Determining the dominant degradation mechanisms in Nitrocellulose [CORRECTIONS]

Amy J. Lai

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Chemistry

University College London

December 18, 2019

Declaration

I, Amy Lai, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abbreviations

%N	percentage nitrogen by mass
2-NDPA	2-Nitrodiphenylamine
AIMD	*ab initio* molecular dynamics
AO	atomic orbital
a.u.	atomic units
B3LYP	Becke, 3-parameter, Lee-Yang-Parr hybrid functional
BCP	bonding critical point
BSSE	basis set superposition error
CH_3/CH_3	NC repeat unit with two $-OCH_3$ capping groups
CH_3/OH	NC repeat unit with $-OCH_3$ capping group on ring 1 and $-OH$ group on ring 2
OH/CH_3	NC repeat unit with $-OH$ capping group on ring 1 and $-OCH_3$ group on ring 2
CCP	cage critical point
CP	critical point
DFT	density functional theory
DFT-D	density functional theory with dispersion correction
DSC	differential scanning calorimetry
DOS	degree of substitution

DPA diphenylamine

EN ethyl nitrate

ESP electrostatic potential

FF force field

G09 Gaussian 09 revision E.01

GGA generalised gradient approximation

GM genetically modified

GTO Gaussian type orbitals

GView Gauss View 5.0.8

HF Hartree-Fock

HMF hydroxymethylfurfural

HOMO highest occupied molecular orbital

IR infra-red spectroscopy

KS-DFT Kohn-Sham DFT

LDA local density approximation

MD molecular dynamics

MEP minimum energy path

MM molecular mechanics

MMFF94 Merck molecular force field 94

MO molecular orbitals

MP2 Møller–Plesset perturbation theory with second order correction

MW molecular weight

NBO natural bond orbital

NC nitrocellulose

NCP nuclear critical point

NG nitroglycerine

NMR nuclear magnetic resonance spectroscopy

PCM polarisable continuum model

PES potential energy surface

PETN pentaerythritol tetranitrate

PETRIN pentaerythritol trinitrate

QM quantum mechanics

QTAIM quantum theory of atoms in molecules

RCP ring critical point

RESP restrained electrostatic potential atomic partial charges

RHF restricted HF

RMS root mean square

ROHF restricted-open HF

UHF unrestricted HF

SB59 1,4-bis(ethylamino)-9,10-anthraquinone dye

SCF self-consistent field

SCRF self-consistent reaction field

SEM scanning electron microscopy

SMD solvation model based on density

S_N2 bi-molecular nucleophilic substitution reaction

STO Slater type orbitals

TG	thermogravitmetric analysis
TS	transition state
UFF	universal force field
UV	ultraviolet
UV-vis	ultraviolet–visible spectroscopy
vdW	van der Waals
ωB97X-D	ωB97X-D long-range corrected hybrid functional
ZPE	zero-point energy

Chapter 1

Theory and Implementation

1.1 Electronic structure methods

Electronic structure methods apply the principles of quantum mechanics to the evaluation of electron position and movement, thereby allowing chemists to derive the properties and interactions of molecules. Despite the long history of research and use of nitrocellulose (NC) in industry, experimental analysis has failed to distinguish the fine mechanistic details of its decomposition. This is partly owed to the variation arising from biodiverse NC source materials, combined with the complexity due to the interplay of many different and simultaneous degradation interactions. Electronic structure methods provide a means to untangle the individual facets of decomposition.

At the most fundamental level, the wave function (Ψ) holds the description of a quantum system. In a non-relativistic system, the probability of a particle possessing a given momentum, or residing in a particular location, is given by the probability density. This can be obtained by multiplication of Ψ with its complex conjugate, $|\Psi^2|$. Integration of $|\Psi^2|$ over a region of space returns the probability that a system will be found within, called the Born interpretation. Values of Ψ are chosen to be orthonormal; integrating $|\Psi^2|$ over all space gives the probability of 1:

$$\langle \Psi_i | \Psi_j \rangle = \delta_{ij} \tag{1.1}$$

where all states are represented by i and j, and:

$$\delta_{ij} = 0 \text{ for } i \neq j$$
$$\delta_{ij} = 1 \text{ for } i = j \text{ the integral is one.}$$

Operators acting on Ψ yield the observable properties of the system. The operator returning the energy of the system is called the the Hamiltonian operator (**H**). Erwin Schrödinger

proposed his equation in 1926, describing a quantum system using its wave function [1]. Schrödinger's time-independent equation is:

$$\mathbf{H}\Psi = E\Psi \tag{1.2}$$

and the energy of the system is given by the expectation value of the Hamiltonian operator:

$$E = \langle \Psi | \mathbf{H} | \Psi \rangle \tag{1.3}$$

where the Hamiltonian operator $\mathbf{H}$ is an eigenvalue of the wave function Ψ, and E is a scalar denoting the energy of the system. A given system may have many acceptable values for Ψ, each with an associated value for E.

The general form of the Hamiltonian is given by:

$$\mathbf{H} = -\sum \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \tag{1.4}$$

where all electrons are represented by i and j, and all nuclei by k and l [2]. $\hbar$ is the reduced Planck's constant ($\hbar = \frac{h}{2\pi} = 1.055 \times 10^{-34}$ Js), m_e is the mass of an electron, m_k is the mass of the nucleus k, e is the charge of an electron, Z_k is the atomic number of k and r_{ik} is the distance between particles i and k. When using atomic units (a.u.), the value of e, m_e and $\hbar$ are reduced to 1. ∇^2 refers to the Laplacian operator, which describes the divergence of the gradient of a field. In Cartesian space, this is defined as the sum of the second derivatives of the gradient with respect to each of the three dimensions (x,y,z):

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{1.5}$$

The first and second terms of equation (1.4) correspond to the kinetic energy of the electrons and the nuclei, respectively. Electron-nuclear attraction is described by the third term; the fourth term describes inter-electronic repulsion and the final term the inter-nuclear repulsion. The final three potential energy terms are identical to their expression in classical mechanics. The kinetic energy terms can be expressed as the eigenvalues of the kinetic energy operator ($\mathbf{T}$):

$$\mathbf{T} = -\frac{\hbar^2}{2m}\nabla^2 \tag{1.6}$$

The total, non-relativistic Hamiltonian can therefore be written in terms of the kinetic energy and potential energy operators:

$$\mathbf{H} = \mathbf{T}_e + \mathbf{T}_N + \mathbf{V}_{e-N} + \mathbf{V}_{e-e} + \mathbf{V}_{N-N} \tag{1.7}$$

where the terms are as they were in equation 1.4. $\mathbf{T}_e$ corresponds to the kinetic energy of the electrons, $\mathbf{T}_N$ the kinetic energy of the nuclei, $\mathbf{V}_{e-N}$ the coulombic interaction between electron and nuclei, $\mathbf{V}_{e-e}$ the electron-electron repulsion and $\mathbf{V}_{N-N}$ the nuclear-nuclear interaction.

