UNIVERSITY OF SOUTHAMPTON

Faculty of Natural Sciences
School of Chemistry

Large-scale Atomistic Simulations of Cellulose and Nitrated Derivatives

by

Catriona Buchanan Gibbon
MChem Chemistry with Maths

A thesis for the degree of
Doctor of Philosophy

December 2023

University of Southampton

Abstract

Faculty of Natural Sciences
School of Chemistry

Doctor of Philosophy

Large-scale Atomistic Simulations of Cellulose and Nitrated Derivatives

by Catriona Buchanan Gibbon

Nitrocellulose is a reactive derivative of cellulose, one of the most commonly occurring natural materials. Nitration of cellulose decreases the stability of the polymer, meaning there is less understanding of the structure and reactions. Insight into the reaction mechanisms can be important as it can provide information about the reaction constants, which can help increase understanding of the conditions needed for storage or use. Despite the importance of the chemical, the literature about nitrocellulose reactions available is limited; as the instability increases as the nitration increases, making it difficult to study the reaction at an atomic level.

This thesis used several different forms of computational chemistry to investigate nitrocellulose and to develop methods that could be used to increase understanding even further.

It began by using molecular dynamic simulations to create large-scale structures of different nitration levels, varying from 0–14.14 wt% nitrogen content. Cellulose is often found in fully crystalline forms, and nitrocellulose is more commonly found to be paracrystalline or amorphous. A protocol has been presented for creating realistic structures of nitrocellulose, focusing on the crystallinity of the systems being simulated. Results are presented and analysed for the creation of structures containing over 52,000 atoms. Comparisons of the different nitration structures meant that some hypotheses could be made as to what causes some different structures to be more likely to become paracrystalline than others.

The actual degradation reactions required quantum mechanical methods, specifically density functional theory (DFT) to study them. First of all these methods were used on nitrate esters, being used as a small analogue for nitrocellulose. Transition state searching methods and minimum energy pathway searches were used on nitrate ester structures to find potential mechanisms for the reaction. Although there were issues getting the calculations to fully converge, some suggested transition states were found for these, and some possible mechanisms were suggested.

Finally, these methods were then considered for larger systems, created through the methods already developed. The ring structure of nitrocellulose, not present in the nitrate esters tested, meant that more care needed to be taken when setting up the products and the reactants for the calculations. This meant that before insightful results could be reached, additional processes for setting up the correct structures needed to be defined. Algorithms were developed for generating these structures.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as: C. Gibbon, P. Di Pietro, M. Storr, D. Broughton and C. K. Skylaris. Using molecular dynamics to simulate realistic structures of nitrocellulose of different nitration levels. *Physical Chemistry Chemical Physics*, 10.1039/d2cp05550c, 2022

Signed:..... Date:.....

Acknowledgements

The first thanks have to go to Professor Chris Skylaris, my academic supervisor, for the support and guidance he has given me. Also to Poppy Di Pietro, and Mark Storr, my industrial supervisors, for all the advice and feedback you've given me.

Thanks to the other members of the Skylaris group, who have been good company over the years. Those who welcomed me - James Womack, Lucas Verga, Tom Ellably, Gabriele Boshchetto, Jamie Whipman, Gabriel Bramley, Dimitrios Matthas, and Han Chen, and those who joined since - Arihant Bhandari, Julian Holland, Davide Sarpa, Nabeel Anwar, Tom Demeyere, Apostolos Kordatos, Loukas Kollias and Brad Ayers. Most of all I want to thank Rebecca Clements, I probably wouldn't have started this without you, but I definitely wouldn't have got to the end of this!

To all the housemates who've helped me survive. James and Lucy, we got through lockdown together. Jo, we survived the trauma of the blue-hair/Pepsi situation. And Kerry has put up with me and my mess for the last few months!

Outside of work, two things have really kept me going, music and Scouting. So I want to thank Southampton University Concert Band, Woking Wind Orchestra, 2nd Guildford (Beaver) Scouts, and most all the team I worked with at 25th Southampton (Beaver) Scouts.

Thank you to Tanya and Amber, for being there with me forever. For listening to me complaining, whether you know what I'm talking about or not, and for giving me two exciting future roles to look forwards to playing: bridesmaid, and godmother!

Thank you to my family. Mum, Dad, Cameron. I honestly don't know if I've got the words to say how much you have supported and helped me.

To my family: Mum, Dad and Cam

Definitions and Abbreviations

Abbreviations:

AIMD	Ab Initio Molecular Dynamics
ASE	Atomic Simulation Environment
BFGS	Broyden-Fletcher-Goldfard-Shanno
BO	Born-Oppenheimer
CF	Coulson-Fischer
CG	Contracted Gaussians
CVFF	Constant Valence Force Field
COSMO	COnductor-like Screening MOdel
DFT	Density Functional Theory
DZ	Double Zeta
EDFT	Ensemble Density Functional Theory
ELF	Electron Localisation Function
GGA	Generalised Gradient Approximation
GTOs	Gaussian Type Orbitals
HF	Hartree-Fock
HK	Hohenberg-Kohn
KS	Kohn-Sham
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator
LDA	Local Density Approximation
LNV	Li-Nunes-Vanderbilt
LST	Linear Synchronous Transit
LYP	Lee–Yang–Parr
MD	Molecular Dynamics
MEP	Minimum Energy Pathway
mGGA	meta Generalised Gradient Approximation
MPI	Message Passing Interface
NEB	Nudged Elastic Band
NGWF	Non-orthogonal Generalised Wannier Function
NWChem	NorthWest Computational Chemistry
ONETEP	Order-N Electronic Total Energy Package
OPLS	Optimised Potentials for Liquid Simulations

OPLS-AA	Optimised Potentials for Liquid Simulations - All Atom
OPLS-UA	Optimised Potentials for Liquid Simulations - United Atom
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
PES	Potential Energy Surface
PIMs	Polymers of Intrinsic Microporosity
QM	Quantum Mechanics
QST	Quadratic Synchronous Transit
RDF	Radial Distribution Function
RHF	Restricted Hartree-Fock
ROHF	Restricted Open-shell Hartree-Fock
RMSD	Root Mean Square Deviation
SAS	Solvent-Accessible Surface
SCF	Self-Consistent Field
SD	Slater Determinant
STOs	Slater Type Orbitals
TDDFT	Time Dependant Density Functional Theory
TFD	Thomas-Fermi-Dirac
UEG	Uniform Electron Gas
UHF	Unrestricted Hartree-Fock
VMD	Visual Molecular Dynamics
XC	Exchange/Correlation

Chapter 1

Introduction

1.1 Introduction to Computational Chemistry

The laws of physics control every particle in the universe and the interactions between them. Therefore, it should be conceivable to use these laws to predict particle behaviours. However, whilst possible in principle, in practice the level of calculations required become too complex for pen and paper when looking at anything more than a couple of electrons or atoms.[1]

In the last century, with the development of computers, running these calculations became more plausible, and from the 1950s onwards, as electronic computers became available for general use, the discipline of computational chemistry started to evolve.[2]

Being able to simulate a chemical has many advantages. Money and time in labs can be saved by running simulations on a range of structures first, to test which are likely to give the best results. Simulations can also be useful in situations where materials of interest are high-cost, or too reactive. However, one of the most important parts of computational chemistry is ensuring that there is a baseline comparison between the model and experimental data, to build trust in the simulation, so computational chemistry cannot exist in isolation.

This work uses a mixture of different types of calculations to focus on different aspects of interest about nitrocellulose, the chemical of interest for this project. This chapter discusses the theory behind these calculations and the strengths and weaknesses of the different methods. Today, there is a large choice of software available, each tailored to a specific objective, so it is important to understand what sort of computational chemistry will fit the requirements of a project.[3,4] The next chapter discusses nitrocellulose, and why the choices of method and software were made to solve the specific issues being investigated.

1.1.1 Types of computational chemistry

This thesis uses two different types of computations - density functional theory calculations, to find energies and relaxed geometries of molecules, and molecular dynamics simulations, which involve evolving systems over time to find equilibrium thermodynamic and dynamical properties.

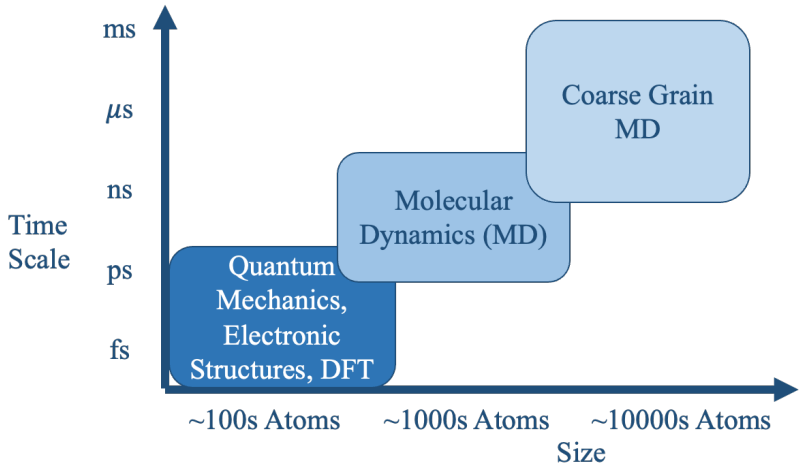


FIGURE 1.1: Different levels of computational chemistry calculations can be performed. The main factors used to judge what level to use are the size of the system, and the time scale being simulated.

One of the fundamental differences in the application of theory between Density Functional Theory (DFT) calculations and Molecular Dynamics (MD) simulations is based on how they interpret the Born-Oppenheimer (BO) approximation. This approximation states that electrons and nuclei differ so greatly in size, mass and speed, that they should be considered separately. In DFT the positions of electrons are calculated using quantum mechanics, whilst nuclei are kept in fixed positions. Molecular dynamics considers the movement of nuclei, and the timescales at which this is often concerned, the electrons are moving so fast that they can be assumed to average themselves out. This has the following consequences that must be considered when choosing which calculation type to use:

- Molecular dynamics simulations can handle larger systems than DFT, as they scale relative to the number of atoms in the system.
- DFT calculations scale with the number of electrons, and most codes are cubic scaling. This means that DFT is much more computationally expensive than molecular dynamics.
- Molecular dynamics simulations apply classical mechanics to the atoms by considering them as balls and springs. It can use time steps in the range of femtoseconds to see how the system evolves with time.

- DFT considers electrons, meaning it is possible to study interactions involving the formation and breaking of bonds. However, some specially developed force fields such as ReaxFF can make this possible in molecular dynamics as well.[5]
- Molecular dynamics needs a list of parameters for the system before it can begin, which are described in the force field (explained further in section 1.2.2 on page 9). This means that either the system being studied needs to already have a force field defined for it, or parameterisation needs to be performed, either using experimental data or DFT calculations.

A third technique that is not used in this project is Ab initio molecular dynamics (AIMD). Instead of using classical mechanics, from Newton's equations of motion, this uses quantum mechanical methods, and Schrödinger's equation.[6] This means that it can be used to study reactions, but it is also limited in size, the same as DFT.

In this thesis, MD will be used to create large-scale structures that are representative of a particular polymer in nature, and DFT will be used to study the reactions that lead to the breaking down of this polymer. To be a realistic representation of the structure specific properties must be reproduced, specifically the crystallinity and the density.

1.2 Molecular Mechanical Methods

Molecular dynamics is used to predict how the structures of molecules behave, how they will interact with other molecules, as well as how systems will progress and change over time. The calculations within molecular dynamics used classical mechanics, as opposed to quantum mechanics (QM), applying Newton's equations of motion to atoms and bonds, whilst treating them as balls and springs respectively. This requires the calculation of several different energies, explained in the following section.

1.2.1 Energy terms

At the centre of molecular mechanics calculations are force fields. A force field is a function that describes the potential energy of a system. Generally, these are written in terms of the cartesian coordinates of the atoms, the bond lengths, bond angles, dihedral angles, electrostatics, repulsion, and dispersion.

These terms are calculated by summing together a range of energy contributions by different physical properties.

$$E_{pot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdW} + E_{electrostatic} \tag{1.1}$$

A force field describes how each of these elements should be calculated, and the parameters required for them, such as bond force constants and equilibrium values.

The terms in Equation 1.1 on the preceding page are considered in two categories - bonded and non-bonded contributions. First the bonded terms, the bond length, the bond angle, the dihedral angles and the improper dihedral angles are discussed in the next section.

1.2.1.1 Bonded energy terms

A bond between two atoms has an equilibrium length. If the atoms get any closer together, they will begin to repel each other, and if they are separated they will try to move back together. This is like a spring and means that we can describe the energy of a bond using Hooke's Law

$$E_{bond\ stretching} = \sum_{bonds} K_r \left(r - r_{eq}\right)^2 \tag{1.2}$$

where r_{eq} is the equilibrium length of the bond, and K_r is the force constant, describing the stiffness of the bond. Figure 1.2 shows that this is not a perfect model for the stretching of a bond, as the bond reaches towards its dissociation length the model breaks down, however as molecular mechanics is not used to study the breaking of bonds, it is acceptable.

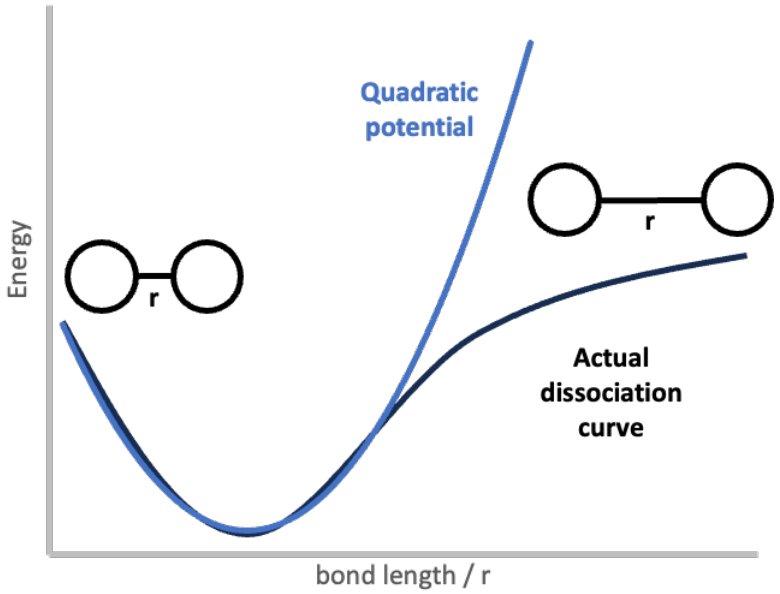


FIGURE 1.2: Bond length model. The difference between the actual dissociation curve (black) at bond length increases, compared to the quadratic harmonic model (blue) being applied to it.

