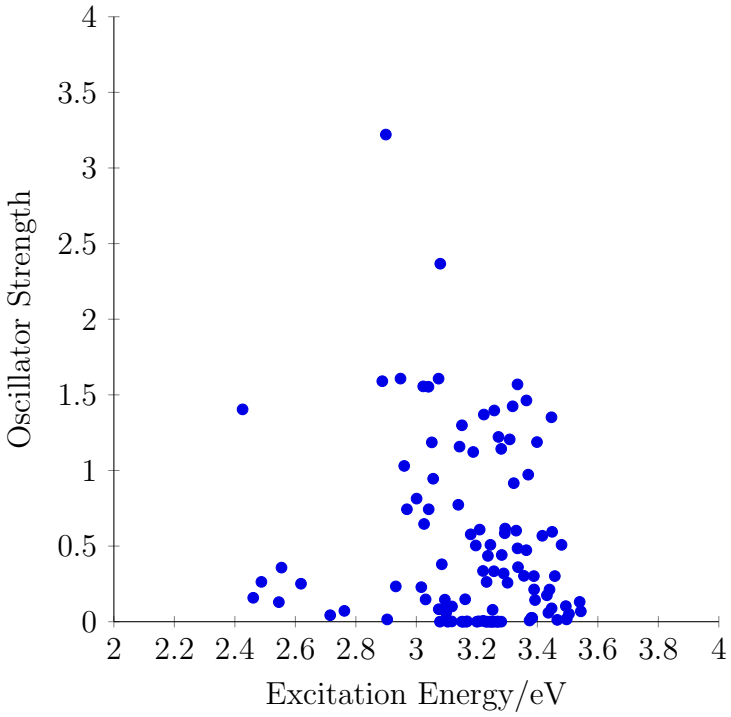


Figure 56: Plot of excitation energy against oscillator strength for the S1 representative set

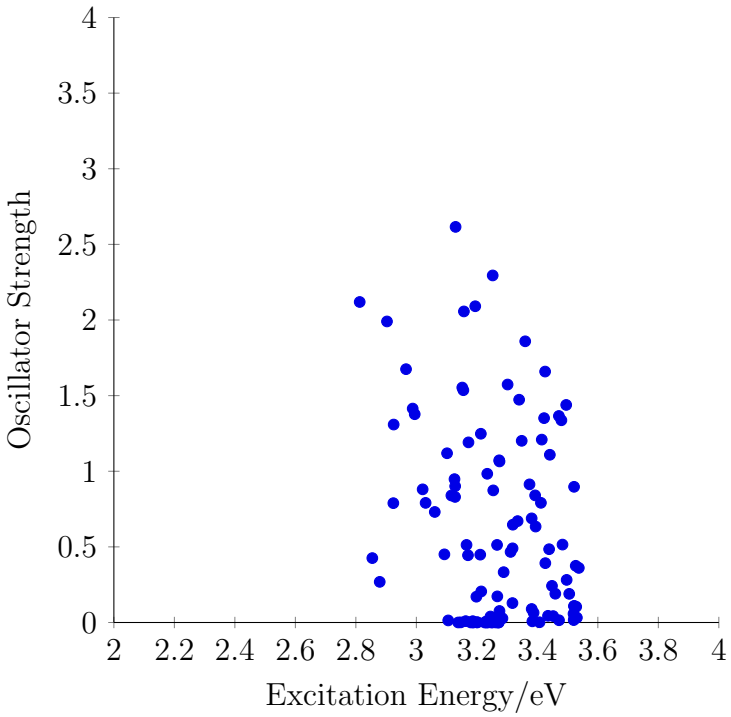


Figure 57: Plot of excitation energy against oscillator strength for the S2 representative set

These plots show that the more structurally diverse CMPs have the largest

spread of excitation energies and the least structurally diverse CMP (S0) has the lowest spread. This trend should be observed in the experimental spectra provided the model of the CMP is suitable. The experimental spectra all show peaks centred at 466nm with a small difference in the spread seen, S0 has the narrowest peak which is as expected from the lower structural diversity seen, and the other 3 systems (YPy, S1 and S2) show similar spreads due to their structural diversity.

4.3 Conclusions

The use of TDDFT methods enables an enhanced understanding of experimentally observed results when considering porous polymers. The two groups of polymers studied here are both amorphous thus require the use of model systems to represent their structures when performing TDDFT calculations. These model systems are carefully selected for the experimental result that is of interest.