1.1.1 Born-Oppenheimer approximation

In a real system, the motion of elections and nuclei are coupled. Electron density flows dynamically in response to the change in nuclear position and repulsion from other electrons. The correlated motion of particles is described by the pairwise attractive and repulsive terms of the Schrödinger equation. However, this interdependency makes defining a wave function difficult. Relative to electronic motion, nuclei move far more slowly, owing to their much greater mass (the mass of a proton is around 1836 times larger than that of the electron). Nuclear positions therefore appear essentially stationary when compared to that of the electrons. Exploiting this property, the Born-Oppenheimer approximation fixes the nuclear positions. In this way, the motion of electrons and nuclei can be decoupled, and the electronic properties of the system may be calculated for the given nuclear coordinates. Dependency on the nuclear kinetic energy term ($\mathbf{T}_N$) is removed, as the nuclei are frozen. The nuclear-nuclear repulsive term ($\mathbf{V}_{N-N}$) becomes a constant for the specified geometry. Equation 1.4 is reduced to its electronic components and nuclear constants which in atomic units can be written as:

$$\mathbf{H} = \mathbf{T}_e + \mathbf{V}_{e-N} + \mathbf{V}_{e-e} + \mathbf{V}_{N-N} \tag{1.8}$$

and the electronic terms can be collected into one term, to simplify notation:

$$\mathbf{H} = \mathbf{H}_{el} + \mathbf{V}_{N-N} \tag{1.9}$$

The Schrödinger's equation can now be written in terms of only the electronic coordinates:

$$(\mathbf{H}_{el} + \mathbf{V}_{N-N})\Psi_{el}(\mathbf{q}_i;\mathbf{q}_k) = E_{el}(\mathbf{q}_i;\mathbf{q}_k) \tag{1.10}$$

where the electronic coordinates are given by $\mathbf{q}_i$, the stationary nuclear positions by $\mathbf{q}_k$ and E_{el} is the electronic energy of the system. The values of $\mathbf{q}_i$ are independent variables, whereas the values of $\mathbf{q}_k$ are parameters.

Given the example of a diatomic molecule, a potential energy curve can be obtained by calculating the value of E_{el} at different inter-nuclear distances. A series of these calculations generates a potential energy profile, allowing identification of an equilibrium bond length at the minimum of the curve. Calculation of E_{el} for all possible nuclear coordinates in a system of three or more atoms facilitates the construction of a hypersurface on which the potential energy is defined by the nuclear geometry, called a potential energy surface (PES). Exploration of the PES allows for discovery of global and local minimum energy structures, intermediate products and transition states on a reaction coordinate, through scrutiny of the energy at a particular set of nuclear coordinates.

Molecular structure theories adopt the Born-Oppenheimer approximation for its effective simplification of the coupled nuclear-electronic motion problem, in addition to its accuracy; this assumption works well for ground state molecules and only introduces very small errors. The model breaks down in the situation where there are multiple PES close in energy to one another, or even intersecting. In these cases the coupled equations must be considered. However for the work within this study, the Born-Oppenheimer approximation is successfully applied for all calculations involving electronic structure determination.

1.1.2 Slater determinants

In a system of multiple electrons, each electron is indistinguishable. If the positions of two electrons are swapped, the distribution of electron density in the system remains the same. The Pauli exclusion principle states that no two identical fermions, such as electrons, may simultaneously occupy the same quantum state within the same system. When considering an atom with two or more electrons, this means that none may have the same set of quantum numbers. As a result, for two equivalent electrons, the wave function of the system is antisymmetric with respect to the exchange of their coordinates:

$$\Psi(1,2,...N) = -\Psi(1,2,..N) \tag{1.11}$$

This requirement is fulfilled by expressing the wave function as a Slater determinant, which changes sign with permutation of the coordinates of two electrons. In the case of a

multi-electronic system, the generalised Slater determinant for N total electrons is as follows:

$$\psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \tag{1.12}$$

where χ_N represents single electron wave functions, or spin-orbitals [3]. In the context of a molecule, the single electron wave functions are molecular orbitals. Rows are labelled by the coordinates of each electron: (1), (2) $\cdots$ (N), whereas each column uses a different orbital function: χ_1, $\chi_2 \cdots$, χ_N. If the labels of (1) and (2) are exchanged, the rows of the determinant are exchanged; a general property of determinants is that the interchange of two rows leads to a change of sign. The expanded form of the determinant (ψ_{SD}) will have the opposite sign when a pair of electronic coordinates are switched, by switching rows within the determinant, thereby fulfilling the antisymmetry requirement. In the dis-allowed case of two equivalent electrons occupying the same spin-orbital, two columns would be identical [4]. The evaluation of the determinant would then be zero, indicating that the probability of two electrons with identical spin occupying the same orbital was zero.

1.1.3 Variational principle

In order to obtain the ground state energy of a system, the wave function giving the lowest energy must be found. This corresponds to the electronic configuration with lowest value of E_{el}. Difficulty then arises, as ground state energy cannot be computed exactly. The variational theorem states that the calculated energy of any guess wave function can only be greater than or equal to the real ground-state energy (E_0) of the system. This provides a criterion for selection of the best guess wave function, as the energy is always bounded from below, where:

$$E = \langle \Psi | \mathbf{H} | \Psi \rangle \geq E_0 \tag{1.13}$$

for a normalised wave function. Thus, when choosing between different trial wave functions, the solution with the lowest energy is the one closest to the exact value.

1.1.4 Hartree-Fock self-consistent field method

In practice, equation 1.10 can only be solved *exactly* in very few circumstances; no exact solutions can be found for problems involving three or more interacting particles, such as in

the case of a helium atom possessing two electrons and a nucleus. For systems of complexity greater than one electron, further approximations must be made.

The Hartree-Fock (HF) approximation was the first practically applicable method for calculation of the ground-state energy of atoms with fixed nuclear positions. The self-consistent field (SCF) method was proposed by Hartree in 1928 [5, 6]. N electrons are treated as individual particles occupying single-electron spin orbitals and move independently of the dynamics of any other fermions in the system. The effective interaction of one electron with all other fermions is averaged and applied as a static external field, in the form of a spherical potential around the electron (called the mean-field approximation). In this way, the N-body problem is reduced to a 1-body problem. The approximation neglects exchange in the electron-electron interaction; the calculated Hartree wavefunction alone does not include any contribution from electron correlation, and incorrectly implies that the electrons are distinguishable. Fock developed this idea by introducing Slater determinant wave functions [7]. The effects of exchange on the coulombic repulsion were incorporated, achieved by taking the trial wave function as a single Slater determinant.