Bond angles can be similarly described as varying harmonically, as shown in figure 1.3, and by using the equation:

$$E_{bond\ angle} = \sum_{bond\ angles} K_\theta \left(\theta - \theta_{eq}\right)^2 \tag{1.3}$$

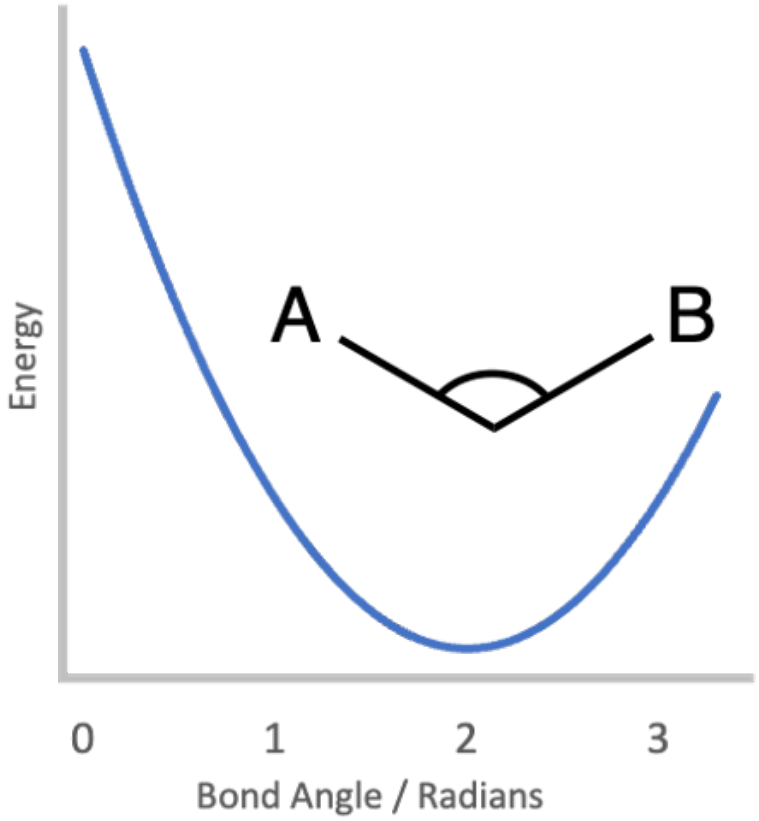


FIGURE 1.3: Bond angle model. The bond angles are varying harmonically with a minimum at θ_0.

Dihedral angles, also known as torsion angles, as shown in figure 1.4, are periodic, so they need a periodic function

$$E_{dihedral} = \sum_{dihedrals} \frac{V_n}{2} \left[1 + cos\left(n\phi - \gamma\right)\right] \tag{1.4}$$

Where V_n is the height of the barrier to rotation, which is affected by the hybridisation of the bond as well as the atom types, n is the number of minima as the bond is rotated, and γ is the energy of the global minimum. Figure 1.5 shows a plot of how the energy changes as the dihedral angle changes on an example of ethane-1,2-diol, showing how the global minimum is when the functional groups are in trans positions, although there are two other minima where the groups are in other non-eclipsed

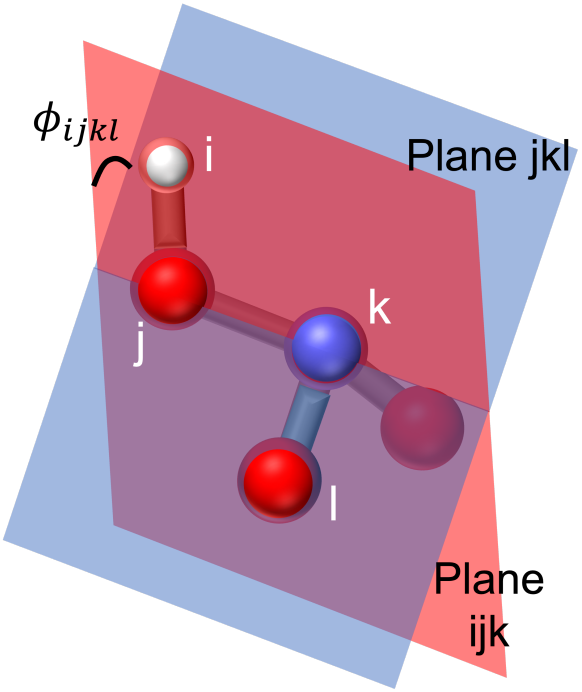


FIGURE 1.4: Diagram of a dihedral angle, shown here on HNO_3.

positions. All the maxima happen when the groups are in eclipsed positions, and the global maximum is where the -OH groups are cis-orientated.

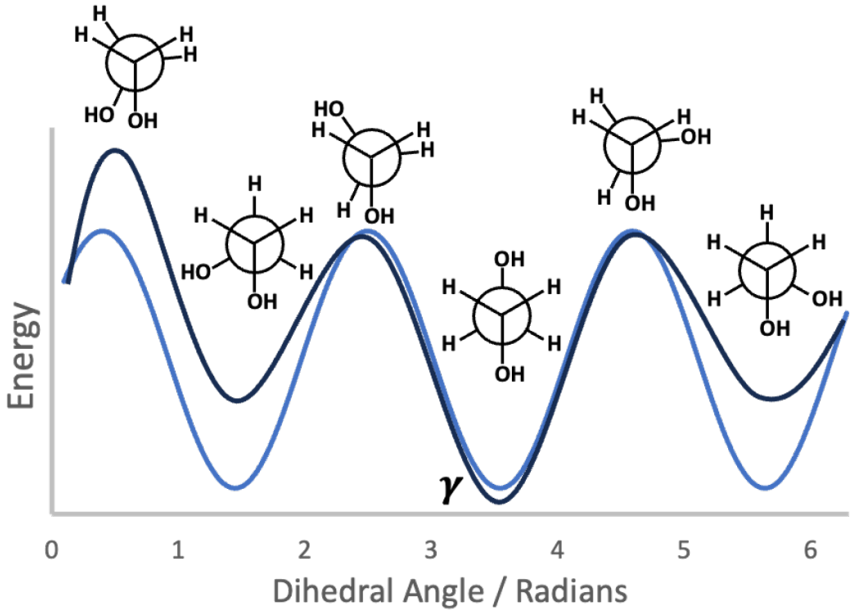


FIGURE 1.5: Dihedral angle model. The actual variation of the energy with the dihedral angle (black) compared to the sin function being used to model it (blue), where γ is the energy of the global minimum.

The final energy term to be defined is the improper dihedral, which can be described as the “out-of-plane bending” (figure 1.6).[7]

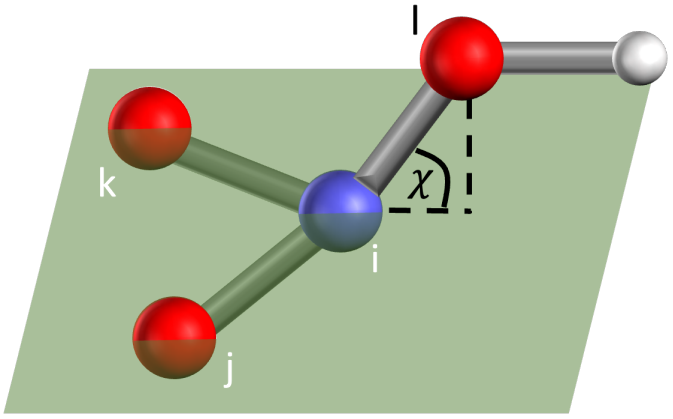


FIGURE 1.6: Diagram of an improper dihedral, modelled here on HNO_3.

The improper dihedral can be modelled in many different ways, for example, in this project the improper dihedral is described using CVFF (Constant Valence Force Field), meaning the potential is defined using 3 factors, shown in equation (1.5).[8]

$$E = K[1 + d\cos(n\phi_{ijkl})] \tag{1.5}$$

where K is a prefactor, d is either +1 or -1, n is an integer, and ϕ_{ijkl} is the improper angle being measured, χ_0 is the equilibrium angle and χ_{ijkl} is the angle being measured.

1.2.1.2 Non-bonded energy terms

The non-bonded terms include repulsion energy and van der Waals interactions. For two atoms to be considered non-bonded, they must be more than two atoms apart (they cannot both be bonded to a mutual atom), or in a completely separate molecule. 1, 4 bonded atoms, connected through a dihedral, have to be considered separately, as they have both bonded and non-bonded characteristics.

The van der Waals energy is repulsive at short interatomic distances, as when they overlap the electron clouds repel each other. These can be modelled using a Lennard-Jones Potential, as shown in figure 1.7 on the following page, using the equation

$$E_{vdW} = \sum_{i<j}^{atoms} \epsilon \left[\frac{\sigma_{ij}}{R_{ij}^{12}} - \frac{\sigma_{ij}}{R_{ij}^{6}} \right] \tag{1.6}$$

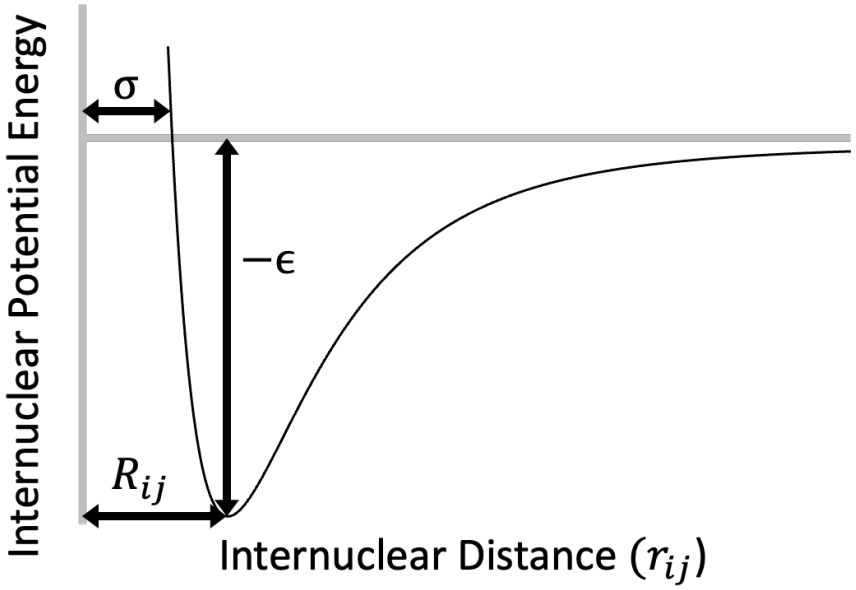


FIGURE 1.7: Lennard-Jones Potential. The bond length affects the van der Waals energy. The scale of the r-axis in this figure will be smaller than the scale in figure 1.2.

where ϵ and σ are coefficients corresponding to the two atoms concerned. On the graph in figure 1.7, σ is a distance equal to twice the van der Waals radius of the atoms, ϵ is the energy well depth, and R_{min} is the equilibrium length.

Although close up, particles will repel each other, when it comes to the longer range they are attracted to each other, as nuclei are charged particles. This means that electrostatic interactions also need to be taken into account. These are modelled using Coulomb's law:

$$E_{electrostatic} = \sum_{i<j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}} \tag{1.7}$$

where q_i and q_j are the partial charges of atoms i and j, R_{ij} is the distance between them, and ϵ is the dielectric constant, also called the relative permittivity, an intrinsic property of the medium.

This electrostatic energy depends on knowing the partial charges. These are not as easy to determine and need to be defined depending on the exact environment of the atom. Figure 1.8 on the next page shows the importance of identifying the correct environment. It highlights two of the carbons, at position 3 and position 6, showing that the partial charge is affected by the number of R groups attached.

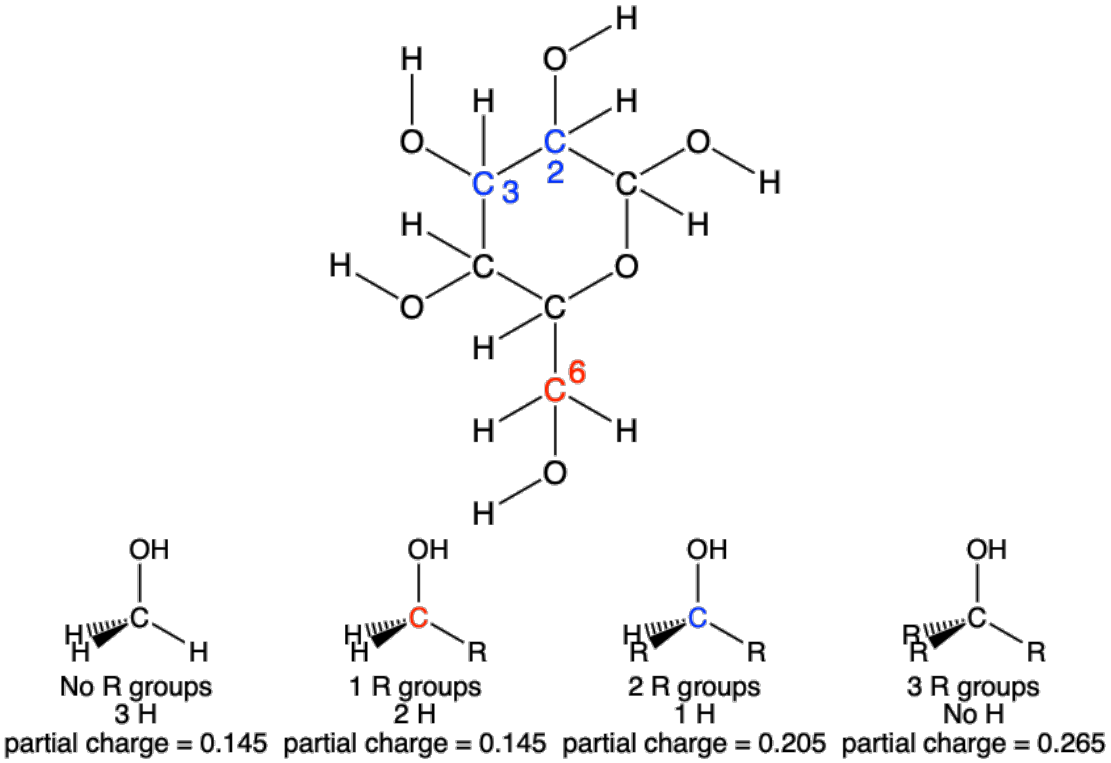


FIGURE 1.8: This shows how the partial charge for an alcohol group can vary depending on its surroundings. In the structure, there are 3 alcohol groups. C_2 and C_3 are both in positions where they are connected to 2 R groups and one H, meaning they will have a partial charge of 0.205, whereas C_2 has two H and only one R, meaning it will have a partial charge of 0.145. The partial charge values are taken from the OPLS-AA force field.[9]

1.2.2 Forcefield parameterisation

A force field is a list of all of the parameters required for the above calculations. These are usually found either through experimental data or quantum mechanical calculations. It also will specify exactly how each of these terms should be calculated.

The development of force fields is one of the main areas of progress in the field of molecular dynamics. It seems there are two main branches down which development tends to lead.[10]

1. Making a force field as generic as possible. The aim of this is to make a force field that can cover almost any structure, all combinations of all elements of the periodic table. Whilst these are useful for getting 'reasonable approximations' of anything, sometimes more accuracy is needed.