The experimental result that required enhanced understanding with PAF-1 was a change in the UV-Vis spectra when placing the polymer in concentrated acid, a biphenyl model system was used to represent the polymer. The protonation of this model system was shown to be stable, with similar energy drops when adding up to 3 protons per model system and a reduced drop when adding a fourth. Therefore, there is a limit to the maximum protonation of the model system and thus polymer.

The protonation of the model system leads to a decrease in the absorption energies calculated when compared to the neutral model system. This decrease did not change the absorption to the degree seen in experiment until 3 protons were added to the model system. In order to get a clearer view of these excitations, density difference plots were generated and they showed a clear charge separation during the excitation with the orbital the excitation was from being on the ring with 2 protons attached and the orbital the excitation was to being on the ring with 1 proton attached. This charge separation and plots suggests that there is potential proton conduction/movement opposite to the movement of electron density in the excitation, more experimental and theoretical work is required understand this but the observation would not have been made without the enhanced understanding of the experimental result using TDDFT.

When considering the CMPs, the model systems were generated via cluster formation simulations performed by theoretical collaborators. The property of these polymers that required enhanced understanding was about the band gaps/excitations of these polymers. The structure types seen during cluster formation simulations had 2 major categories (branched and ring structures), each of these broad categories had differing excitation behaviours. The ring structures had

lower excitation energies than the branched structures across the 4 CMPs studied.

DD plots were again used to provide a clearer picture of the excitation process in these systems. These plots show that the excitations are localised on the pyrene rings of these model systems, thus the local environment of these rings is important for controlling these excitations. It may be possible to tailor the band gap/excitations by functionalising the linkers included in these CMPs, again further experimental/theoretical work is required to confirm this suggestion.

When considering a representative set of the diversity of structures generated by the cluster formation, an UV-Vis spectra can be generated, these show that the position of the peak changes very little between the 4 CMPs but the broadness of the peak increases with higher structural diversity and this is seen in the experimental spectra of the frameworks.

5 Conclusions

The extensive assessment of the 3 global and 3 range separated hybrids (B3LYP, PBE0, B97-2, CAM-B3LYP, LC-ωPBE and ωB97-XD) over a range of ground state and excited state properties has shown some interesting results. It highlights that each functional tested has properties where it performs accurately and those with less accurate performance. B3LYP performs well for ionisation energies, vibrational wavenumbers, electron affinities and local excitations but less well for isotropic polarisabilities, Rydberg excitations and charge transfer excitations. PBE0 performs well for isotropic polarisabilities and local excitations but less well for electron affinities and atomisation energies. B97-2 performs well for ionisation energies, atomisation energies, vibrational wavenumbers, isotropic polarisabilites and local excitations but less well for diatomic bond lengths, Rydberg excitations and charge transfer excitations. CAM-B3LYP performs well for diatomic bond lengths and all excitations but less well for isotropic polarisabilities. LC-ωPBE performs well for reaction barriers, local excitations and Rydberg excitations but less well for ionisation potentials, electron affinities, diatomic bond lengths, vibrational wavenumbers. ωB97-XD performs welll for most properties but less well ionisation potentials and vibrational wavenumbers.

The increased computational cost and complexity associated with using a range separated hybrid is important for some properties but less so for others. It is important to use range separated hybrids for the accurate calculation of excitations (other than local excitations, where all functionals tested performed well), atomisation energies and reaction barriers. The performance of LC-ωPBE is poorer than the other range separated hybrids in most excitations excluding Rydberg excitations.

However, the performance of LC-ωPBE may be improved by tuning the parameters of the range separation. Initially the parameters used where the same as those of CAM-B3LYP ($a = 0.19$,$a+b = 0.65$ and $\omega = 0.33$ a_0^{-1}), generating CAM-ωPBE. This new functional has poor performance across most properties with exceptions of (list properties with good performance). The parameters where optimised using a set of atomisation energies whilst applying the constraint $a + b = 0.65$ (keeping the $a + b$ value the same as CAM-B3LYP). This lead to an optimised verison of CAM-ωPBE (CAM-ωPBEa). The performance of this optimised functional was improved over CAM-ωPBE. However, the performance was still poor when compated to CAM-B3LYP and LC-ωPBE.