For an electron in orbital χ_i in the field of nuclei M and other electrons χ_j, the Hamiltonian operator is comprised of three terms, corresponding to the three contributions to the energy. The core Hamiltonian operator, $\mathbf{H}^{core}$ comprises of the kinetic energy of each electron, and the electron-nuclear interaction:

$$\mathbf{H}^{core}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{1A}} \tag{1.14}$$

where Z_A is the nuclear charge and r_{1A} is the separation of electron (1) with nucleus A. In a mono-electronic system, this would be the only operator present.

The coulomb operator, $\mathbf{J}_j$ corresponds to the averaged interaction potential between each pair of electrons in the same orbital, and with other electrons in other orbitals χ_j:

$$\mathbf{J}_j(1) = \int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_j(2) \tag{1.15}$$

where $d\tau_i$ indicates the integration is over the spatial and spin coordinates of electron i, and r_{12} is the distance between the two electrons.

The exchange operator $\mathbf{K}_j$, is only non-zero for electrons with the same spin, arising

due to the antisymmetry of the wavefunction:

$$\mathbf{K}_j(1)\chi_i(1) = \Big[\int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_j(2)\Big]\chi_j(1) \tag{1.16}$$

defined in terms of its effect when acting on χ_i. The Hamiltonian operator written in terms of its core, coulomb and exchange contributions is as follows:

$$\Big[\mathbf{H}^{core}(1) + \sum_{j=1}^{N}\{\mathbf{J}_j(1) - \mathbf{K}_j(1)\}\Big]\chi_i(1) = \sum_{j=1}^{N}\varepsilon_{ij}\chi_j(1) \tag{1.17}$$

and can be simplified to:

$$\mathbf{F}_i\chi_i = \sum_j \varepsilon_{ij}\chi_j \tag{1.18}$$

where $\mathbf{F}_i$ is the Fock operator, and ε_{ij} is the energy of orbital χ_j. The operator is a one electron Hamiltonian for an electron in a multi-electron system. For "closed shell" problems where there are no unpaired electrons, the operator has the form:

$$\mathbf{F}_i(1) = \mathbf{H}^{core}(1) + \sum_{j=1}^{N/2}\{2\mathbf{J}_j(1) - \mathbf{K}_j(1)\} \tag{1.19}$$

$\mathbf{H}^{core}$ consisting of the kinetic energy terms can be solved exactly, the electron-electron repulsion $\mathbf{J}_j$ must apply the mean-field approximation but the exchange component $\mathbf{K}_j$ is solved iteratively. Starting with an initial guess wave functions for the occupied orbitals χ_i, solution of the one-electron HF eigenvalue equations generates a new set of orbitals. Applying the variational principle, the spin orbitals are varied to minimise the energy. This process propagates using the newly generated orbitals for the next optimisation, until the difference between the final solution and its previous iteration falls within an acceptable threshold and is "self-consistent".

1.1.5 Open shell systems

The forced pairing of electrons of opposing spin into a shared orbital is referred to as the *restricted* scheme (figure 1.1). For closed shell systems, this treatment is appropriate. For species with unpaired electron spin such as in transition metal complexes or radicals, an alternative model allowing singly occupied orbitals must be adopted. The *restricted-open* scheme maintains electron pairing within orbitals except in the case of the highest occupied molecular orbital (HOMO), which is singly occupied. An alternative model is the

unrestricted scheme, where all electrons are unpaired and reside in their own orbitals. A caveat of the unrestricted model is its susceptibility to spin contamination, which has consequences at large bond separations. This artificial mixing of spin states leads to a lowering of the obtained energies when compared to the restricted variants [8].

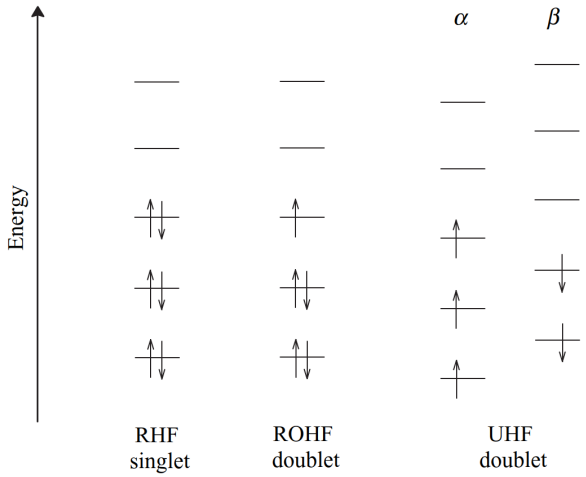

Figure 1.1: The electron ordering schemes corresponding to the restricted Hartree-Fock (HF), restricted-open HF (ROHF) and unrestricted HF (UHF) methods of calculation for closed and open shell systems [4].

1.1.6 Electron correlation

The energy difference between the real energy and the result obtained from HF is called the correlation energy, $E_{corr.}$.

$$E_{corr.} = E_{exact} - E_{HF} \tag{1.20}$$

The $E_{corr.}$ term can be divided into two components. The static correlation component arises as a system cannot be fully described by a single set of molecular orbitals (MO)s, and the dynamic correlation contribution derives from the neglect of instantaneous electron repulsion interactions. The latter includes the description of instantaneous dipolar interactions, leading to van der Waals forces, which are lost when the electron repulsion terms are averaged. Post-HF methods such as perturbation theory and coupled-cluster techniques aim to account for the difference by inclusion of the contribution from correlation as an additive term, or *via* multi-electron wave functions. However, these methods become prohibitively expensive with increasing numbers of electrons, such that the system size is limited to small molecules for calculations of high accuracy. The high computational demand associated with handling a many-electron wave function is circumvented in density functional theory (DFT), by expression of the total energy in terms of electron density.

1.2 Density functional theory

There are two approaches for solving the Schrödinger equation for a polyatomic system with many electrons. *Ab initio* methods generate solutions from "first principles", without information gained from experimental results. By contrast, *semi-empirical* methods deal with parameters fitted to experimental quantities, such as enthalpies of formation or dipole moments. DFT derives from the Thomas-Femi-Dirac model, whereby electron correlation is modelled via functionals of the electron density, $\rho(r)$. Currently, it forms the most widely used approach for quantum mechanics (QM) problems. When compared to HF and post-HF methods, DFT provides increased computational efficiency. Modern hybrid functionals are able to produce results on the order of MP2 accuracy, utilising only the resource of a HF calculation. The NC monomers, dimer and trimer models examined in this study consist of between 30 - 75 atoms; DFT provided the best pay-off between accuracy and efficiency for application on a system of this size.