2. Ability to give accuracy in the calculation of specific properties, such as molecular structure, conformational properties and heats of formation. force fields developed in this area can provide results within the same degree of accuracy as experimental error.

One force field previously used for similar molecules is the OPLS (**O**ptimized **P**otentials for **L**iquid **S**imulations) force field (examples of this being used will be discussed later in section 2.4).[11,12] The energies are calculated as follows:[13]

$$\begin{aligned} E(r^N) = &\sum_{bonds} K_r \left(r - r_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 \\ &+ \sum_{dihedrals} \left(\frac{V_1}{2} \left[1 + cos\left(\phi - \phi_1\right)\right] + \frac{V_2}{2} \left[1 - cos\left(2\phi - \phi_2\right)\right] \right) \\ &+ \sum_{dihedrals} \left(\frac{V_3}{2} \left[1 + cos\left(3\phi - \phi_3\right)\right] + \frac{V_4}{2} \left[1 - cos\left(4\phi - \phi_4\right)\right] \right) \\ &+ \sum_{i>j} f_{ij} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j e^2}{4\pi\epsilon_0 R_{ij}} \right) \end{aligned} \tag{1.8}$$

Where the non-bonded terms are $A_{ij} = \sqrt{A_{ii}A_{jj}}$ and $B_{ij} = \sqrt{B_{ii}B_{jj}}$ and a "scale factor" f_{ij} is used to mean that 1-4 non-bonded interactions are scaled down to half:

$$f_{ij} = \begin{cases} 0.5 & \text{for 1-4 interactions} \\ 1 & \text{for other combinations of } ij \end{cases} \tag{1.9}$$

It can be seen that the calculation for the dihedral angles is being defined by a Fourier series, and the non-bonded terms have been combined into one summation, with this "scale factor".

The original OPLS force field (OPLS-UA) uses a 'partially united-atom' model. The focus in development was very much on the non-bonded parameters, and so it was found acceptable to include non-polar hydrogens with the carbons they were bonded to, to simplify the calculation. This means that a methyl group would be considered as one CH_3 unit in the ball and spring image, as opposed to one carbon and three hydrogens. Hydrogens in polar environments, such as in a hydroxyl group, would still be considered separately.[9]

Whilst this might be an advantage when it comes to calculation times, as computational resources have increased, the need to use this united atom model has decreased, and so a variation of the OPLS force field which is commonly used today is the OPLS all-atom force field, OPLS-AA.[9] Now, all atoms can be calculated separately.

There are also a few 'reactive' force fields that have been specially developed, such as ReaxFF, which include connection-dependent terms.[5] Bond order is calculated empirically through interatomic distances, and electronic interactions are treated implicitly, meaning that reaction chemistry is being calculated without QM calculations required. ReaxFF has been developed for hydrocarbons, and also has capabilities for nitramines, so it has the potential to be used for the study of nitrocellulose.[14,15] A more

thorough review of forcefields that have been used to study nitrocellulose is performed in chapter 2 in section 2.4.1.

1.2.3 Molecular dynamic algorithms

As stated previously, molecular dynamics apply classical mechanics equations to molecules. At every step, the forces on the atoms in a system are calculated, which can then be combined with previously calculated positions and velocities, to calculate new positions and velocities. Using these new values, the atoms are moved one step into new positions, then the process is repeated, and a new set of forces are calculated. Based on Newton's second law ($F = ma$), the trajectory is found using

$$m\frac{d^2 r_i}{dt^2} = F_i\left(r_1, r_2, ..., r_i, ..., r_n\right) \qquad i =\in [1, n] \tag{1.10}$$

for a particle of mass m, with force F_i being applied in r_i direction.

The force is the negative gradient of the potential energy V, which is the sum of the potential energy for all intermolecular and intramolecular interactions.

$$F_i = -\nabla_{r_i} V \tag{1.11}$$

These equations are solved using finite difference methods, which calculate each point after small time steps of Δt, calculating the new velocities, acceleration and force at each point. These calculations use truncated Taylor expansions. We can do this truncation because the time steps are small. This means that the value of each term of the Taylor series decreases exponentially. Therefore, to limit computational expense, the Taylor expansion is truncated, resulting in an error related to the level of truncation $\mathcal{O}(\Delta T^n)$, when n is the lowest power of Δt not being calculated. [16]

1.2.3.1 Verlet algorithm

The first example that is going to be discussed is the Verlet algorithm. This uses the positions, $r(t)$, and the acceleration, $a(t)$, at time t, and the previous step, $t - \Delta t$, to find new positions at $t + \Delta t$. [17]

$$r\left(t + \Delta t\right) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \tag{1.12}$$

$$r\left(t - \Delta t\right) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \frac{1}{6}b(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \tag{1.13}$$

Equations (1.12) and (1.13) both currently contain $v(t)$ velocity terms, but by rearranging for $r(t+\Delta t)$, the velocity and the later terms of the Taylor expansion can be cancelled out, giving:

$$r\left(t+\Delta t\right) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{1.14}$$

with the acceleration being

$$a(t) = -\left(\frac{1}{m}\right)\nabla V(r_1, r_2, ..., r_n) \tag{1.15}$$

using the equations based on $F = ma$ derived in equations (1.10) and (1.11).

The velocity may have been cancelled out here, but it can be necessary to calculate it, as the velocity is required for other physical quantities such as kinetic energy.

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{1.16}$$

Notice here that the truncation error for the velocity is in the order of $\mathcal{O}(\Delta t^2)$, because it is calculated in a non-cumulative way from the positions, as opposed to the position updates which are in the order of $\mathcal{O}(\Delta t^4)$.

With the kinetic energy, the temperature can then also be calculated.

1.2.3.2 Leap-frog algorithm

Other algorithms are about the same as the Verlet algorithm on the balance of efficiency to accuracy.[4] For example, the Leap-Frog algorithm, where the velocity is calculated at $\frac{1}{2}\,\Delta t$ steps. Starting from the initial positions, the velocity is calculated.

$$v\left(t - \frac{\Delta t}{2}\right) = \frac{r(t) - r(t-\Delta t)}{\Delta t} \tag{1.17}$$

$$v\left(t + \frac{\Delta t}{2}\right) = \frac{r(t+\Delta t) - r(t)}{\Delta t} \tag{1.18}$$

These can be added together and rearranged to give

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{\Delta t}{m}F(t) \tag{1.19}$$

From this, the new positions can be calculated as

$$r(t+\Delta t) = r(t) + \Delta t v\left(t + \frac{\Delta t}{2}\right) \tag{1.20}$$

For the calculation of kinetic energy and temperature, the velocity needs to be known at the same time as the positions, so we can also calculate the velocity at time t as

$$v(t) = \frac{v\left(t - \frac{\Delta t}{2}\right) + v\left(\Delta t + \frac{\Delta t}{2}\right)}{2} \tag{1.21}$$

The advantage of the Leap-Frog algorithm over the Verlet is that it calculates the velocities. However, it does need half time-steps to be calculated which is a disadvantage.

1.2.4 Periodic boundary conditions

One way that the simulation time can be reduced is by implementing periodic boundary conditions. When these are correctly used a system can be used to approximate an infinite system. This is illustrated in figure 1.9.

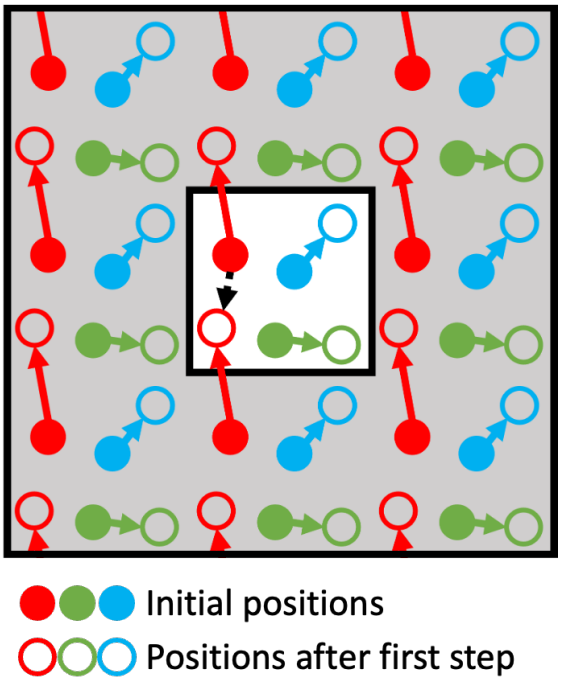


FIGURE 1.9: A two-dimensional image illustrating periodic boundary conditions. The white inner region shows the region involved in the calculation, and the grey outer area represents a layer of the virtual 'infinite' system being approximated. The filled circles are the initial positions of the atoms, and the hollow circles are the positions after a hypothetical step.

The white region of figure 1.9 is the simulation cell. When the red atom moves out of the simulation cell it is reflected in through the periodic boundary conditions. The number of atoms in the system must remain constant - if atoms are being lost there is an issue with the PBCs.[18]

The grey region is also important - based on the distance cutoff a “ghost region” must be considered, as the interactions through the boundary conditions should also be calculated. For example, after the step has taken place, the red atom is far away from the blue atom, so their interaction will not be significant. However, it is much closer to the “ghost” of the blue atom, so this interaction is still significant.

It is also possible to apply periodic boundary conditions to only specific dimensions. This means that PBCs can be used in the study of fibres and microfibrils - an arrangement of chains can be set up so that the transverse section is the desired shape and size. If periodic boundary conditions are applied along the length of the chains a fibre of infinite length can be simulated.[19] The degree of polymerisation (the number of repeat units along the chain) does make a difference to the simulation results, particularly the crystallinity of the structures being created.

1.2.5 Ensembles

Another thing that needs to be chosen before starting a molecular dynamics simulation is the ensemble. Ensembles are collections of possible systems, which are different microscopically but have specified macroscopic or thermodynamic properties fixed.[20]

They are referred to by the properties that they are maintaining, as their fixed properties, and the most common are characterised as follows:[21]

- Canonical ensemble: NVT. As shown in 1.10. The amount in the system, the volume of the system, and the temperature are all fixed, allowing for variation in the energies
- Microcanonical ensemble: NVE. Each system is thermally isolated, meaning the amount, volume and energy of the system stays the same, but the temperature is allowed to change.
- Isobaric-Isothermal ensemble: NPT. The amount, pressure and temperature are all fixed, leaving the volume to be allowed to change.
- Grand canonical ensemble: μVT. Along with the volume and temperature, the chemical potential μ is fixed. This allows for variation in the amount, meaning that matter can pass between the replications.

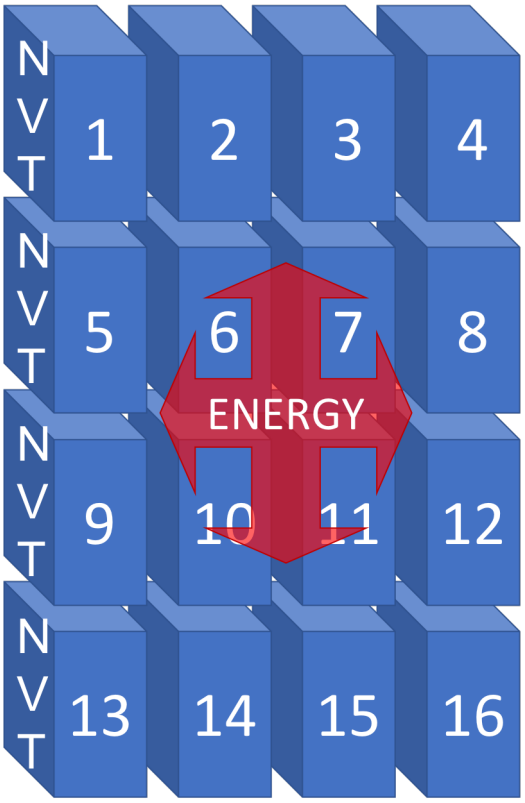


FIGURE 1.10: 16 states in a canonical NVT ensemble. The temperature is constant across all of them, as they are in thermal contact. The volume and amount in each replication are also remaining constant. At the same time, heat energy is being transferred between them, so although the total energy of the system is constant, the energy of each replication varies.

In some cases, these can be specified more precisely - for example in the case of a study by Zhang *et al.* into liquid/liquid interfaces in 1995.[22] In this study, systems were made with layers of two different liquids and vacuums. So within this, they made several custom ensembles including $NVAE$ - where the volume is fixed, but also the surface area, as this would affect the interface between the layers, NP_nAH_{pn} - P_n means the *normal* pressure is fixed, the pressure across the layers, and H_{pn} is the enthalpy calculated using this normal pressure.

Once it is decided what parameters are being controlled, then how they are being controlled needs to be decided.

1.2.6 Thermostats and barostats

It is not too difficult to picture how the volume of a system can be maintained throughout a calculation, but more conceptually difficult to picture how the temperature and pressure are maintained.

A very simple way that this could be achieved is to rescale the velocities at certain intervals, so that the total kinetic energy of the system is correct according to the desired temperature.[23]

One method that uses this is the Berendsen thermostat.[24] It uses a system weakly coupled to a heat bath, and adjusts the temperature, T, so that the deviation from the temperature of the bath, T_0, is reduced exponentially, relative to a time constant τ, using the equation[25]

$$\left(\frac{dT}{dt}\right)_{bath} = \frac{T_0 - T}{\tau} \tag{1.22}$$

This results in a system that is not truly NVT, as the kinetic energy fluctuations are being supressed. However for large enough systems it is a reasonable approximation.

Another popular thermostat, Nosé-Hoover, deals with the temperature by reformatting the equations of motion to make all the thermodynamic variables, except the energy, into constants of motion. These can then be solved, giving results as before, with the thermodynamic properties fixed.

Unfortunately, in cases of systems with slow energy transfer, a deterministic thermostat like the Nosé-Hoover, will struggle. These situations can be solved by using a stochastic thermostat, such as the Andersen or Langevin thermostat.

The Langevin thermostat uses stochastic differential equations. It can cope with larger time steps than the deterministic thermostats mentioned here.

Barostats treat the pressure in the same way as thermostats treat the temperature.

1.2.7 Equilibration protocols

Before long 'production' molecular dynamics simulations are run on a system, it is important to relax its total energy. For this, often a specific routine of short simulations will be undertaken, generally referred to as a protocol.