In order to understand the underlying cause of the poor performance of CAM-ωPBE and CAM-ωPBEa. A study of the quantity of HFx required for optimally performing global hybrid of a series of underlying functionals (LDA, PBE, BLYP

and TPSS) by assessing the performance of a global hybrid with increasing quantities of HFx over a set of atomisation energies. This lead to some interesting results, each underlying functional had a differing quantity of HFx in its optimal global, with LDA requiring the highest quantity and TPSS requiring the least.

It was also seen that the optimal quantity of HFx differed for some global hybrids (especially for PBE and LDA) when considering different molecule types (cyclic and non cyclic). This shows that some molecule types have differing behaviour when considering the quantity of HFx required (cyclic requires more HFx for PBE based). This shows that what is meant by a diverse set may require redefinition as it could be possible to lose details when considering too large a set or under or over representing certain categories of molecules (such as cyclic and non cyclic molecules). This observation could lead to utilising more of the data that is present in current benchmark sets to enable more accurate calculation of properties with functionals tuned to the categories of molecules. Current benchmark data could be split into different subsets and this could provide information that is currently being missed due to the drive to tuning for average performance.

In order to properly assess the optimal global hybrid based on PBE and BLYP, a similar assessment was performed over the full G-1 and G-2 sets of atomisation energies. The optimal global hybrid for each underlying functional had a differing quantity of HFx (values of HFx in each optimal). This difference sheds some light on the issues seen with the initial CAM-ωPBE and CAM-ωPBEa functionals, due to insufficient HFx included when $a + b = 0.65$ was enforced.‘

With the observation about the differing quantity of HFx required for an optimally performing global hybrid of PBE and BLYP, another optimisation of the range separation parameters for CAM-ωPBE with the condition $a + b < 1$ was undertaken. The optimisation used a series of atomisation energies, including pyridine due to the differing behaviour of cyclic systems observed when considering global hybrids of PBE. The optimised parameters were $a = 0.05$, $a + b = 0.80$ and $\omega = 0.40\mathrm{a}_0^{-1}$, referred to as CAM-PBE. The performance of CAM-PBE was much improved over CAM-ωPBEa; especially for excited state properties where CAM-PBE has a comparable performance to CAM-B3LYP, which suggests that an attenuated range separated hybrid of PBE may have merits for the calculation of excited state properties. However, there is poor performance when calculating ionisation energies and electron affinities.

Thus a set of parameters were optimised on electron affinities ($a = 0.10$, $a + b = 0.80$ and $\omega = 0.15\ \mathrm{a}_0^{-1}$), referred to as CAM-PBEea. However, whilst these parameters improved the performance ofelectron affinities and ionisation potentials, they had a detrimental effect on the performance of other properties where CAM-PBE had good performance. Therefore whilst CAM-PBE has good

performance generally it should be noted aboutthe poor performance of electron affinities and ionisation potentials.

The benchmark set of excited state structures and emission energies developed here is attempting to fill noticeable gap in literature. As nothing similar to benchmark sets seen in literature for ground state properties (those included in the GMTNK55 database) exists for these properties. The need for this benchmark set is due to it being difficult to experimentally obtain data (short lived etc), with the exception of diatomics.

Diatomics are used as a preliminary study to highlight issues that may be faced when looking at more complicated systems. However, they are of limited use as diatomics can be brute forced (due to one possible coordinate) to get the correct behaviour of each state. The diatomics studied showed that it is important to use basis sets with diffuse functions when considering higher energy excitations (>9 eV). Another issue it highlighted was that there can be a potential issue when states cross during optimisation which will require careful monitoring to ensure the correct state is followed during the excited state geometry optimisation when developing the benchmark set and when using the benchmark set to assess approximate E_{xc}s.

Selection of molecules based on existing benchmark sets for absorptions, this includes a variety of small organic molecules, including 2 groups of closely related molecules (polyacetylene and acene groups).

The method used (RI-CC2) has been validated against higher level theory or experimentally derived results for some molecules and states from literature. RICC2 shows good performance for these states along with existing use of RI-CC2 as reference results for a molecule in the set. Once the method was validated, excited state geometries and emission values were calculated for each state in the benchmark set, combined to produce the benchmark set. The benchmark set developed here can and should be expanded upon. This could happen via higher level theory being used to calculate the emission energies at the structures obtained here or by expanding the set to include a greater diversity of molecules (include molecules which contain heavier atoms such as sulphur or phosphorous).