1.2.1 Hohenberg-Kohn formalism

Modern DFT is based on two fundamental theorems proposed by Hohenberg and Kohn in 1964 [9]. The first theorem states that for the ground state of system, there exists a unique energy and non-degenerate electron density. The density can therefore be used to determine the Hamiltonian of a system, thereby also describing its ground state energy $E[\rho(r)]$, wave function and other properties of the system. The energy is a functional of the density:

$$E[\rho(r)] = \int \rho(r)V(r)\,d(r) + F[\rho(r)] \tag{1.21}$$

where $V(r)$ is the external potential, with the first term of the equation arising from the interaction of electrons with $V(r)$ (usually a coulombic attraction between electrons and nuclei). $F[\rho(r)]$ is a universal functional of the density, representing the total kinetic energy and electron-electron repulsion. It is not possible to explicitly express $F[\rho(r)]$ in terms of $\rho(r)$, so its exact form is not known.

The second theorem states that the ground state energy can be obtained *via* minimisation of $E[\rho(r)]$. Since equation 1.21 gives the exact energy of the original Hamiltonian, by applying the variational principle, the lowest possible value of $E[\rho(r)]$ gives the real solution for the ground state energy, and therefore $\rho(r)$. It is not possible to verify that the found $\rho(r)$ giving the lowest value of $E[\rho(r)]$ corresponds to a wave function obeying the Pauli

exclusion principle requirement for antisymmetry. This problem and the unknown identity of $F[\rho(r)]$ were addressed by the Kohn-Sham equations.

1.2.2 Kohn-Sham DFT

The Kohn-Sham scheme establishes a system with N non-interacting electrons, in a similar manner to HF [10]. The wave function is described by a single Slater determinant of one-electron orbitals, and the electron density is set to be identical to that of the exact ground state wave function. Using this approximation, the energy of the system can again be divided up into its component contributions:

$$E[\rho] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_V[\rho(r)] + E_{XC}[\rho(r)] \tag{1.22}$$

where E_{KE} is the kinetic energy of the non-interacting electrons, E_H the Hartree electrostatic energy corresponding to the electron-electron repulsion between electrons, E_V the interaction between the electrons and the external potential due to the nuclei, and E_{XC} the exchange-correlation energy, encapsulating non-classical exchange and correlation contributions not accounted for by the other terms. Referring back to the Born interpretation (equation 1.1), the density can be obtained from the sum of the square moduli of the wave function:

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 = |\Psi|^2 \tag{1.23}$$

Aside from the kinetic energy term in equation 1.22, the remaining terms can be summarised into an effective potential v_{eff}:

$$v_{eff} = v_H(r) + v_V(r) + v_{XC}(r) \tag{1.24}$$

where $v_H(r)$ is the electron repulsion potential and $v_V(r)$ the electron-ion potential. The exchange-correlation potential $v_{XC}(r)$ is the functional derivative of E_{XC}. If E_{XC} is known then $v_{XC}(r)$ can be computed from the following equation:

$$v_{XC}(r) = \frac{\delta E_{XC}}{\delta \rho} \tag{1.25}$$

The kinetic energy term in equation 1.22 can be expressed in terms of the one-electron

wave function:

$$E_{KE}[\rho(r)] = -\frac{1}{2}\nabla^2 \sum_{i=1}^{N} \langle \psi_i | \nabla^2 | \psi_i \rangle \tag{1.26}$$

Combining equations 1.24 and 1.26, a new Hamiltonian can be written, only considering the non-interacting system:

$$\mathbf{H} = -\frac{\nabla^2}{2} + v_{eff} \tag{1.27}$$

Using the Kohn-Sham formulation of the Schrödinger equation, the one-electron orbitals $\psi_i(r)$ have the form:

$$\Big(-\frac{\nabla^2}{2} + v_{eff}(r)\Big)\psi_i(r) = \varepsilon_i \psi_i(r) \tag{1.28}$$

where $v_{eff}(r)$ is the effective potential, $\psi_i(r)$ the Kohn-Sham orbitals and ε_i the Kohn-Sham orbital energies.

The Kohn-Sham equations are solved self-consistently. Evaluation of equation 1.22 gives the total electronic energy. A guess density is supplied for the initial evaluation of equation 1.28, to generate a set of orbitals. This in turn informs the next iteration with an improved density value, until convergence is reached.

1.2.2.1 Exchange-correlation functionals

Local density approximation

In practice, the exact solution for E_{XC} in equation 1.24 is not known, so an approximation is used. The simplest method is the local density approximation (LDA), based upon a uniform electron gas. ε_{XC} is calculated per electron as function of the density, and integration over all space gives the E_{XC} for the whole system:

$$E_{XC}^{LDA}[\rho(r)] = \int \rho(r)\varepsilon_{XC}[\rho(r)]\,dr \tag{1.29}$$

This method has demonstrated good results for structural properties such as bond lengths and lattice constants, often improving on HF results. The model falls down when considering systems with many molecules, overestimating binding energies and atomisation energies, and performing worse than HF for open shell atoms.

Generalised gradient approximation

To account for inhomogenous electron density present in real systems, the local gradient of the density $\nabla\rho(r)$ can be taken into account at each coordinate, on top of the existing dependency on the density $\rho(r)$. This gradient corrected approach is called the generalised

gradient approximation (GGA):

$$E_{XC}^{GGA}[\rho(r)] = \int \rho(r)\nabla[\rho(r)]\,dr \tag{1.30}$$

The results obtained using GGA greatly improves upon LDA, such as in the calculation of bond dissociation energies. The GGA is split into its exchange and correlation contributions, which can be solved individually:

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \tag{1.31}$$

The function of $\nabla\rho(r)$ is not uniquely defined and no true form is known; further developments saw the proposal of numerous gradient correction schemes, often fitting to experimental parameters. Functionals that have been well established in literature include:

- B88: the exchange functional developed by Becke, and contains an empirical parameter β=0.0042 fitted to give the best agreement with the HF energies of the noble gases [11].
- LYP: a widely used GGA correlation functional, developed by Lee, Yang and Parr, with empirical fitting to the helium atom. Combination with the B88 exchange functional gives the BLYP method [12].
- PBE: the exchange-correlation functional developed by Perdew, Burke and Ernzerhof derived purely from *ab initio* calculations [13]. The PBE family of functionals have parameterisations optimised for different materials and interfaces, such as small molecules, crystal structures and metal surface energies [14, 15, 16].