Two example protocols are going to be discussed here:

1. A 12-Step procedure, based on the work of Larsen *et al.* from 2011, which was used to study the density and crystallinity at equilibrium of porous materials, specifically PIMs (polymers of intrinsic microporosity), and was chosen to look at because it shows how heating and cooling, and the annealing cycles suggested in Chapter 3, in figure 3.16 on page 94[26]

2. A protocol from Bregado, which is for amorphising paracrystalline structures, chosen to discuss as an example protocol that has been used successfully on cellulose, one of the structures of interest to this project, and it could potentially be extended to test on nitrocellulose.[19]

1.2.7.1 12-Step protocol

TABLE 1.1: 12-Step Protocol.[26]

step	conditions	duration / ps
1	NVT, 600 K	50
2	NVT, 300 K	50
3	NPT, 300 K, $0.2P_{max}$ bar	50
4	NVT, 600 K	50
5	NVT, 300 K	100
6	NPT, 300 K, $0.6P_{max}$ bar	50
7	NVT, 600 K	50
8	NVT, 300 K	100
9	NPT, 300 K, P_{max} bar	50
10	NVT, 600 K	50
11	NVT, 300 K	100
12	NPT, 300 K, 1 bar	800

This 12-Step Protocol has a variable P_{max}. There are essentially 4 loops of heating to 600 K and cooling back to 300 K using NVT and then an NPT with different pressure values. It is an abridged version of a 21-Step "slow decompression" protocol also used by Larsen *et al.*.[26]

Several tests were run with different P_{max} to see how the density of the final product was affected by the pressures being applied during the main protocol.

1.2.7.2 Amorphous protocol

TABLE 1.2: Amorphous Protocol.[19]

step	conditions	duration / ps
1	Minimisation	-
2	NVT, 800 K	25
3	Minimisation	-
4	NVT, 750 K	50
5	Minimisation	-
6	NVT, 700 K	50
7	Minimisation	-
8	NPT, 700K, 1 atm	50
9	Minimisation	-
10	NPT, 700 K, 1 atm	24 ns

This protocol was used on microfibril cellulose structures to help them lose their crystalline structure.[19] Every step is at a very high temperature, which also will help to lose the regular structure, and the minimisation between each step is to ensure that the most relaxed, lowest energy-disordered structure is being found.

There are several ways in which molecular dynamics can be analysed to check the progress of the simulation, and for the results to be made comparable to experimental data.

1.2.8 Methods for analysing results

1.2.8.1 Calculation of density

The density of a structure is a property that is similarly straightforward to find for an experimental structure or a simulated structure. This means that if a system aims to recreate a known experimental structure, comparing the density is a reasonable diagnostic technique.

When an NPT calculation is run, the volume is not fixed, which means that the system's size will change, and therefore the density will change. If the densities found in our simulations are similar to those found experimentally then the simulations can be assumed to be representative of natural nitrocellulose.

This study will aim to use molecular dynamics simulations to create large-scale realistic nitrocellulose systems and test how well they might represent natural systems. It also will study how varying the nitration affects the structure.

1.2.8.2 Root-mean-square deviation

Root Mean Square Deviation (RMSD) can be calculated to see the progress of a reaction towards equilibrium.[27]

RMSD is the measure of deviation from a certain structure, usually taken from the original geometry. It is calculated using the following equation:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} \left(r_{i,sim} - r_{i,expt}\right)^2}{N}} \tag{1.23}$$

where N is the number of atoms in the system being studied, $r_{i,sim}$ and $r_{i,expt}$ are the positions of atom i at this step of the simulation and in the geometry for comparison (or expected geometry) respectively.

When RMSD is calculated, it can be used to monitor the progress of the simulation. As a simulation reaches equilibrium, there will be fewer dramatic changes in the geometry, so the RMSD will stabilise.

In this project, the RMSD is used to check if the structures have stopped undergoing major changes. The structures are compared to their initial structure, as there is no experimental data that can be used.

1.2.8.3 Calculation of hydrogen bonds

In the case of cellulose, the number of hydrogen bonds present can give us information about how amorphous a system has become. The hydrogen bonds are often responsible for the interchain interactions, and so can be diagnostic of how strongly crystalline a structure is.[28]

These bonds can be calculated using the *hbonds* plugin in VMD (see Section 1.4.3.1 on page 51), by finding all cases where there are two atoms, a hydrogen donor (D) and a hydrogen acceptor (A), such that the distance D-A is less a specified length (default 3 Å), and the angle D-H-A is less than a specified angle (default 20°).

1.2.8.4 Calculation of dihedral angles

Another factor that can give us information about the amount of disorder in a system is to measure the torsions around the glycosidic bond. The dihedral angles between $O_5 - C_1 - O_4 - C_4$ and $C_1 - O_4 - C_4 - C_5$ (as labelled in figure 1.11 on the following page) can be measured to show how much flexibility there is in the structure.

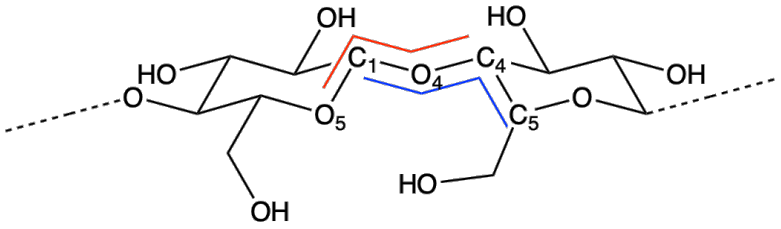


FIGURE 1.11: Cellulose dimer with dihedrals at glycosidic bond.

In a crystalline structure, the dihedral angles will be uniformly the same throughout the structure, and as the system becomes less crystalline there will be a range of dihedral angles.[19]

1.2.8.5 Radial distribution function

The Radial Distribution Function (RDF), is a probability distribution that describes how atoms are arranged relative to each other. It can be described using the following equation:

$$n(r)dr = g(r)4\pi r^2 n_0 dr \tag{1.24}$$

when $n(r)dr$ tells us the average number of atoms within a shell at radius r from a chosen atom with thickness dr, where n_0 is the average density of atoms. This n_0 value is used to scale the results so that a completely random ideal gas would have $g(r) = 1$.

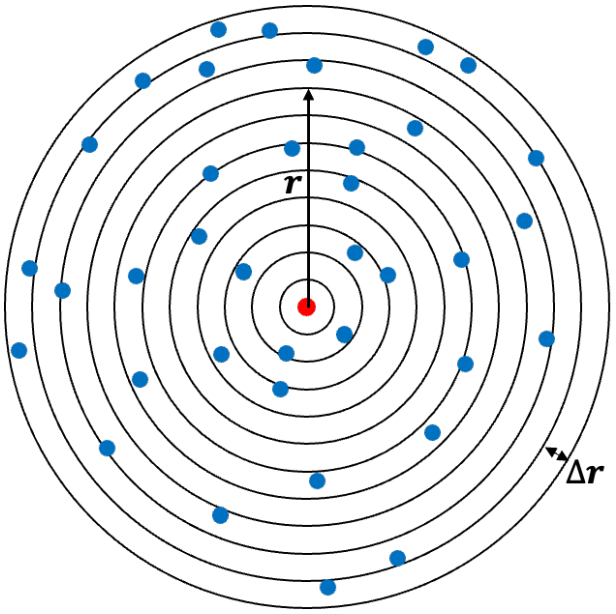


FIGURE 1.12: A two-dimensional example showing how the radial regions are formed, for the counting of the atoms. This is measuring the RDF of the red atom relative to the blue atoms.

For a 3-dimensional system, the RDF can be calculated using the following algorithm:[4]

1. Set values for dr, the range $(0, L)$, and the number of values to be calculated, such that $dr = L/Nbins$.
2. Looping over every value r in the desired range
   (a) Count the number of atoms within the range r to $r + dr$. This means that you are counting all atoms in a shell with width dr.
   (b) Divide by N, the total number of atoms in the system.
   (c) Divide this value by $4\pi r^2 dr$, the volume of the shell. This accounts for the fact that the further out you get, the larger the volume of the shell.
   (d) Divide this value by the number of atoms that should theoretically be in this region if they were evenly distributed over the total space.

It produces a characteristic graph, that can be compared to experimental data, obtained through X-Ray Diffraction. The width and number of peaks can tell us about the regularity of the structures. This can give information about the arrangement of the structure, as discussed further in section 3.5.3 and demonstrated in figure 3.8 on page 84.

1.3 Quantum Mechanical Methods

As discussed earlier, Quantum Mechanical methods, such as DFT, calculate the position of electrons in a system, and so are more useful when studying the making and breaking of bonds. Due to scaling relative to the number of electrons, these calculations tend to be run on smaller systems, but this project will attempt to run these calculations on larger systems.

Before we start describing the development of quantum mechanical theories, there are several principles that need to be defined.

1.3.1 Key definitions

1.3.1.1 Wavefunctions

The *many-body* wavefunction of a system (Ψ) is a mathematical function used to describe the locations of the particles (electrons and nuclei) in the system

$$\Psi\left(x_1, y_1, z_1, x_2, y_2, z_2, ..., x_N, y_N, z_N\right) = \Psi\left(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N\right) \tag{1.25}$$

In the case of a single particle, the wavefunction $\psi(\mathbf{r})$, means we can find the probability of the particle being at point $\mathbf{r}$ by squaring the wave function, $|\psi(\mathbf{r})|^2$. Likewise for a *many-body* system $|\Psi(\mathbf{r}_1, \mathbf{r}_2..., \mathbf{r}_N)|^2$ is the probability of there being simultaneously particle 1 at point $\mathbf{r}_1$, particle 2 at point $\mathbf{r}_2$, up to particle N at point $\mathbf{r}_N$.[29]

1.3.1.2 Born Oppenheimer (BO) approximation

This approximation was mentioned earlier in the *Introduction to Computational Chemistry* section, and states that due to the large size difference between nuclei and electrons (a nucleus is around 1800 times the size of an electron), it is possible to separate the motion of nuclei from the motion of electrons. This means in the case of quantum mechanics, where we are studying the electron density, the nuclei can be thought of as stationary. This simplifies calculations, as it allows us to leave out the kinetic energy of the nuclei and the nucleus-nucleus repulsion terms.

1.3.1.3 The Schrödinger equation

The Schrödinger equation (equation (1.26)) is the most important equation in quantum mechanics

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi \tag{1.26}$$

where $\hat{H}$ is the Hamiltonian operator, i is the imaginary unit, $\hbar$ is the reduced Planck constant ($h/2\pi$), and Ψ is the wavefunction of the system.

A time independent Schrödinger equation can be written as the eigenequation

$$\hat{H}\Psi = E\Psi \tag{1.27}$$

where E is the eigenvalues of the Hamiltonian, and also the total energy of the system. The Hamiltonian operator is separable into several components that can be considered independently.

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{ne} \tag{1.28}$$

where T_e and T_n are kinetic operators for the electrons and the nuclei respectively, and the V terms are potential energies of electron-electron, nuclei-nuclei and nuclei-electron iterations.[30]

1.3.1.4 Hartree products

Assuming that the Hamiltonian for each individual electron can be written as $\hat{h}_i$ and the full electronic Hamiltonian is the sum of these one-electrons Hamiltonians, you can write equation (1.29).

$$\hat{H}_{elec}\left(1,2,...,N_{elec}\right) = \hat{h}\left(1\right) + \hat{h}\left(2\right) + ... + \hat{h}\left(N_{elect}\right) \tag{1.29}$$

where

$$\hat{h}_i = -\frac{1}{2}\nabla^2 + \sum_{I=1}^{N_{elec}} \frac{-Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{1.30}$$

The individual one-electron Hamiltonians can be solved as eigenfunctions to find spin orbitals and energy in the form of the eigenfunctions and eigenvalues respectively.

$$\hat{h}_i \chi_j\left(x_i\right) = \epsilon_j \chi_j\left(x_i\right) \tag{1.31}$$

The Hartree-product is the name given to the simplest possible many-electron wavefunction

$$\psi_{HP} = \left(x_1, x_2, ..., x_N\right) = \chi_i(x_1)\chi_j(x_2)...\chi_k(x_n) \tag{1.32}$$

This is a simplification - it assumes there is no correlation between the motion of electrons, and it also assigns specific electrons to specific spin orbitals, which is not correct, as the particles should be indistinguishable.

1.3.1.5 Variation principle

An expectation value of the energy, $\langle E \rangle$ can be written as

$$\langle E \rangle = \ dfrac \int \psi^*\left(x\right) \hat{H} \psi\left(x\right) dx \int \psi^*\left(x\right) \psi\left(x\right) dx \tag{1.33}$$

where ψ is an approximate wavefunction. When the wavefunction is normalised the denominator is equal to 0, so the equation becomes

$$\langle E \rangle = \int \psi^*\left(x\right) \hat{H} \psi\left(x\right) dx \tag{1.34}$$

This expectation value is going to be greater than or equal to the energy of the system

$$\langle E \rangle \geq E_{exact} \tag{1.35}$$

This means that the wavefunction with the lowest expectation value is the best wavefunction to represent the state of a system, as it is the closest to the exact energy.

1.3.1.6 Molecular orbitals

Molecular orbitals are written using two components - a spatial orbital, describing the location in cartesian coordinates (x, y, z); and a spin function, which describes the spin coordinate which can be either up or down.

Each of these parts is represented by a different Greek letter. $\chi(x)$ is the spin orbital. $\phi(\mathbf{r})$ is the spatial orbital. $\sigma(\omega)$ is the general symbol for the spin function, although as there are only two different possible values for this, often it is just written directly as α or β.[31]

$$\chi(x)_{UP} = \phi(\mathbf{r})\,\alpha(\omega) \qquad \chi(x)_{DOWN} = \phi(\mathbf{r})\,\beta(\omega) \tag{1.36}$$

These molecular orbitals describe single electrons at a time. When molecular orbitals are combined they must obey the rule of antisymmetry (derived from relativistic quantum field theory) - which states that if the position of any two fermion (fundamental particles with half spin, such as electrons and protons) are swapped, the wavefunction must change sign. Equation 1.37 shows antisymmetry being obeyed when electron i and electron j are swapped.

$$\Psi(\chi_1, ..., \chi_i, ..., \chi_j, ..., \chi_N) = -\Psi(\chi_1, ..., \chi_j, ..., \chi_i, ..., \chi_N) \tag{1.37}$$

The fact that antisymmetry must be obeyed means that the molecular orbitals cannot simply be multiplied together.

1.3.1.7 Pauli exclusion principle

The Pauli Exclusion principle states that no two electrons can have the same combination of quantum numbers. In this case, this means that each spin orbital can only contain a maximum of two electrons, and when there are two in a spin orbital they must be in opposite spin states.

1.3.2 Using the Schrödinger equation

Looking back to equation (1.28), it becomes more apparent why the equation gets more difficult to solve - the larger a system is, the more interactions there are that are going to need calculating.