An assessment of the performance of a series of approximate E_{xc}s for emission energies/excited state geometries was undertaken. Singlet emissions are much more dependent of the quantity of HFx included in the functional that the triplet emissions. The trends observed are preserved between at the benchmark structures and at the structures obtained via excited state geometry optimisations using the related approximate E_{xc}. However there seems to be no approximate E_{xc} of those tested which has a great performance for the emission energies from the benchmark set either at the structures from the benchmark set or from structures

optimised using the approximate E_{xc}. One suggestion to take from this is that when approximate E_{xc}s are developed using ground state properties they are not accurate for excited state properties. This idea extends further as range separated hybrids perform better for absorptions than global hybrids. However, this is not the case for emission energies and thus excited state geometries. This means that the methods of improving the accuracy of excited state energies in the ground state do not necessarily improve the accuracy of excited state properties away from the ground state structure. Therefore a fresh view of developing approximate E_{xc}s for the calculation of excited state properties is needed. This could potentially come in different ways, one obvious suggestion is to use the newly generated benchmark set of excited state geomteries and emission energies when developing approximate E_{xc} as an additional exact data for approximate E_{xc} to reproduce.

The use of TDDFT methods enables an enhanced understanding of experimentally observed results when considering porous polymers. The two groups of polymers studied here are both amorphous thus require the use of model systems to represent their structures when performing TDDFT calculations. These model systems are carefully selected for the experimental result that is of interest.

The experimental result that required enhanced understanding with PAF-1 was a change in the UV-Vis spectra when placing the polymer in concentrated acid, a biphenyl model system was used to represent the polymer. The protonation of this model system was shown to be stable, with similar energy drops when adding up to 3 protons per model system and a reduced drop when adding a fourth. Therefore, there is a limit to the maximum protonation of the model system and thus polymer.

The protonation of the model system leads to a decrease in the absorption energies calculated when compared to the neutral model system. This decrease did not change the absorption to the degree seen in experiment until 3 protons were added to the model system. In order to get a clearer view of these excitations, density difference plots were generated and they showed a clear charge separation during the excitation with the orbital the excitation was from being on the ring with 2 protons attached and the orbital the excitation was to being on the ring with 1 proton attached. This charge separation and plots suggests that there is potential proton conduction/movement opposite to the movement of electron density in the excitation, more experimental and theoretical work is required understand this but the observation would not have been made without the enhanced understanding of the experimental result using TDDFT.

When considering the CMPs, the model systems were generated via cluster formation simulations performed by theoretical collaborators. The property of these polymers that required enhanced understanding was about the band

gaps/excitations of these polymers. The structure types seen during cluster formation simulations had 2 major categories (branched and ring structures), each of these broad categories had differing excitation behaviours. The ring structures had lower excitation energies than the branched structures across the 4 CMPs studied.

DD plots were again used to provide a clearer picture of the excitation process in these systems. These plots show that the excitations are localised on the pyrene rings of these model systems, thus the local environment of these rings is important for controlling these excitations. It may be possible to tailor the band gap/excitations by functionalising the linkers included in these CMPs, again further experimental/theoretical work is required to confirm this suggestion.

When considering a representative set of the diversity of structures generated by the cluster formation, an UV-Vis spectra can be generated, these show that the position of the peak changes very little between the 4 CMPs but the broadness of the peak increases with higher structural diversity and this is seen in the experimental spectra of the frameworks.

References

1. T. W. Keal and D. J. Tozer, *J. Chem. Phys.*, 2005, **123**, 121103.
2. J.-X. Jiang, A. Trewin, D. J. Adams and A. I. Cooper, *Chemical Science*, 2011, **2**, 1777.
3. P. W. Atkins, *Molecular Quantum Mechanics*, Oxford University Press, 1983.
4. F. Jensen, *Introduction to Computational Chemistry*, John Wiley and Sons Ltd., 2nd edn., 2007.
5. J. C. Slater and H. C. Verma, *Phys. Rev.*, 1929, **34**, 1293–1295.
6. P. A. M. Dirac, *Proc. R. Soc. A*, 1926, **112**, 661–677.
7. C. C. J. Roothaan, *Rev. Mod. Phys.*, 1951, **23**, 69–89.
8. J. Cizek, *J. Chem. Phys.*, 1966, **45**, 4256.
9. R. Bartlett, *Annu. Rev. Phys. Chem.*, 1981, **32**, 359–401.
10. J. Stubber and J. Paldus, *Journal of Molecular Structure: THEOCHEM*, 2002, **591**, 219–230.
11. G. D. Purvis and R. J. Bartlett, *J. Chem. Phys.*, 1982, **76**, 1910–1918.