1.2.3 Hybrid functionals

In DFT, the influence of the exchange contribution to E_{XC} is significantly larger than that of correlation, which only adds minor corrections [17]. In hybrid DFT, a portion of exact exchange is introduced into the DFT exchange energy *via* linear combinations of HF and GGA exchange. This assists to counteract the self-interaction problem, a significant source of error in most approximate exchange-correlation functionals used for Kohn-Sham DFT (KS-DFT) calculations. This spurious interaction of an electron with itself does not appear in HF where the exchange is defined exactly, and a cancelling of the coulomb and exchange-correlation energies occurs [18]. The description of the exchange-correlation is imperfect in

DFT; the exchange and coulomb energy terms do not completely cancel out, and an energy contribution is experienced even in a one electron system. An observable manifestation of this in larger systems is the underestimation of reaction barriers [19].

Perhaps the most widely used functional of this category is Becke, 3-parameter, Lee-Yang-Parr hybrid functional (B3LYP) [20, 21]. It is employed in the current study to examine the degradation reactions of NC. The formalism is as follows:

$$E_{XC}^{B3LYP} = E_X^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_X(E_X^{B88} - E_X^{LDA}) + E_c^{LDA} - a_C(E_C^{LYP} - E_C^{LDA}) \quad (1.32)$$

a_0, ca_X and a_C are parameters fitted to experimental atomisation energies, ionisation potentials, proton affinities and atomic energies. The semi-empirical coefficients have values of 0.20, 0.72 and 0.81, respectively, indicating that the B3LYP hybrid functional contains 20% HF exchange, 72% B88 exchange and 81% electron correlation contribution from LYP. B3LYP is of particular interest in the study of NC as it has been applied in comparable computational investigations on glucose and its analogues, which forms the base structure of the NC monomer [22, 23].

There are documented cases where B3LYP performs extremely poorly, such as in the description of $\pi \rightarrow \sigma$ structural transformations [24], and medium-long range dispersion interactions [25]. In the first case, no similar transitions are expected for the reactions studied in this work. When considering the medium-range electron correlation reactions that may occur for the investigated species, the distance at which the intermolecular interactions take place and at which B3LYP performance suffers (1.5-3.5 $\AA$), is of relevance. The calculations falling within this range will therefore be performed with both B3LYP and the ωB97X-D long-range corrected hybrid functional (ωB97X-D), in order to validate and compare results.

Grimme proposed an empirical DFT method with dispersion correction to include van der Waals interactions, described as density functional theory with dispersion correction (DFT-D) [26, 27]. The general form of the DFT-D scheme is used to calculate the total energy of the system, with the addition of an empirical atomic-pairwise dispersion correction:

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \quad (1.33)$$

The ωB97X-D dispersion corrected functional was presented by Chai *et. al*, based on the

ωB97X, which is used to calculate the KS-DFT energy, E_{KS-DFT} [28, 29]. The approach is founded on Becke's B97-D GGA functional [30, 27]. For the B97-D functional, expansion coefficients were added to the original B97 and optimised to reduce the density functional at small separations where lower electron correlation was expected. At medium-long ranges, the density functional description was dictated by the semi-empirical dispersion correction term. In the case of ωB97X-D, the unscaled dispersion correction is defined as follows:

$$E_{disp} = -\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{C_6^{ij}}{R_{ij}^6}f_{damp}(R_{ij}) \tag{1.34}$$

where N is the number of atoms in the system, C_6^{ij} is the dispersion coefficient for atom pair ij, and R_{ij} is the interatomic distance. At short interatomic distances, dispersion is zero. The asymptotic, pairwise van der Waals (vdW) potentials are maintained by the damping function:

$$f_{damp}(R_{ij}) = \frac{1}{1+a(R_{ij}/R_r)^{-12}} \tag{1.35}$$

The function decreases to 1 as R_{ij} increases, but also diminishes quickly at small values of R_{ij} to prevent the divergence observed for the undamped case (figure 1.2). R_r is the sum of the vdW radii of the atoms ij, whilst a determines the strength of the dispersion corrections. The ωB97X-D functional incorporates 100% long-range exchange, 22% short-

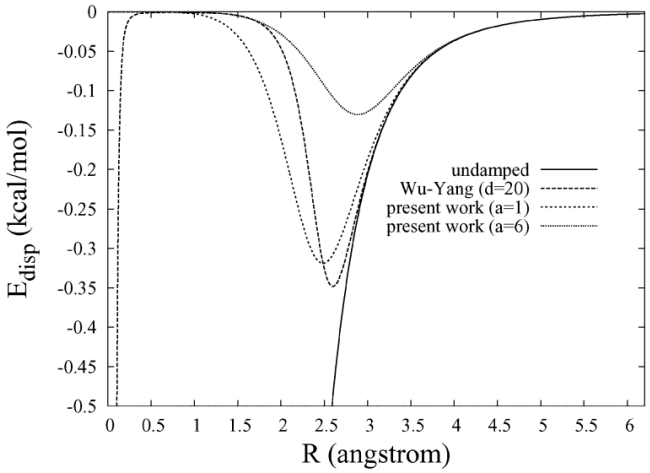

Figure 1.2: Dispersion energy of Ne_2 with and without the damping function (equation 1.35). The function of Wu and Yang ($f_{damp}^{WY}(R) = 1/(1+e^{-\frac{dR}{R_r-1}})$) is also compared [31]. Reproduced from the work of Chai *et. al* [28, 29] with permission from the PCCP Owner Societies.

range exchange, a modified B97 short-range exchange and B97 correlation, with empirical

dispersion corrections. When the constraint ω is reduced to 0, ωB97X-D is equivalent to the B97 functional with the addition of the dispersion correction. The performance of ωB97X-D will be compared to that of B3LYP in this work with particular attention paid to the reaction energies.

1.2.4 Basis set approximation

A basis set is the collection of mathematical basis functions, used in linear combination, to construct the molecular orbitals (MO). An individual MO can be defined as:

$$\psi_i = \sum_{\mu=1}^{N} c_{\mu i}\chi_\mu \tag{1.36}$$

where $c_{\mu i}$ is the molecular orbital expansion coefficient and χ_μ corresponds to the one-electron occupied orbitals, often atomic orbitals, also called *basis functions*. N is the total number of basis functions, which are all chosen to be normalised. The smallest possible basis set is a single basis per occupied orbital on each atom in a molecular system, termed the "minimal basis". The HF limit is when the addition of further bases does not lower the energy of the system any further.