The kinetic energy for a system of N electrons and M atoms can be expressed as

$$\text{kinetic energy} = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_{I=1}^{M} \frac{\hbar^2}{2M_I}\nabla_I^2 \tag{1.38}$$

and the Coulomb potentials are calculated using Coulomb's Law as follows[29]

$$(\text{potential energy})_{ee} = \sum_{i<j} \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.39a}$$

$$(\text{potential energy})_{nn} = \sum_{I<J} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_j|} \tag{1.39b}$$

$$(\text{potential energy})_{en} = -\sum_{i,I} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{1.39c}$$

Using these formulae, equation (1.28) can be expanded into

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + \sum_{i<j} \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} + \sum_{i,I} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{1.40}$$

where i and j are indexes for counting electrons, I and J are indexes for counting the nuclei, m_e and e are the mass and charge of an electron, M and Z are the mass and charge of nuclei and $\mathbf{r}$ and $\mathbf{R}$ are positions vectors of the electron and nuclei respectively. The Laplacian operator is a second derivative which in Cartesian coordinates is written as:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{1.41}$$

The number of terms in this equation increases as the number of electrons and nuclei increases - for a hydrogen atom there is one electron and one nucleus, and so there are three terms.

$$\hat{H} = -\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}} - \frac{\hbar^2}{2M}\nabla_{\mathbf{R}} - \frac{e^2}{|\mathbf{r} - \mathbf{R}|} \tag{1.42}$$

Adding an electron to the system, to look at helium, two new terms are added, the interaction between the second electron and the nucleus, and a repulsion term between the two electrons. A third electron will add 3 terms (electron 3/nucleus, electron 3/electron 1 and electron 3/electron 2). This shows clearly how the size of this calculation can increase exponentially.

The first way that we can simplify the system is to apply the Born Oppenheimer approximation. This means that the kinetic energy term of the nuclei becomes zero, and the internuclear interactions become constant, reducing the number of interactions to be considered. However, this is not enough on its own to make the problem easily soluble for systems of any significant size.

1.3.2.1 Slater determinant

Consider a two-electron system with two electrons x_1 and x_2 with orbitals χ_i and χ_j. Either electron 1 is in χ_i and electron 2 is in χ_j, or electron 2 in in χ_i and electron 1 is in χ_j.

$$\begin{aligned} \phi_{HP}(x_1, x_2) &= \chi_i(x_1)\chi_j(x_2) \\ \phi_{HP}(x_1, x_2) &= \chi_i(x_2)\chi_j(x_1) \end{aligned} \tag{1.43}$$

For antisymmetry to apply, if these are combined linearly, a negative sign must be involved. For an N electron system, there is a normalisation factor of $N!^{-\frac{1}{2}}$, meaning for a two-electron system this is $2^{-\frac{1}{2}}$.

$$\phi_{HP}(x_1, x_2) = 2^{-\frac{1}{2}}\Big(\chi_i(x_1)\chi_j(x_2) - \chi_i(x_2)\chi_j(x_1)\Big) \tag{1.44}$$

This is known as the Slater Determinant (SD) and can be extended to N electron systems as shown in equation (1.45).

$$\phi_{SD}(x_1, x_2, ..., x_N) = (N!)^{-\frac{1}{2}} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \dots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \dots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \dots & \chi_k(x_N) \end{vmatrix} \tag{1.45}$$

1.3.2.2 Slater determinant energy

The energy of a Slater Determinant can be found using the variation principle (equation (1.35) on page 24). It is established that the total energy is the sum of the kinetic energy and the interaction energies.

$$E_{SD} = E_{kin} + E_{ee} + E_{eN} \tag{1.46}$$

The terms for the kinetic energy and the nucleus-electron interactions can be summed together for each electron in an N electron system, and with the required equations in Chapter 1, on page 25 - the kinetic energy in equation (1.38) and nucleus-electron interactions are found using equation (1.39), becoming

$$E_{eN} = -\sum_{i,I} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{1.47}$$

This, and equation (1.35) means that equation (1.46) can be rewritten as

$$E_{SD} = \sum_{i=1}^{N_{elec}} \chi_i^*(x)\chi_i(x)dx + E_{e-e} \tag{1.48}$$

When the electron-electron interactions begin to be considered, the calculations become even more complicated. To help find this the spin orbitals need to be integrated out, using orthonormality, so that it is in terms of spatial orbitals. To do this it is assumed that they are orthonormal. This means that when two spin orbitals have the same spin function, the integral of the spin functions is 1, and if they have different spins, the integral is 0.

$$\begin{gathered} \int \sigma_q^*(\omega_1)\sigma_t(\omega_1) = \delta_{qt} \\ \delta_{qt} = \begin{cases} 1 & q = t \\ 0 & q \neq t \end{cases} \end{gathered} \tag{1.49}$$

We then reach the following equation (equation (1.50)), where J_{ij} is the electrostatic interactions between one electron in orbital ψ_i and every other electron in orbital ψ_j (equation (1.51)), and K_{ij} is an exchange interaction between the electrons of orbital ψ_j that have the same spin (equation (1.52)).

$$E_{SD} = 2\sum_{i=1}^{\frac{N_{elec}}{2}} h_i + \sum_{i=1}^{\frac{N_{elec}}{2}} \sum_{j=1}^{\frac{N_{elec}}{2}} \left(2J_{ij} - K_{ij}\right) \tag{1.50}$$

$$J_{ij} = \int \int \frac{|\phi_i(r_1)|^2}{r_{12}} dr_1 dr_2 \tag{1.51}$$

$$K_{ij} = \int \int \frac{[\phi_i^*(r_1)\phi_j(r_2)][\phi_i^*(r_2)\phi_j(r_1)]}{r_{12}} dr_1 dr_2 \tag{1.52}$$

This halves the number of exchange integrals to be calculated, as they are only occurring where the electron spins are the same.

1.3.3 Hartree-Fock theory

Hartree-Fock theory (HF) theory uses wavefunction theory to solve the electronic Hamiltonian equation (equation (1.26)).

HF uses an approximation known as “mean-field approximation”, or “independent electron approximation”, where each electron is considered individually as being subjected to a mean field of all of the other electrons. This is a simplification, as the actual positions of the surrounding electrons are no longer relevant to the calculation, just the average overall density. The energy difference between the exact energy and the energy being calculated using HF theory, caused by this approximation, is known as the correlation energy.

$$E_{correlation} = E_{exact} - E_{HF} \tag{1.53}$$

Overall HF theory can be written in the following equation

$$E_{elec}^{Hartree-Fock} = E_{kin} + E_{e-N} + E_{coul} + E_{N-N} + E_X \tag{1.54}$$

Where E_{kin} is the electron kinetic energy, E_{e-N} is the electron-nuclei interactions, E_{coul} is electron-electron interactions, E_{N-N} is nuclear repulsion terms and E_X is the exchange energy, the interaction between the electrons and the mean field of other electrons. The correlation energy of electrons which comes from the actual positions is not accounted for here.

The Hartree-Fock energy can only be solved iteratively, as the equation to find a system’s orbitals depends on knowing the orbitals. Therefore it is solved using a Self-Consistent Field, SCF, procedure, as shown in figure 1.13 on the facing page.[32]

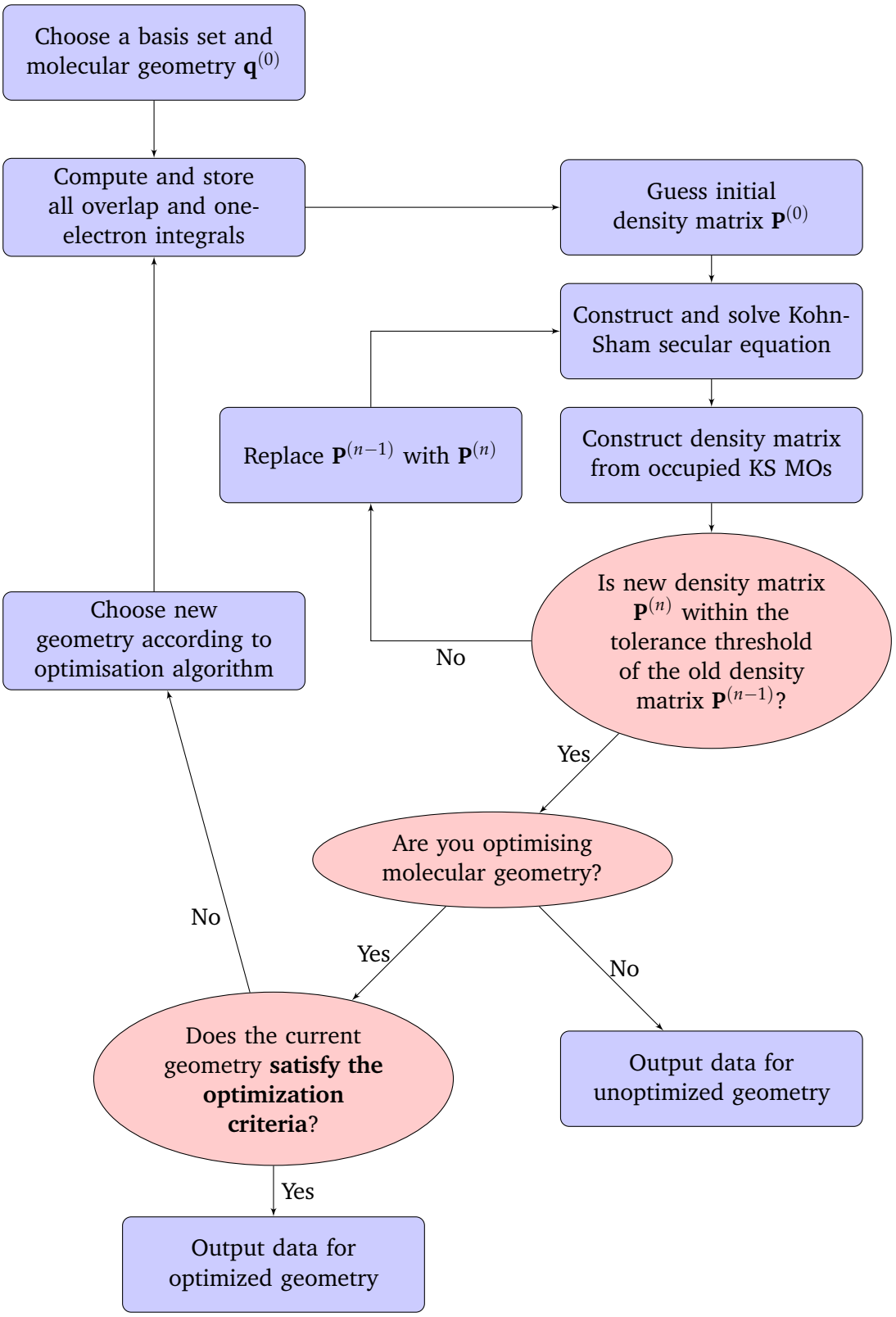


FIGURE 1.13: Flow chart demonstrating Kohn Sham SCF procedure.[27]

1.3.4 Restricted and unrestricted calculations

There are three types of Hartree-Fock calculations. Restricted Hartree-Fock (RHF) can only be applied to systems with an even number of electrons, which must be paired up (closed-shell) - each spatial orbital contains two electrons, in opposite spin states. These are shown in figure 1.14.

Restricted Open-shell Hartree-Fock (ROHF) calculates for each electron, as opposed to pairs. This means it can be used for systems containing unpaired electrons.

Unrestricted Hartree-Fock (UHF) calculations consider the spin orbitals separately, so can be used on systems containing unpaired electrons. Doing this can make a difference because a 1sα electron is going to be affected by other possible α electrons, differently from how a 1sβ electron would be. The fact that the spin affects the energy leads to spin contamination, which could have an important effect on the calculations.

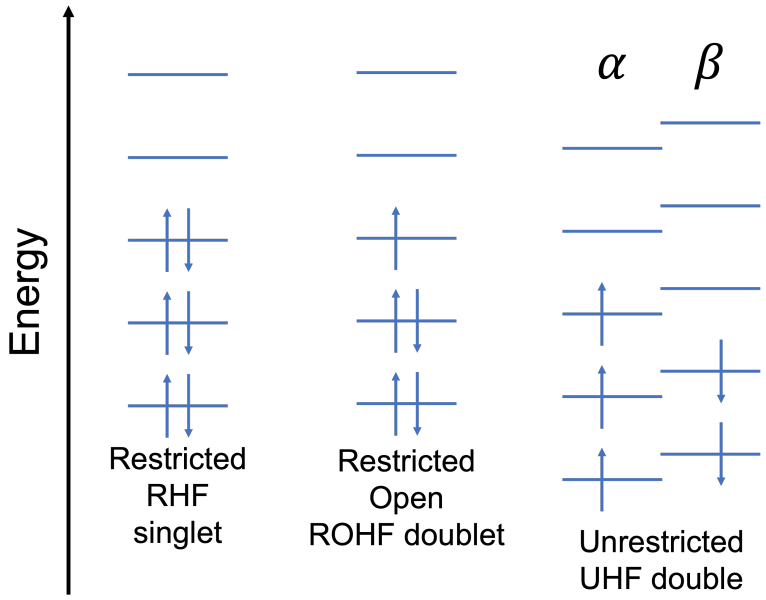


FIGURE 1.14: Electron configurations for restricted open and closed, and unrestricted calculations.

1.3.5 Density functional theory

Density Functional Theory (DFT) is one of the more common quantum mechanics approaches. Unlike HF theory, which calculates energy based on wavefunctions, DFT uses the density of the electrons and increases the accuracy of the calculations by including an approximation for electron correlation.

The first suggestion that density could be used to calculate the energy of systems came from Thomas and Fermi in 1927. They used the concept of the Uniform Electron Gas

(UEG) model (also known as the Jellium model), where electrons are considered as point negative charges with a uniform background positive charge, to give the following equation

$$E_{TF}[n] = C_1 \int d^3r\, n(\mathbf{r})^{\frac{5}{3}} + \int d^3r\, V_{ext}(\mathbf{r}) n(\mathbf{r}) + \frac{1}{2} \int \int d^3r\, d^3r' \frac{n(\mathbf{r}) n(\mathbf{r'})}{|\mathbf{r} - \mathbf{r'}|} \tag{1.55}$$

where $C_1 = \frac{3}{10} \left(3\pi^2\right)^{\frac{2}{3}}$. The first term here is the kinetic energy of the non-interacting electrons, and the second and third terms show the electron-electron repulsion, and nucleus-electron interactions being calculated classically. However, this model does not include exchange or correlation energies.

Dirac improved a little upon the Thomas-Fermi equation (equation (1.55)), by adding a term for the exchange energy

$$E_{TFD}[\rho] = E_{TF}[n] + C_2 \int d^3r\, n(\mathbf{r})^{\frac{4}{3}} \tag{1.56}$$

where $C_2 = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{\frac{1}{3}}$. Correlation energy is still not accounted for here.[33]

Hohenberg and Kohn proved that there was a direct relationship between the density and the Hamiltonian. This means that when we have the density defining the number of electrons, the properties of the system can be found.

As before, the energy is the sum of the kinetic energy terms $T_e[\rho]$ and the interactions $U_{ee}[\rho]$.

$$E[\rho] = T_e[\rho] + U_{ee}[\rho] + V_e xt[\rho] \tag{1.57}$$

The terms that involve the nuclei are referred to as external energies, as after the BO approximation is applied, the calculations are only on the electrons, making the nuclear interaction “external”.