12. J. A. Pople, M. Head-Gordon and K. Raghavachari, *J. Chem. Phys.*, 1987, **87**, 5968–5975.

13. C. Moller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

14. G. A. Arteca, F. M. Fernández and E. A. Castro, in *Rayleigh-Schrodinger Perturbation Theory (RSPT)*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1990, pp. 45–71.

15. J. C. Slater, *Phys. Rev.*, 1930, **36**, 57–64.

16. S. F. Boys, *Proc. R. Soc. A*, 1950, **200**, 542–554.

17. W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1969, **51**, 2657–2664.

18. R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.

19. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

20. P. Hohenburg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

21. W. Kohn and L. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

22. M. Levy, *Proc. Natl. Acad. Sci. U.S.A.*, 1979, **76**, 6062.

23. E. Lieb, *Int. Jour. Quant. Chem.*, 1983, **24**, 243–277.

24. J. P. Perdew and K. Schmidt, *AIP Conf. Proc.*, 2011, **577**, 1–20.

25. P. A. M. Dirac, *Proc. Cambridge Philos. Soc.*, 1930, **26**, 376.

26. M. Gell-Mann and K. A. Brueckner, *Phys. Rev.*, 1957, **106**, 364.

27. P.-F. Loos and P. M. W. Gill, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2016, **6**, 410–429.

28. D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566–568.

29. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

30. J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1992, **45**, 13244–13249.

31. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

32. A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

33. C. Lee, W. Yang and R. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

34. J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.

35. J. Tao, J. P. Perdew, V. N. Straoverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.

36. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

37. J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

38. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

39. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Chem. Phys.*, 1994, **98**, 11623–11627.

40. C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6169.

41. Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2007, **120**, 215–241.

42. J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.

43. J. Heyd and G. E. Scuseria, *J. Chem. Phys.*, 2004, **121**, 1187–1192.

44. H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540–3544.

45. O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 234109.

46. J. D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

47. M. J. G. Peach, P. Benfield, T. Helgaker and D. J. Tozer, *J. Chem. Phys.*, 2008, **128**, 044118.

48. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

49. A. D. Laurent, C. Adamo and D. Jacquemin, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14334–14356.

50. A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009–4037.

51. M. Casida and M. Huix-Rotllant, *Annu. Rev. Phys. Chem.*, 2012, **63**, 287–323.

52. C. Adamo and D. Jacquemin, *Chem. Soc. Rev.*, 2013, **42**, 845–856.

53. E. Runge and E. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.

54. R. van Leeuwen, *Phys. Rev. Lett.*, 1999, **82**, 3863–3866.

55. M. Marques, C. Ullrich, F. Nogueira, A. Rubio, K. Burke and E. Gross, *Time-Dependent Density Functional Theory*, Springer-Verlag, 2006.

56. A. Görling, *Phys. Rev. A*, 1998, **57**, 3433–3436.

57. E. Gross, J. Dobson and M. Petersilka, *Density functional theory of time-dependent phenomena, Topics in Current Chemsity*, Springer, 1996.

58. M. Casida, in *In Recent Advances in Density-Functional Methods*, 1995, vol. 3.

59. G. Onida, L. Reining and A. Rubio, *Reviews of Modern Physics*, 2002, **74**, 601–659.

60. A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, 1971.

61. S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.*, 1999, **314**, 291 – 299.

62. A. Dreuw, G. R. Fleming and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2003, **5**, 3247.

63. B. D. Dunietz, A. Dreuw and M. Head-Gordon, *J. Phys. Chem. B*, 2003, **107**, 5623.

64. D. J. Tozer, R. D. Amos, N. C. Handy, B. J. Roos and L. Serrano-Andres, *Mol. Phys.*, 1999, **97**, 859.

65. A. Dreuw, J. L. Weisman and M. Head-Gordon, *J. Chem. Phys.*, 2003, **119**, 2943.

66. A. Dreuw and M. Head-Gordon, *J. Am. Chem. Soc.*, 2004, **126**, 4007.

67. A. L. Sobolewski and W. Domcke, *Chem. Phys.*, 2003, **294**, 73.

68. Z. L. Cai, K. Sendt and J. R. Reimers, *J. Chem. Phys.*, 2002, **117**, 5543.

69. S. Grimme and M. Parac, *Chem. Phys. Chem.*, 2003, **3**, 292.

70. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, 2009.