Slater type orbitals (STO) with exponential dependence $Ae^{-\alpha r}$, are the intuitive choice for atomic functions; α controls how quickly the function decays (radial extent), with higher values for higher effective nuclear charge. They are extremely similar in their mathematical expression to the real atomic orbital, however, mathematically challenging to implement in molecular orbital calculations. Gaussian type orbitals (GTO) of dependence $Ae^{-\alpha r^2}$, offer a more practically viable approximation. Linear combinations of primitive gaussians are used to form the actual basis functions. The constructed basis functions are then called contracted gaussians, and have the form:

$$\chi_\mu = \sum_p d_{\mu p} g_p \tag{1.37}$$

where $d_{\mu p}$ are fixed constants. The expansion to the MO therefore takes the form:

$$\psi_i = \sum_\mu c_{\mu i}\Big(\sum_p d_{\mu p} g_p\Big) \tag{1.38}$$

In this study, Pople style basis sets will be employed, with general format **X-YZg** [32]. Split valence basis sets such as these, describe the core electrons using fewer basis functions than

the interacting valence electrons; the valence electrons play a much more significant part in bonding and intermolecular interactions. An example is **6-31+G***, where the core electrons are described using a single contracted gaussian consisting of six primitive functions. The valence orbital is split into two contracted gaussians, where one is described by three primitive gaussians, and the other, one primitive gaussian. The **+** sign corresponds to diffuse functions, the extension of the basis function as it tails away from the nucleus. This is particularly important when describing the behaviour of ions. The * indicates the addition of polarisation functions, in this case, d-functions to all non-H atoms. Higher order modifications also add p-functions to H atoms, additional d-functions to non-H atoms, d-functions to H atoms, with increasing levels of polarisation. This mixing of orbitals facilitates a better description for the correct shape of the MO. Increasing the number of basis functions on each occupied orbital allows for expansion or contraction, increasing flexibility in response to the environment.

Basis set superposition error BSSE is a false lowering of the energy that can occur when two species in a system with finite bases sets approach one another to form a complex. Particle A borrows the extra basis functions belonging to particle B and an artificial stabilisation is observed. The error arises from the inconsistency in treatment between the individual particles at large separations, and the complex at short distances. This is particularly pertinent when performing PES scans where bond breaking or formation is expected, and in the consideration of reaction energies. The effect is particularly pronounced for smaller basis sets. Counterpoise correction can be used to circumvent basis set superposition error (BSSE); each calculation is performed twice, once using the mixed basis set of the complex and once with the separate bases of the two species, and this energy difference is subtracted from the uncorrected energies [33, 34]. However this comes at the expense of higher computational cost. Preliminary investigations into the species examined in this study showed that the use of counterpoise correction offered minimal improvements to the energy, and the additional CPU time was not proportionate to the marginal improvements. Calculations were performed at sufficiently high basis that the effects of BSSE were insignificant, and counterpoise correction was not used for further investigations.

1.3 Implementation & Analysis

The QM calculations employed in this study typically involved a geometry optimisation, followed by a frequency calculation conducted in Gaussian 09 revision E.01 (G09) [35].

This section will provide an generic overview of the methods and techniques used to prepare input or analyse the optimised geometry output from a QM calculation. Subsequent investigations into transition state structures, reaction coordinates and analysis of critical bonding points will be explained. The details of individual calculation schemes are included in the methodology section at the the start of each chapter. This includes any details of the QM methods and basis sets chosen, individual optimisation procedures, types of calculation or special keywords used, and any non-standard software or technical details of importance.

1.3.1 Molecular mechanics geometry optimisation

In the case of the larger models in this study (60 atoms and above), QM geometry optimisation was a laborious and expensive process, particularly with the presence of numerous floppy O−H bonds and small methyl group rotations. In order to reduce the required time and computational cost, these structures were optimised using a low-level molecular mechanics geometry calculation to minimise large forces in the system, prior to the finer refinements made during the QM optimisation. Molecular mechanics MM uses classical Newtonian mechanics to model chemical structures and properties; the electrons are not explicitly handled, bypassing the Schrödinger's equation [36, 4]. Molecules are instead treated as a "ball on a spring", where the strength of interaction is defined by the atom types and force constants.

Each molecular mechanics (MM) approach is defined by its force field, where the energy of the system is described by a parametric function, fitted to experimental or high level computational data. Different force fields are therefore fitted to different chemical systems.

The general functional form of a force field is:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{1.39}$$

where E_{FF} is the energy of the system. The first three terms can be referred to as the 'bonding' terms: E_{str} is the energy function for bond stretching between two atoms, E_{bend} the energy required to bend a bond angle and E_{tors} is the energy of rotation around a bond. The remainder are thus called 'non-bonding' terms: E_{vdw} and E_{el} refer to the energy of non-bonding van der Waals interactions and of electrostatic interactions, respectively. E_{cross} defines the coupling of the first three terms.

In this study, the universal force field (UFF) [37] and Merck molecular force field 94 (MMFF94) [38, 39, 40, 41, 42] force fields are tested on NC dimer and trimer structures. UFF is an empirical, all-element model, in principle able to describe the whole periodic table. It has been parameterised on individual elements, hybridization and connectivity; all the atom types in the NC unit are included, and UFF has successfully been applied in the study of cellulose dissolution [43]. MMFF94 is a forcefield parameterised from high level QM calculations for small molecules, organics and proteins. It has been used to sample different torsional conformers of cellulose [44] and to study the dynamics of bio-active polysaccharides [45].

1.3.2 Potential Energy Surface scans

Geometry scans or PES scans were used to probe the local energy landscape corresponding to a specific change in geometry. During the course of a scan, a structural property - such as a selected bond length, angle or dihedral is adjusted in incremental steps, as specified by the given scan parameters. In the case of a relaxed scan, at each step, the adjusted parameter is frozen and a geometry optimisation is performed, allowing the rest of the system to relax around the modified bond. Each scan yields a PES of the explored pathway. In the case of bond breaking or formation, the presented energy profile is a reaction co-ordinate diagram. In these cases, an energy maximum followed by a trough indicates a transition state and intermediate reaction product, respectively. The structural co-ordinates at the peak of the curve can be extracted and used for subsequent frequency calculations or a transition state search, to validate the mechanistic pathway. Intrinsic reaction coordinate calculations also test the reactant and product along the reaction pathway, to ensure that they lie on the same PES. To explore the predicted degradation mechanisms of NC, the scanned parameters were assigned to the bonds undergoing the most significant transformation during a particular step of the mechanism. In the case that more than one significant bond was altered, multiple scans with different bond specifications were compared. In this way, multi-dimensional scans could be performed. Two-dimensional scans were used to probe simultaneous processes in the system, however, these proved computationally intensive due to the high number of coordinate points to be calculated. Relaxed geometry scans were performed on the optimised reactant geometry using the Opt=ModRedundant keyword to specify the scanned coordinate, the number of scan steps and the method for energy optimisation. A rigid scan consisted of a single point energy calculation of the structure at each

step, rather than full relaxation of the wider system, as in the relaxed scans.