The external terms of equation (1.57) can be grouped together into what is known as the universal functional $F_{HK}[\rho]$.

$$E[\rho] = F_{HK}[\rho] + \int \hat{V}_{ext}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} \tag{1.58}$$

This universal function has been proven to exist, however, no expression has yet been defined for it. It contains the following energy terms

$$F_{HK}[\rho] = E_{kin}[\rho] + E_{coul}[\rho] + E_X[\rho] + E_C[\rho] \tag{1.59}$$

E_{kin} and E_X do have known equations, just not in terms of the density. E_{coul} does have an exact expression in terms of density that can be used, but the other energies must be approximated.[34]

1.3.6 Kohn Sham DFT

The next people to make a major breakthrough in the development of these methods were Kohn and Sham.[35] They were searching for a method to eliminate the need for the universal function. They used a combination of density and molecular orbitals to consider a system of non-interacting electrons where the density is the same as the system would be if they were interacting. This now means that the kinetic energy can be calculated exactly, although it does bring the number of variables needing to be calculated from 3 to 3N.

The energy contributions can now be written again as

$$E_{KS}[\rho] = T_s[\rho] + \int \left[\hat{V}_{ext}(\mathbf{r}) + \hat{U}_{el}(\mathbf{r})\right] \rho(\mathbf{r}) d\mathbf{r} + E_{XC}[\rho] \tag{1.60}$$

when $T_s[\rho]$ is the kinetic energy of the theoretical system of non-interacting electrons, $\hat{V}_{ext}$ is external potential and $\hat{U}_{el}$ is the interaction of the electrons with nuclei and E_{XC} is the exchange-correlation energy. If J is used to represent the Coloumbic interactions, the equation can be further simplified to give us

$$E_{XC}[\rho] = (T_e[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) \tag{1.61}$$

Where the first set of brackets is the kinetic energy, including a correction term, as non-interacting electrons will move slower than interacting electrons. The other set of brackets includes the coulombic interactions, exchange and correlation.

A new Schrödinger equation can be written, after three operators are defined - $\hat{V}_{ext}$, the external potential; $\hat{U}_{el}$, the interactions between the electrons and nuclei; and $\hat{V}_{XC}$, which is the functional derivative of the energy with respect to density.

$$\hat{V}_{ext} = \sum_a \frac{-Z_a}{|R_a - r|} \tag{1.62}$$

$$\hat{U}_{el}(r) = \int \frac{\rho(r')}{|r' - r|} dr' \tag{1.63}$$

$$\hat{V}_{XC}(r) = \frac{\delta E_{XC}[\rho(r)]}{\delta \rho(r)} \tag{1.64}$$

The Schrödinger equation becomes

$$\left[-\frac{1}{2}\nabla_i^2 + \hat{V}_{eff}(\mathbf{r})\right] \Phi_i^K(\mathbf{r}) = \epsilon_i \Phi_i i(\mathbf{r})^{KS} \tag{1.65}$$

where $-\frac{1}{2}\nabla_i^2 + \hat{V}_{eff}(\mathbf{r})$ is the Kohn-Sham Hamiltonian, $\hat{h}_{KS}$.

Kohn Sham DFT improves on previous methods, by considering a fictitious system of non-interacting electrons.[34] This is not a perfect solution, it still requires correction terms.

1.3.7 Exchange and correlation functionals

Most of the elements in the equation (1.60) can be solved exactly, except for E_{XC}, which contains the exchange energy, correlation energy, and the kinetic energy from KS DFT that describes the difference between the real and fictitious system.[36]

Several methods have been found to account for this, and they are often described as levels of a Jacob's Ladder, illustrated in figure 1.15, increasing with accuracy the higher up the ladder they are found.[37]

They all are based on the basic form shown in equation (1.66).

$$E_{XC} = \int \rho(\mathbf{r}) \epsilon_{XC}[\rho(\mathbf{r})] d\mathbf{r} \tag{1.66}$$

when $\epsilon_{XC}[\rho](\mathbf{r})$ is exchange-correlation energy for each particle of a UEG of density, and $\rho(\mathbf{r})$ is the number of electrons per unit volume. The exchange and correlation energies can be separated, giving

$$E_{XC} = E_X + E_C = \int \rho(\mathbf{r}) \epsilon_x[\rho(\mathbf{r})] d\mathbf{r} + \int \rho(\mathbf{r}) \epsilon_c[\rho(\mathbf{r})] d\mathbf{r} \tag{1.67}$$

The "hell" of the Jacob's Ladder (figure 1.15) is the Hartree world, and so on the lowest "rung", furthest from the "heaven" of chemical accuracy, is the Local Density Approximation (LDA). It is based simply on the UEG, assuming a constant electron density over a system, which is not ideal, and why LDA is on the lowest rung.[38]

$$E_{XC}^{LDA}[\rho] = E_X^{LDA}[\rho] + E_C^{LDA}[\rho] \tag{1.68}$$

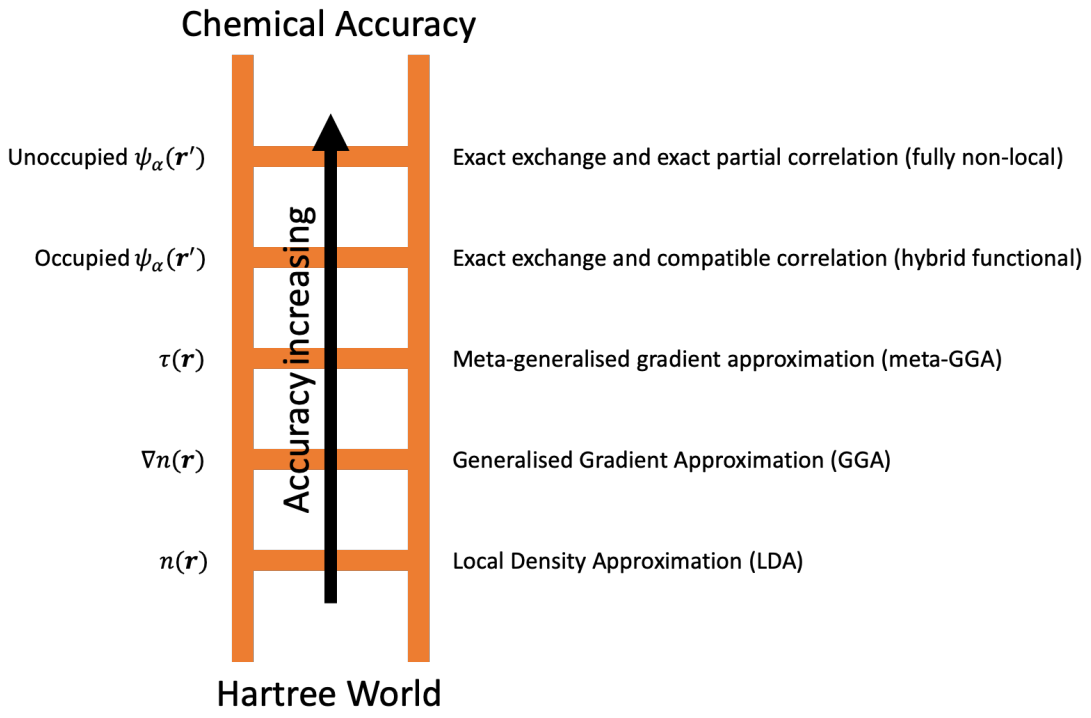


FIGURE 1.15: Jacob's Ladder of exchange correlation functionals.[37]

The exchange energy can be derived using UEG:

$$E_X^{LDA}[\rho] = -C_x \int \rho^{\frac{4}{3}}(\mathbf{r}) d\mathbf{r} \tag{1.69}$$

Different LDA methods use different values for the correlation energy, found by comparing the results of UEG with advanced quantum mechanics methods.

The next step up is the Generalised Gradient Approximation (GGA).[39] This allows for a gradient term, meaning the electron gas is no longer uniform. Most GGAs are found in the form of the LDA functional with a correction term:

$$E_{XC}^{GGA}[\rho] = E_{XC}^{LDA}[\rho] + \Delta E_{XC}\left[\frac{|\nabla\rho(\mathbf{r})|}{\rho^{\frac{4}{3}}(\mathbf{r})}\right] \tag{1.70}$$

The parameters of GGAs are often found by fitting to experimental results, or results of more accurate calculations, like LYP, although others have no empirically derived results, like PBE.

Adding a Laplacian of the density $\nabla^2\rho(\mathbf{r})$ takes the method up another step to being a meta-GGA (mGGA), although adding kinetic energy density, τ, is more common, as it is more numerically stable to calculate.

The final stage is hybrid functionals. These use fractions of HF exchange energy:

$$E_{XC} = (1-a)E_{XC}^{DFT} + aE_X^{HF} \tag{1.71}$$

For example, one of the best-known and most-used hybrid functionals is B3LYP, which is expressed as

$$E_{XC}^{B3LYP} = (1-a)E_X^{LDA} + aE_{XHF} + b\Delta E_x^B + (1-c)E_C^{LDA} + cE_C^{LYP} \tag{1.72}$$

When $a = 0.1161$, $b = 0.9262$ and $c = 0.8133$.

The choice of which functional is important - literature should be used to find which functionals are most suitable for the system being experimented on, and ideally, calibration checks should be run, and to find which is the most suitable. New functionalities are constantly being developed.[40,41] For this project, several different functionals will be compared to see which provide the best results for the sized systems we are using, and the computational programs (for example ONETEP, described later in section 1.4.1.2 has only recently had hybrid functionals implemented).

1.3.8 Basis sets

In the textbook "Essentials of Computational Chemistry, Theories and Models", Cramer says "Atoms and molecules are the basis set in which chemical events are expressed".[27] It is understood that atoms and molecules are the foundation on which calculations of the chemical world are based, and similarly, basis sets contain the building blocks for chemical computations.

Basis sets are linear combinations of basis functions, where basis functions tend to be atomic orbitals. A basis set ψ_i is described as

$$\psi_i(\mathbf{r}) = \sum_{\alpha=1}^{N_{BF}} G_\alpha(\mathbf{r}) C_{\alpha i} \tag{1.73}$$

where N_{BF} is the number of basis functions being used, $G_\alpha(\mathbf{r})$ is a known basis function, and $C_{\alpha i}$ is a coefficient to be determined by calculations.

There are two key places where looking at the behaviour of a function being chosen, to see how close to reality it is, is important - near the nucleus, where the function should ideally reach a cusp; and far from the nucleus, where there should be decay at rate $e^{-a\mathbf{r}}$.

Slater functions (equation (1.74)) are used to make Slater type orbitals (STOs) and Gaussian functions (equation (1.75)) are used to form Gaussian-type orbitals (GTOs).

Of these two only the slater function fits these two qualities, as it resembles the shape of a hydrogen 1s orbital, but Gaussians are considerably easier to solve. This means in the computational quest to balance accuracy and computational cost, GTOs are more frequently used.

$$\text{STO}: \quad s_v(\mathbf{r}) = e^{-\zeta_v|\mathbf{r}|} = e^{-\zeta_v\sqrt{(x^2+y^2+z^2)}} \tag{1.74}$$

$$\text{GTO}: \quad g_v(\mathbf{r}) = e^{-\zeta_v\mathbf{r}^2} = e^{-\zeta_v\left(x^2+y^2+z^2\right)} \tag{1.75}$$

Often multiple primitive Gaussian functions are added together making contracted Gaussians (CGs)

$$\text{CG}: \quad G_\alpha(\mathbf{r}) = \sum_{v=1}^{N_\alpha} c_v g_v(\mathbf{r}) \tag{1.76}$$

The simplest form of CGs are “single zeta” basis sets, STO-nG, where n is the number of primitive Gaussians being included, making a closer approximation of an STO. The STO-nG basis functions for a 1s orbital of hydrogen are shown in figure 1.16.

$STO-1G:$	$e^{-0.27\mathbf{r}^2}$
$STO-2G:$	$0.68e^{-0.15\mathbf{r}^2} + 0.43e^{-0.86\mathbf{r}^2}$
$STO-3G:$	$0.44e^{-0.11\mathbf{r}^2} + 0.53e^{-0.41\mathbf{r}^2} + 0.15e^{-2.23\mathbf{r}^2}$

FIGURE 1.16: STO-nG functions for the 1s orbital of a hydrogen atom.

STO-nG basis sets are a single CG. To increase accuracy, multiple CGs per orbital can be used, “double zeta” (DZ) basis sets use two CGs, “triple zeta” three, and so on.

It is also possible to use a different number of CGs for different orbitals, normally one CG for “core” atomic orbitals, and more for the valence orbitals. This is done because the core electrons are closer to the nucleus, and will be drawn in, less able to get involved in interactions, meaning they are not worth describing in as much detail. These “split valence” basis sets are named in the form

$$C - V_1V_2...V_nG$$

where C is the number of primitive Gaussians in the core CG, and V_n is the number of primitive Gaussians in the nth valence shell.

These basis sets can also be improved by things like adding polarisation functions. Adding specifically tailored CGs, such as CGs of higher angular momentum than the valence orbitals, means that chemical bonds can form in any direction, increasing the "flexibility" of atoms. Generally, these are represented on the basis set name as a *, such as in 6-31G*.

Another type of orbital that can be added for increased accuracy is a diffuse function. These are CGs extending further out than the molecular orbitals, improving the accuracy of calculations on species where the extended electronic densities are important, such as anions and molecules forming hydrogen bonds. These are shown in the basis set name as a +, such as a 4-31+G.

A balanced basis set includes polarisation and diffuse functions.

1.3.9 Solvent models

It is one thing to apply DFT in a vacuum, to a completely isolated molecule in the gas phase - with no other molecules around. However, it is another to apply calculations to a system of a molecule with a solvent. Generally, this can be done in one of two ways - implicitly or explicitly.

The most conceptually straightforward method is to run a calculation on a solute molecule surrounded by several solvent molecules. To do this type of explicit calculation, every single interaction between all the solvent molecules surrounding the molecule and the molecule itself must be calculated. This will increase the number of interactions, and as the number of calculations that are needed scales cubically, will greatly increase the computation time. For example, in figure 1.17 there are only five solvent molecules, and already each electron in the ethane would need to be calculated with respect to sixty-seven other electrons in this configuration. There will also be many other configurations, with different orientations of the ethane.

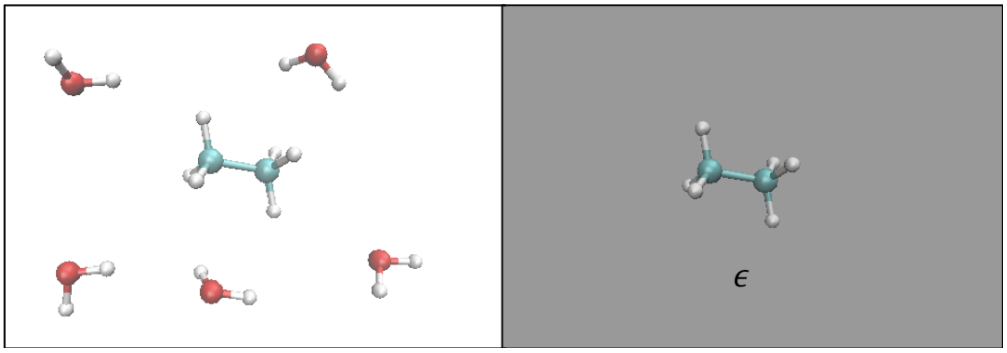


FIGURE 1.17: An explicit system of ethane with water, and an implicit one.