71. *TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007,TURBOMOLE GmbH, since 2007; available from* `http://www.turbomole.com`.

72. J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M. L. Klein and J. P. Perdew, 2015.

73. Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.

74. Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2006, **110**, 13126–13130.

75. Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2006, **110**, 5121–5129.

76. T. Koopmans, *Physica*, 1934, **1**, 104–113.

77. R. Baer and D. Neuhauser, *Phys. Rev. Lett.*, 2005, **94**, 043002.

78. E. Livshits and R. Baer, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2932.

79. L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **7**, 291–309.

80. L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670.

81. L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.

82. M. J. G. Peach, T. Helgaker, P. Salek, T. W. Keal, O. B. Lutnaes, D. J. Tozer and N. C. Handy, *Phys. Chem. Chem. Phys.*, 2006, **8**, 558–562.

83. C. Clavaguéra-Sarrio, N. Ismail, C. J. Marsden, D. Bégué and C. Pouchan, *Chemical Physics*, 2004, **302**, 1–11.

84. A. J. Sadlej, *Collect. Czech. Chem. Commun.*, 1988, **53**, 1995.

85. M. J. G. Peach, M. J. Williamson and D. J. Tozer, *J. Chem. Theory Comput.*, 2011, **7**, 3578–3585.

86. M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer and W. Thiel, *J. Chem. Phys.*, 2008, **128**, 134110.

87. M. R. Silva-Junior, S. P. A. Sauer, M. Schreiber and W. Thiel, *Mol. Phys.*, 2010, **108**, 453–465.

88. M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer and W. Thiel, *J. Chem. Phys.*, 2010, **133**, 174318.

89. M. J. G. Peach and D. J. Tozer, *J. Phys. Chem. A*, 2012, **116**, 9783–9789.

90. M. J. G. Peach, N. Warner and D. J. Tozer, *Mol. Phys.*, 2013, **111**, 1271–1274.

91. M. J. G. Peach, A. Cohen and D. J. Tozer, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4543.

92. L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221–7230.

93. L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 1997, **106**, 1063–1079.

94. C. Adamo, M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, 2000, **112**, 2643–2649.

95. J. C. Slater, *The Self-Consistent Field for Molecular and Solids, Quantum Theory of Molecular and Solids*, McGraw-Hill, 1974, vol. 4.

96. V. N. Straoverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129–12137.

97. F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2002, **117**, 7433–7447.

98. M. W. D. Hanson-Heine, M. W. George and N. A. Besley, *J. Chem. Phys.*, 2013, **138**, 064101.

99. *NIST webbook http://webbook.nist.gov/chemistry/.*

100. A. M. Teale and D. J. Tozer, *J. Chem. Phys.*, 2005, **122**, 034101.

101. A. Köhn and C. Hättig, *J. Am. Chem. Soc.*, 2004, **126**, 7399–7410.

102. I. Fdez. Galván, M. E. Martín, A. Muñoz-Losa, M. L. Sánchez and M. A. Aguilar, *J. Chem. Theory Comput.*, 2011, **7**, 1850–1857.

103. M. Boggio-Pasqua, M. J. Bearpark, M. Klene and M. A. Robb, *J. Chem. Phys.*, 2004, **120**, 7849.

104. W. H. Adams, *Phys. Rev.*, 1962, **127**, 1650–1658.

105. D. Wu, F. Xu, B. Sun, R. Fu, H. He and K. Matyjaszewski, *Chemical Reviews*, 2012, **112**, 3959–4015.

106. T. Ben and S. Qiu, *CrystEngComm*, 2013, **15**, 17–26.

107. T. Ben, H. Ren, S. Ma, D. Cao, J. Lan, X. Jing, W. Wang, J. Xu, F. Deng, J. Simmons, S. Qiu and G. Zhu, *Angew. Chem. Int. Ed.*, 2009, **48**, 9457–9460.

108. W. J. Hehre and J. A. Pople, *J. Am. Chem. Soc.*, 1972, **94**, 6901–6904.

109. J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, 2005, **105**, 2999–3094.

110. A. I. Cooper, *Advanced Materials*, 2009, **21**, 1291–1295.

111. J. Thomas, C. Mollart, L. Turner, P. Heasman, P. Fayon and A. Trewin, *J. Phys. Chem. B*, 2020, **124**, 7318–7326.