1.3.3 Transition state searches

Transition state searches are called through the Opt=TS, QST2 or QST3 keywords in G09. The Opt=TS method attempts to optimise an input "guess" geometry to a maximum. The guess geometry can be drawn by hand, obtained *via* a PES scan, from coordinates in literature or generated using the QST2 function. A transition state (TS) can be identified by the single imaginary frequency corresponding to the vibration transitioning from product to reactant. Often, a TS calculation alone will not be able to isolate the correct transition state and is usually used in conjunction the QST2 or QST3 methods. The QST2 option is able to generate a transition state geometry using the Synchronous Transit Quasi-Newton (STQN) method [46]. Here, the transition geometry is assumed to be around midway between a given reactant and product; the calculation interpolates between the starting and end point, probing the energy profile. Thus, the corresponding atom labels must match in both the reactants and products. QST3 performs a similar function, but also considers an input guess transition state structure. It is widely acknowledged that transition state searching is challenging. In addition to the techniques above, the task requires a certain measure of chemical intuition when constructing guess structures, as the calculation is sensitive to small changes in geometry.

1.3.4 Solvent model

To account for solvent effects in the reactions of NC, a self-consistent reaction field (SCRF) method was applied [47]. These methods treat the solvent as a uniform dielectric constant (ε) in the background.

The solvated molecule is placed in a cavity within the SCRF. The solute dipole induces a dipole in the solvent medium surrounding the cavity, which in return leads to an electric field within the cavity called the reaction field. The shape of the cavity is dictated by the the specific SCRF approach; in this study, the polarisable continuum model (PCM) is used [48, 49]. PCM models the cavity as interlocking van der Waals radii around each atom. The charges at the van der Waals surface are split into a series of small rectangles and the value of the point charge at that surface element is calculated by:

$$q_i = -\left[\frac{\varepsilon - 1}{4\pi\varepsilon}\right] E_i \Delta S \tag{1.40}$$

where ε is the dielectric, E_i the electric field gradient and ΔS the area of the surface element. The coulombic contribution of each surface element to the other charges are calculated and iteratively refined until self-consistent [50]. This calculated charge potential (ϕ_r) is then added to the solute Hamiltonian:

$$\mathbf{H} = \mathbf{H}_0 + \phi_r \tag{1.41}$$

and the SCF reaction continues as normal. New surface charges are calculated at each SCF cycle until the overall surface charge is also self-consistent for the optimised system. As this method only includes the electrostatic contribution, an obvious limitation to this technique is that any reactions involving direct interaction with solvent molecules, such as stablisation by solvation shells *via* van der Waals or hydrogen bonding interactions, will be missed. In these cases the explicit solvent molecules should be added. However, for models where explicit interaction with the solvent is not required but solvent dieletric properties strongly influence the reaction environment, these methods can be very effective. This method has demonstrated successful application in the study of cellulose and glucose reactions in solvent [51, 52]. For calculations in solvent, the SCRF=PCM keyword will therefore be included in G09 calculation specifications, with water (ε=78.4) as the default solvent.

1.3.5 Topology analysis using the quantum theory of atoms in molecules

Topology analysis is a method of obtaining useful properties from the 3D representation of the electron density obtained from a QM calculation. A quantitative way to obtain information on the topology of the electron density, is by taking the first derivative of the gradient ($\nabla(\rho)$). In 1991, Bader proposed a technique to analyse the electron density using the quantum theory of atoms in molecules (QTAIM) [53]. The points in the topological landscape at which $\nabla(\rho)$ is zero (excluding points at infinity) signify a stationary point. With respect to intermolecular interactions, these correspond to interaction centres, and are deemed Critical Points (CP). The matrix of partial second derivatives of the gradient is referred to as the *Hessian*. The Hessian is a (3×3) symmetric matrix; diagonalisation sets the off-diagonal terms to zero and generates the principal axes of curvature. The sum of the diagonal terms returns the laplacian of the electron density ($\nabla^2(\rho)$); evaluation of $\nabla^2(\rho)$ identifies the characteristic of the critical point (CP). In the context of electron density, the CP can be classified into four types based on the number of negative eigenvalues of the Hessian (table 1.1):

Table 1.1: Features of different types of critical point from QTAIM topological analysis.

Critical Point	Label	Negative eigenvalues	Attribute	Representation
Nuclear (NCP)	(3,-3)	3	Local maximum	Atomic nuclei
Bonding (BCP)	(3,-1)	2	2^{nd} order saddle point	Bonding site
Ring (RCP)	(3,+1)	1	1^{st} order saddle point	Steric point or centre of ring system
Cage (CCP)	(3,+3)	0	Local minimum	Centre of cage system

Nuclear critical points (NCP) are so called, as they are generally located at atomic nuclei positions. All three eigenvalues of the Hessian matrix of the function are negative, corresponding to a local maximum on the electron density landscape. The number of NCP is usually equal to the number of atoms, though there are exceptions, such as in Li_2 which has a greater number of NCP than atoms, and for KrH^+, with a lower number. When two eigenvalues of the Hessian matrix of a function are negative, a second-order saddle point is present. These sites are usually located between attractive atom pairs, and so are referred to as a bonding critical point (BCP). As the electron density at the BCP is closely related to bond strength, the magnitude of $\nabla^2(\rho)$ will give an indication of the bonding type [54]. If one eigenvalue of the Hessian is negative, there exists a first-order saddle point. This generally appears in the center of ring systems and highlights a local steric effect, hence it is known as a ring critical point (RCP). When none of the eigenvalues are negative, it corresponds to the local minimum. For electron density analysis, these points generally appear in the center of cage systems, such as in pyramidal structure. These are referred to as a cage critical point (CCP).

The positions of CPs are searched by the Newton method, an initial guess point must be assigned; the minimisation will always converge to the CP that is closest to the guess point. By assigning different guesses and iterating over each of them, all CPs can eventually be found. Once searches of CPs are finished, the Poincaré-Hopf relationship [55] can be used to verify that the obtained topology is self-consistent, and that all CPs are found. For an isolated system, the relationship states that:

$$n(3,-3)\text{–}n(3,-1)+n(3,+1)\text{–}n(3,+3)=1 \tag{1.42}$$

The gradient path linking BCP, or two local maxima, can be referred to as a "bond path". Presence of a bond path indicates an atomic interaction path available for a variety bonding interactions; these will lie along covalent bonds, as well as non-covalent interactions such as intra molecular bonding or H-bonding. A network of bond paths is known as molecular graph. Scrutiny of the graph returns a reliable image of the geometry of the molecule.

Bibliography

[1] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049–1070, Dec 1926.

[2] Christopher J. Cramer. *Essentials of computational chemistry: Theories and models*, chapter 9, pages 338–340. Number 2. Wiley, 2004.

[3] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929.

[4] Frank Jensen. *Introduction to Computational Chemistry*, chapter 2, pages 22–35. Wiley, 2 edition, 2007.

[5] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part iii. term values and intensities in series in optical spectra. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(3):426–437, 1928.

[6] D. R. Hartree. The distribution of charge and current in an atom consisting of many electrons obeying dirac's equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 25(2):225–236, 1929.