To perform a calculation like this generally a volume of solvent is geometrically optimised, and then the solute molecule is added in. Unfortunately, this often produces

a static phase for the solvent, which is not representative of a real solvent. For example, water has a static permittivity ϵ_0 that is twenty times larger than its actual electronic contribution ϵ_∞ in the case of real-life solvent-solute interactions.[42]

On the other hand, in implicit solvent calculations, the solvent is considered as a bulk "continuum", the properties of which are made to represent the solvent. The idea of using a continuum model was first introduced over a century ago, and so in that time, many different models have evolved.[43]

Implicit solvent methods cannot tell us about how the solute molecules interact with individual solvent molecules. However, if the focus is not on the interactions, but on how things like polarisation affect how the solute molecule behaves, there are many other advantages to this type of calculation. One of the main advantages of a continuum, alongside the reduction of time and cost of the calculations, is the removal of orientational dependence - in an explicit solvent, the molecule can be in different orientations relative to the solvent molecules, each of which should be calculated individually for total accuracy, whereas in a continuum the solvent properties are constant everywhere ruling out the many degrees of freedom.

This is important for knowing the solvation energy, described in equation (1.77).

$$\Delta G^\circ_{sol} = G^\circ_{solution} - G^\circ_{gas} \tag{1.77}$$

All implicit solvent methods start by inserting a solute molecule into a solvent setting. This is done by creating a cavity in the solvent. The interactions are considered to happen at the cavity surfaces, the solvent-accessible surface (SAS).

There are many different methods for including solvents in calculations. For example, COSMO, the COnductor-like Screening MOdel. Most solvation modelling algorithms have the dielectric constant set so that it is equal to 1 at grid points within the space being occupied by the cavity, and equal to the dielectric constant outside.

$$\epsilon(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in C \\ \epsilon & \mathbf{r} \notin C \end{cases} \tag{1.78}$$

In the case of COSMO and other conductor-like models, the dielectric constant is set to be infinity.

$$\epsilon(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in C \\ \infty & \mathbf{r} \notin C \end{cases} \tag{1.79}$$

And a scaling factor is set:

$$f(\epsilon) = \frac{(\epsilon - 1)}{(\epsilon - x)} \tag{1.80}$$

A conductor-like model is more computationally robust than a dielectric model. Initially, there were concerns that COSMO would only work for mediums with high dielectric constants, although this has fortunately now been proven to be unfounded.[44]

1.3.10 Linear scaling

Normal DFT must build NxN matrices, meaning that the calculation times for molecules with N electrons scales at $\mathcal{O}(N^3)$. This generally causes normal DFT calculations to only be worthwhile for molecules of a hundred atoms or less.

This is not particularly necessary. There is a limit to the localisation; an electron is not going to react with another that is hundreds of atoms away. To solve this systematically, several variations have been developed to minimise the number of calculations. It is possible to get theoretically linear ($\mathcal{O}(N)$) calculations, although the pre-factor is often fairly large, meaning that the system must be large before the linear scaling is more efficient.

Work was done by Hierse and Stechel in 1994, which studied heptane and decane.[45] It was suggested that the results of a DFT calculation on a smaller molecule could be used to find reasonable results for a larger molecule.

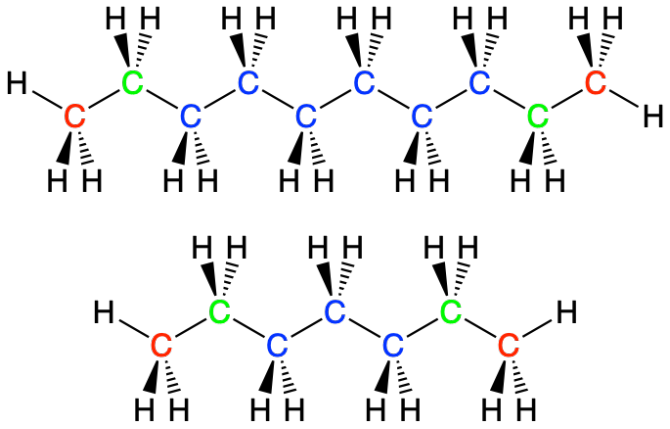
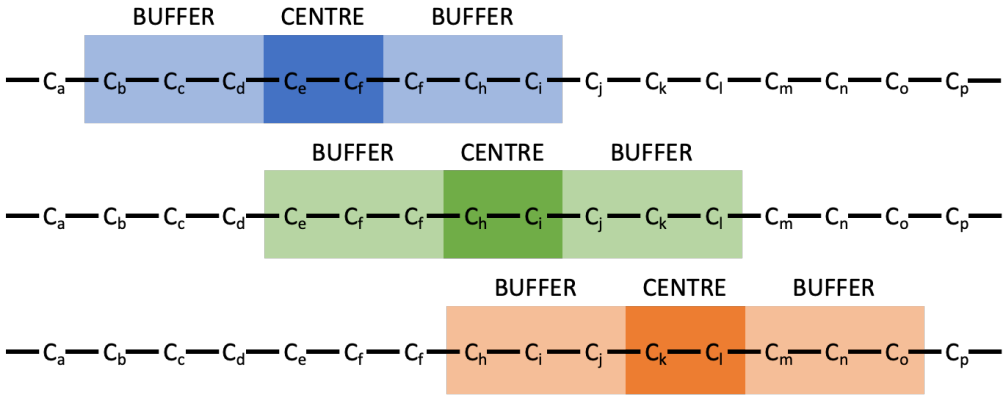


FIGURE 1.18: Molecules of decane and heptane. The colours show the carbons that are in near-identical environments.

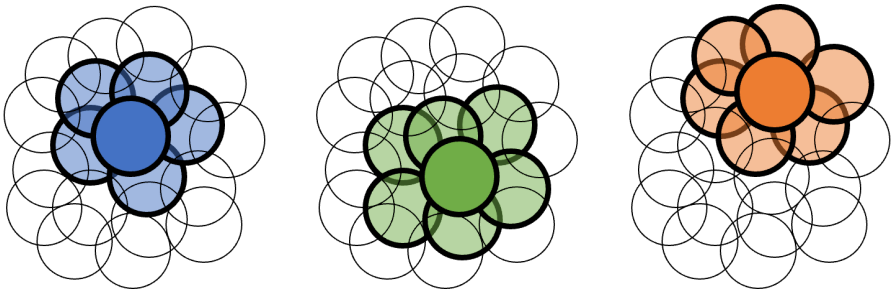
The first idea here is that in a self-repeating carbon chain, in theory, the middle molecules will be in very similar environments, so instead of needing to calculate every single molecule, some of the information can just be copied and pasted. Hierse and Stechel tested this by calculating the bond orbitals, and in decane, only the two outer bonds showed any variation, the 6 inner bonds were all extremely similar, and also

very similar to the inner bonds of the heptane molecule. These observations were then applied to see if they could be used to obtain information for dodecane. They showed that for these specific molecules, using initial guesses of smaller molecules worked fine to calculate larger ones, as the accuracy obtained is similar to that of regular DFT.

This, of course, does not work so well for molecules that do not repeat. For a chain, a divide-and-conquer method can be used.[46,47] This considers sub-volumes for each central unit, as shown in figures 1.19a and 1.19b.



(A) A schematic diagram of how the divide-and-conquer can work on a chain of units. Each C_x is a unit of the chain, but they are not all necessarily the same. This image based on one from Goedecker (1999).[48]



(B) A schematic diagram of how the divide-and-conquer method could be used on a quantum system of overlapping regions. The calculation will be run for all the thickly outlined regions - the centre region (solid filled) and the buffer (surrounding lighter region).[33]

FIGURE 1.19: Two different examples of how divide-and-conquer can be used.

This means that the cost of calculation is in the order of $\left(N_{small} + N_{buffer}\right)^3$ which is better than N_{full}^3.

Another method that is used in several computational is the orbital minimisation approach. This uses Wannier functions, functions that decay exponentially, which improves the localisation in the calculations.[49]

They are combined as a functional

$$\Omega = 2\sum_n \sum_{i,j} c_i^n H'_{i,j} c_j^n - \sum_{n,m} \sum_{i,j} c_i^n H'_{i,j} c_j^m \sum_l c_l^n c_l^m \tag{1.81}$$

with c_i^n as the coefficient for the nth Wannier orbital with respect to the ith basis function and $H'_{i,j}$ is the matrix elements in a shifted Hamiltonian $H - \mu I$.[50,51]

Using linear scaling techniques is extremely useful as it can reduce the timescales, and means that calculations can be run on larger-scale systems with less computational resources, meaning that the intermolecular interactions of systems of large molecular can be studied at a quantum level.

1.3.11 Calculation types

1.3.11.1 Single point energy

The most basic is a single point energy calculation, where the energy of a specific geometry is evaluated, using DFT, and an iterative process to find the probability density of the electrons. This is the data that will be outputted in the flow chart in figure 1.13 on page 29, if the question "Are you optimising molecular geometry?" is answered "No".

1.3.11.2 Geometry optimisation (relaxation)

Energy minimisation is the process of finding the optimal geometry, the geometry with the minimum energy.

If there were only two variables that could be changed, this process can be pictured in three-dimensional space. Figure 1.20 shows a potential energy surface (PES), where two variables on the x- and y-axis can be changed, and the height along the z-axis represents the energy at this point. The minimum of the graph represents the minimum energy.

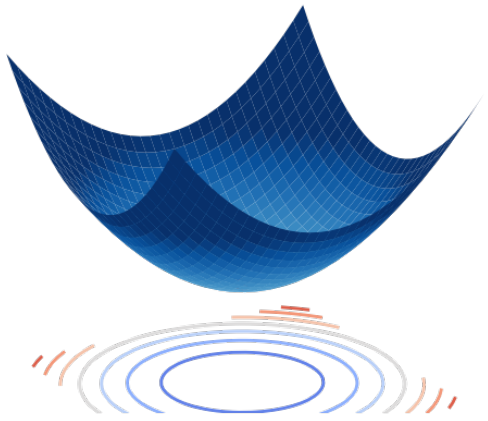


FIGURE 1.20: Minimum on a two-dimensional surface. On a $3N$-dimensional PES the minimum will be the optimal geometry.

This is a very simplified picture - for an N atom system there will be $3N$ variables, and every atom can move in the x, y, and z directions.

One way of computationally finding these minima, as used in the program NWChem (described further in Section 1.4.1.1 on page 49) is a Driver method, a Quasi-Newtonian optimisation which involves line searches and approximate energy Hessian updates. Essentially this method is a steepest descent method, using line searches. This means it finds the direction that has the steepest gradient, and then travels in the opposite direction for as long as possible, until the minimum on that line has been reached, and the gradient begins increasing again.

These optimisation calculations find local minimum points on the PES; however, they are not always the minima that chemically make the most sense for the structure - the output structures should be visualised and checked to ensure that they are chemically correct.

One example of a widely used geometry optimsation algorithm is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.[52,53] The disadvantage of this method usually is that it will take a lot of memory, as it requires a very large matrix. However, there are improvements, such as L-BFGS, shown in figure 1.21 on the facing page.

The improvement in memory usage comes from only storing the last m iterations, $\{s_j, y_j\}_{j=k-m}^{k}$, and using this to calculate the new inverse Hessian H_k as follows

$$\begin{aligned} H_{k+1} = &\left(V_k^T \cdots V_{k-m}^T\right) H_0 \left(V_{k-m} \cdots V_k\right) \\ &+\rho_{k-m} \left(V_k^T \cdots V_{k-m+1}^T\right) s_{k-m} s_{k-m}^T \left(V_{k-m+1} \cdots V_k\right) \\ &+\rho_{k-m+1} \left(V_k^T \cdots V_{k-m+2}^T\right) s_{k-m+1} s_{k-m+1} \left(V_{k-m+2} \cdots V_k\right) \\ &+ \cdots \\ &+\rho_k s_k s_k^T \end{aligned} \tag{1.82}$$

When $\rho_k = 1/(y_k^T s_k)$, $V_k = I - \rho_k y_s s_k^T$, $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, and I is the identity matrix.

1.3.11.3 Minimum energy pathways and transition state searches

As a minimum on the PES represents an optimised geometry, a transition state is found at a saddle point - the lowest point over a barrier that must be crossed. Mathematically a saddle point is where the gradient is negative in all directions except for one, which is positive.

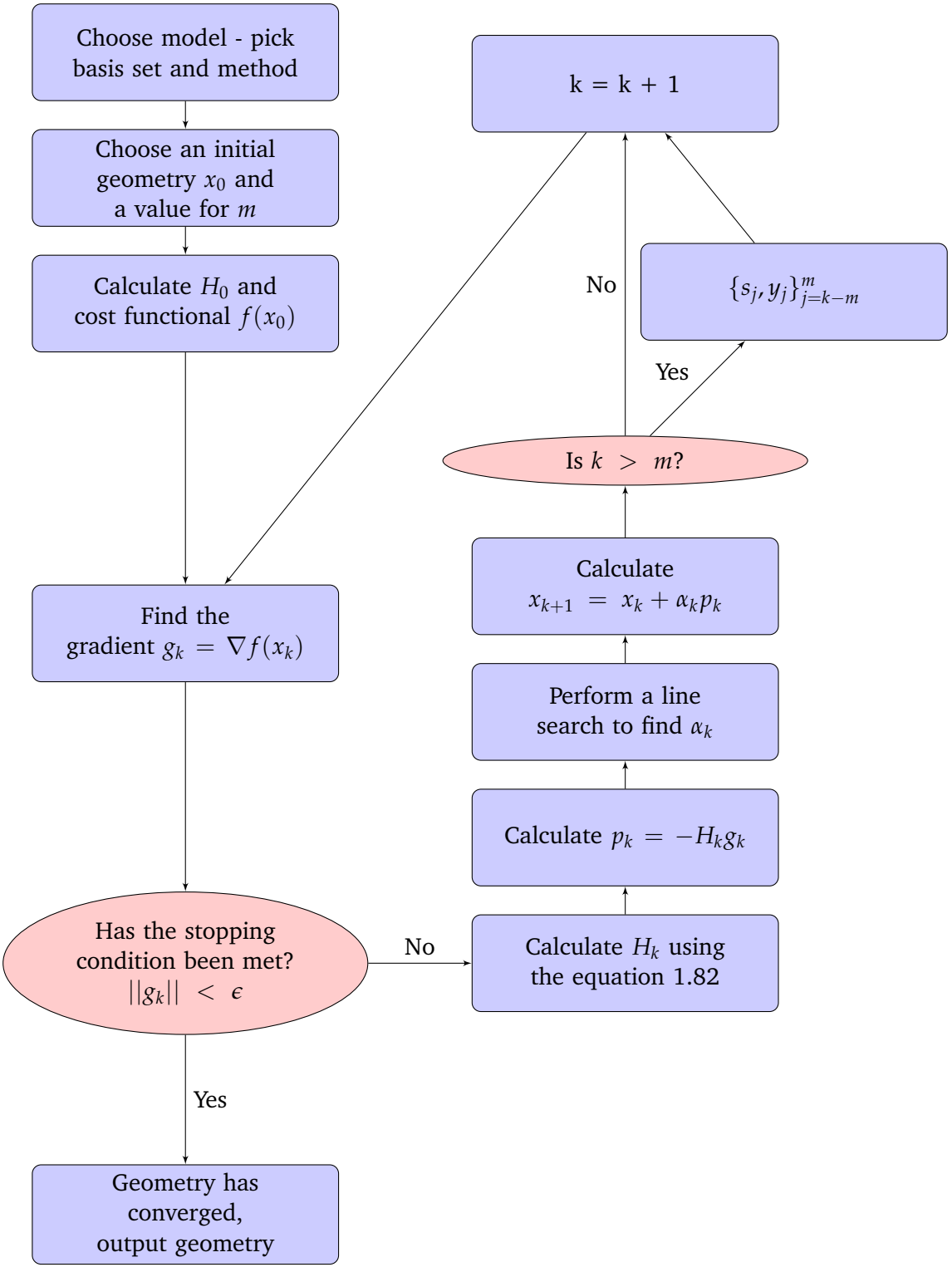


FIGURE 1.21: Flow chart showing L-BFGS method for geometry optimisation.[54]

Finding a transition state can be important for studying reactions, as it can tell us about the activation energy for a reaction, as this is the energy barrier needing to be overcome for a reaction to proceed.