[7] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1):126–148, 1930.

[8] Ambili S. Menon and Leo Radom. Consequences of spin contamination in unrestricted calculations on open-shell species: Effect of hartree-fock and møller-plesset contributions in hybrid and double-hybrid density functional theory approaches. *J. Phys. Chem. A*, 112(50):13225–13230, December 2008.

[9] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

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

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

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

[13] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[14] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Phys. Rev. B*, 59:7413–7421, Mar 1999.

[15] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, Apr 2008.

[16] Zhigang Wu and R. E. Cohen. More accurate generalized gradient approximation for solids. *Phys. Rev. B*, 73:235116, Jun 2006.

[17] Jürgen Gräfenstein and Dieter Cremer. The self-interaction error and the description of non-dynamic electron correlation in density functional theory. *Theoretical Chemistry Accounts*, 123(3):171–182, 2009.

[18] Junwei Bao, Laura Gagliardi, and Donald G. Truhlar. Self-interaction error in density functional theory: An appraisal. *J. Phys. Chem. Lett.*, 9(9):2353–2358, May 2018.

[19] Igor Ying Zhang, Jianming Wu, and Xin Xu. Extending the reliability and applicability of b3lyp. *Chem. Commun.*, 46(18):3057–3070, 2010.

[20] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, April 1993.

[21] Axel D. Becke. A new mixing of hartree-fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, January 1993.

[22] Udo Schnupf, J. L. Willett, Wayne B. Bosma, and Frank A. Momany. Dft study of α- and β-d-allopyranose at the b3lyp/6-311++g** level of theory. *Carbohydrate Research*, 342(2):196–216, 2007.

[23] F. A. Momany, M. Appell, G. Strati, and J. L. Willett. B3lyp/6-311++g** study of monohydrates of α- and β-d-glucopyranose: hydrogen bonding, stress energies, and effect of hydration on internal coordinates. *Carbohydrate Research*, 339(3):553–567, 2004.

[24] Susan N. Pieniazek, Fernando R. Clemente, and Kendall N. Houk. Sources of error in dft computations of cc bond formation thermochemistries: *pi→sigma* transformations and error cancellation by dft methods. *Angewandte Chemie International Edition*, 47(40):7746–7749, September 2008.

[25] Peter R. Schreiner. Relative energy computations with approximate density functional theory–a caveat! *Angewandte Chemie International Edition*, 46(23):4217–4219, June 2007.

[26] Stefan Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25(12):1463–1473, September 2004.

[27] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, November 2006.

[28] Jeng-Da Chai and Martin Head-Gordon. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.*, 128(8):084106, February 2008.

[29] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10(44):6615–6620, 2008.

[30] Axel D. Becke. Density-functional thermochemistry. v. systematic optimization of exchange-correlation functionals. *J. Chem. Phys.*, 107(20):8554–8560, November 1997.

[31] Qin Wu and Weitao Yang. Empirical correction to density functional theory for van der waals interactions. *J. Chem. Phys.*, 116(2):515–524, December 2001.

[32] R. Ditchfield, W. J. Hehre, and J. A. Pople. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.*, 54(2):724–728, January 1971.

[33] Frans B. van Duijneveldt, Jeanne G. C. M. van Duijneveldt-van de Rijdt, and Joop H. van Lenthe. State of the art in counterpoise theory. *Chem. Rev.*, 94(7):1873–1885, November 1994.

[34] Ł. M. Mentel and E. J. Baerends. Can the counterpoise correction for basis set superposition effect be justified? *J. Chem. Theory Comput.*, 10(1):252–267, January 2014.

[35] 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 revision e.01. Gaussian Inc. Wallingford CT 2009.

[36] J. B. Foresman and Æ Frisch. *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc.: Wallingford, CT, 2 edition, 1996.

[37] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.

[38] Thomas A. Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996.

[39] Thomas A. Halgren. Merck molecular force field. ii. mmff94 van der waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, 17(5-6):520–552, 1996.

[40] Thomas A. Halgren. Merck molecular force field. iii. molecular geometries and vibrational frequencies for mmff94. *Journal of Computational Chemistry*, 17(5-6):553–586, 1996.

[41] Thomas A. Halgren and Robert B. Nachbar. Merck molecular force field. iv. conformational energies and geometries for mmff94. *Journal of Computational Chemistry*, 17(5-6):587–615, 1996.

[42] Thomas A. Halgren. Merck molecular force field. v. extension of mmff94 using experimental data, additional computational data, and empirical rules. *Journal of Computational Chemistry*, 17(5-6):616–641, 1996.

[43] Faranak Bazooyar, Frank A. Momany, and Kim Bolton. Validating empirical force fields for molecular-level simulation of cellulose dissolution. *Computational and Theoretical Chemistry*, 984:119–127, 2012.

[44] Mckay W. Easton, John J. Nash, and Hilkka I. Kenttämaa. Dehydration pathways for glucose and cellobiose during fast pyrolysis. *J. Phys. Chem. A*, 122(41):8071–8085, October 2018.

[45] Anne Imberty and Serge Pérez. Structure, conformation, and dynamics of bioactive oligosaccharides: Theoretical approaches and experimental validations. *Chem. Rev.*, 100(12):4567–4588, December 2000.

[46] Niranjan Govind, Max Petersen, George Fitzgerald, Dominic King-Smith, and Jan Andzelm. A generalized synchronous transit method for transition state location. *Computational Materials Science*, 28(2):250–258, 2003.

[47] O. Tapia and O. Goscinski. Self-consistent reaction field theory of solvent effects. *null*, 29(6):1653–1661, June 1975.

[48] S. Miertuš, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum. a direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, 55(1):117–129, 1981.

[49] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chem. Rev.*, 105(8):2999–3094, August 2005.

[50] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2 edition, 2001.

[51] Heather B. Mayes and Linda J. Broadbelt. Unraveling the reactions that unravel cellulose. *J. Phys. Chem. A*, 116(26):7098–7106, July 2012.

[52] Zhenhuan Li, Kunmei Su, Jun Ren, Dongjiang Yang, Bowen Cheng, Chan Kyung Kim, and Xiangdong Yao. Direct catalytic conversion of glucose and cellulose. *Green Chem.*, 20(4):863–872, 2018.

[53] Richard F. W. Bader. A quantum theory of molecular structure and its applications. *Chem. Rev.*, 91(5):893–928, July 1991.

[54] Chérif F. Matta and Russell J. Boyd. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, chapter 1, page 11. WILEY-VCH Verlag, 2007.

[55] P. Balanarayan and Shridhar R. Gadre. Topography of molecular scalar fields. i. algorithm and poincaré-hopf relation. *J. Chem. Phys.*, 119(10):5037–5043, August 2003.