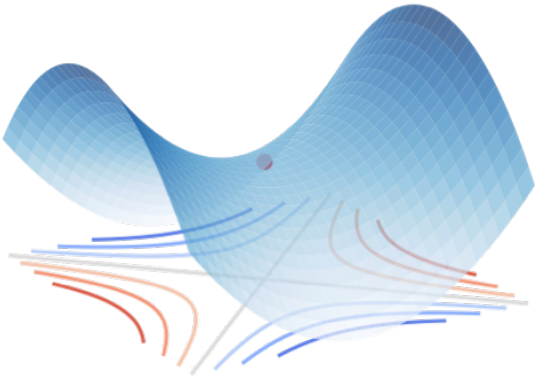


FIGURE 1.22: Saddle point on a two-dimensional surface. On a $3N$-dimensional PES this minimum will be the geometry of a saddle point (although not guaranteed to be the saddle point on the minimum energy pathway (MEP) between the intended reactants and products).

The Driver Algorithm used by programs such as NWChem can be adapted to find a saddle point rather than a minimum. In this project, it is often the case that NEB calculations fail to converge, but the maximum point along the path is then subjected to a driver saddle point search, which increases the chance that the saddle point being found is the saddle point along the MEP.

Of course, this needs to be applied along with visual inspection and good chemical knowledge. Tests can also be run to confirm that a saddle point has indeed been found. In NWChem a frequency test can be run, that calculates all of the vibrational frequencies for the structure. The frequency found is the negative of the gradient, meaning that if the point is correctly a saddle point, then all frequencies should be positive, except for one that is negative, and the 6 degrees of freedom which are being held fixed, so their frequencies will be zero.

The NWChem calculations also output trajectories for each of these vibrations, so the trajectory of the imaginary frequency should show the path over the saddle point, which again should be looked at to check if it makes chemical sense.

1.3.11.4 LST/QST optimisation

LST/QST are another way to look for a saddle point. LST stands for Linear Synchronous Transit, and QST stands for Quadratic Synchronous Transit.

LST begins with a linear line search between the reactants and the products, shown by the red line in figure 1.23 on the next page. The maximum point along this path becomes a first estimate for the transition state.[55]

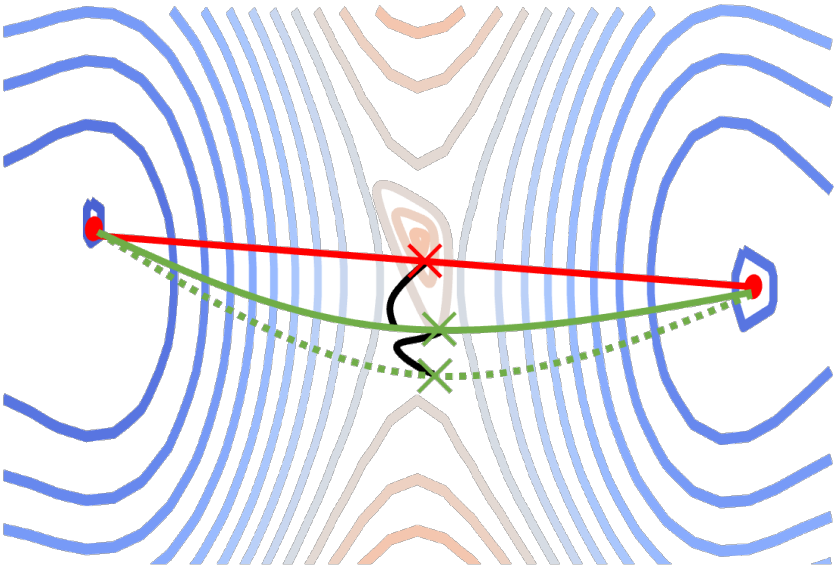


FIGURE 1.23: Sketch diagram of LST/QST. Blue is low energy and red high on the contour map. The linear LST step is shown in red, the optimisation of the maximum points in black, and the QST steps in green (solid and then dotted).

The estimated transition state from LST is then optimised to a local minimum (making sure it is not just going back to the reactants or products), along the black line of figure 1.23.

Next, a new estimate is found by another line search, this time along a quadratic curve passing through the reactant, product, and transition state guess found previously. The new transition state estimate is found by optimisation of the maximum along this line. This quadratic process can be repeated. In figure 1.23 the QST steps are shown by the green lines, solid for the first step, and dotted for the second.

1.3.11.5 Slowest ascent

Many techniques to find the minimum energy pathway (MEP) and saddle point have been developed over the years, two of which are described here. For all methods the start points (reactants of the reaction being studied), and in cases that involve two points, the endpoints (products) are all optimised geometries, so are all in local minima on the PES.

Possibly the most elementary and a good starting point is the route of slowest ascent.[56] This method, demonstrated in figure 1.24, starts at the minimum. From there the gradient is measured in every direction, and a step is taken in the direction that has been determined to have the minimum gradient. This is repeated until hopefully a maximum is reached, and the path begins to descend to the other side.

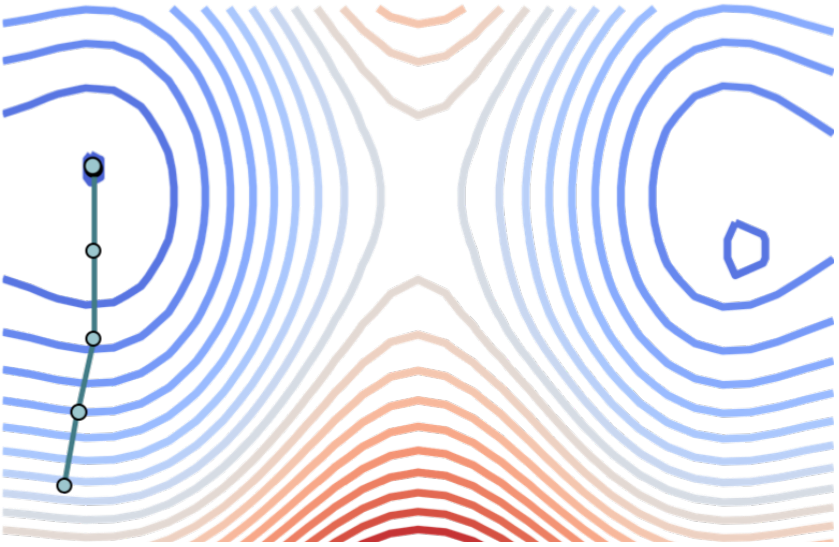


FIGURE 1.24: Sketch diagram of slowest ascent path. Blue is low energy and red high on the contour map. In the case shown here, it will not make it over the energy barrier. This is a sketch, so the path may not be the exact path that a slowest ascent path would take.

As figure 1.24 shows, whilst this might be a simple concept, it can often go awry, and is not guaranteed to give successful results. The calculation can become stuck without ever making it over the energy barrier.

1.3.11.6 Nudged elastic band theory

One of the most well-used techniques for MEP searches is the Nudged Elastic Band (NEB) theory.

This is a "chain of states method", which takes the direct path between the initial and final states of a system, and creates a chain of points (beads) with a considered to be attached by springs. Using springs adds elasticity to the path. These are illustrated in figure 1.25 where the black dots and dashed line show the initial path.

These images are all given an object function, relating the energies of the images to the positions of nearby images, and the forces acting on them. The force considered for each image is shown in equation (1.83), where k_i is the spring constant for the spring section after the image at i, $\vec{R_i}$ is the image at i, and $\vec{\nabla} V$ is the potential energy at $\vec{R_i}$.[57]

$$\vec{F_i} = -\vec{\nabla} V\left(\vec{R}_i\right) + k_{i+1}\left(\vec{R}_{i+1} - \vec{R}_i\right) - k_i\left(\vec{R}_i - \vec{R}_{i-1}\right) \tag{1.83}$$

This object function is then minimised for all images at once, and repeated with the new positions until the image converges. This can take more than 500 iterations.[58]

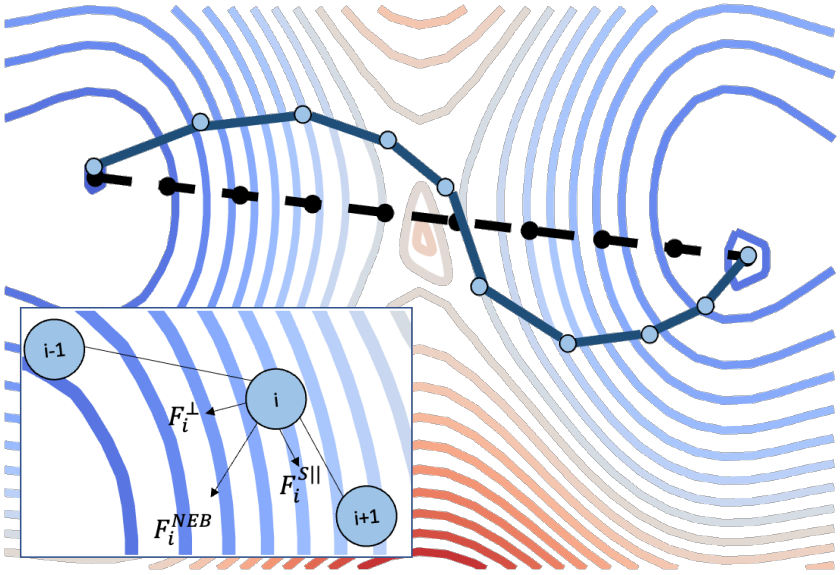


FIGURE 1.25: Sketch diagram of an NEB path. Blue is low energy and red high on the contour map. Here the dashed line is the initial path, and the blue line is a final path, with the circles representing the positions of each bead. This is a sketch, so the path may not be the exact path that the NEB path would take.

This works generally, although there can be issues, such as slipping, where the forces can push the points apart, over the top of a key part of the topology.

To solve this issue the 'nudge' comes in. This means that the forces are being projected out at an angle so that they will not interfere with the relaxation of the images, shown in figure 1.25.

To do this the energy and the gradients need to be found for each initial image. Then a tangent to the path is estimated for the point on the path i, taking account of the position of $i-1$ and $i+1$. The parallel component of the spring force is added as well.

There are variations of NEB, such as Climbing Image NEB (CI-NEB), where the image with the highest energy is no longer considered as being affected by the spring force and is instead subjected to a saddle point search.[59]

1.3.12 Using results to find reaction parameters

The activation energy of a reaction is the amount of energy that is required to get a reaction to happen, including getting it to overcome the energy barrier.

This is related to the rate constant, k, of a reaction by equation (1.84).

$$k = Ae^{-E_A/RT} \tag{1.84}$$

where A is a pre-exponential factor, E_A is the activation energy, R is the ideal gas constant, and T is the temperature of the system. These variables are usually determined experimentally.

By considering the linear plot of $\ln k$ against $1/T$, this can be fitted to $y = mx + c$, giving us the following two relationships

$$E_a = \Delta H^{\ddagger} + RT \tag{1.85}$$

$$A = \frac{k_B T}{h} e^{\left(1 + \Delta S^{\ddagger}/R\right)} \tag{1.86}$$

These lead to the Eyring equation

$$k = \frac{k_B T}{h} e^{-\Delta G^{\ddagger}/RT} \tag{1.87}$$

where k_B is the Boltzmann constant and h the Planck constant.[60] $\Delta G^{\ddagger}$ is the Gibbs free energy of activation of the transition state, which can be defined by in terms of enthalpy and entropy as

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{1.88}$$

This means that an entropic term now seems to have been added. However, in the Arrhenius equation, the entropic term is accounted for by the pre-exponential factor A, so it has not appeared from nowhere.

These equations can be used to find many properties of a reaction, even including viscosity and diffusion.[61]

1.4 Computational Software and Tools Used

Several different computational programs were used for the work in this thesis. Some of their differences and advantages are discussed here. The specific reason why each program is used will be discussed in the later chapters.

1.4.1 DFT programs

1.4.1.1 NWChem

The NorthWest Computational Chemistry (NWChem) software package is a DFT computational code, which is cubically scaling to the number of electrons in a system.[62,63]

Input files must include start-up directives, the definition of the chemical system, the specification of various parameters for the calculations, and task directives. It can run almost any electronic structure theory calculation needed in modern computational chemistry. DFT in NWChem uses Gaussian basis sets and is kept up to date with most exchange-correlation functionals available in the literature.

While it can be used as a text-based interface, it can also be used through external packages, such as the python library ASE (Atomic Simulation Environment).[64]

1.4.1.2 ONETEP

ONETEP is a linear scaling DFT package.[65,66] ONETEP uses

$$\rho\left(\mathbf{r},\mathbf{r'}\right)=\sum_{\alpha\beta}\phi_{\alpha}\left(\mathbf{r}\right)K^{\alpha\beta}\phi_{\beta}^{*}\left(\mathbf{r'}\right) \tag{1.89}$$

to write its density matrix, where ϕ are localised functions known as non-orthogonal generalised Wannier Functions (NGWFs), mentioned previously in section 1.3.10, and $K^{\alpha\beta}$ is the density kernel.[67] Truncating the kernel using

$$K^{\alpha\beta}=0 \quad \text{when} \quad r_{cut}<\left|\mathbf{R}_{\alpha}-\mathbf{R}_{\beta}\right| \tag{1.90}$$

where $\mathbf{R}_{\alpha}$ and $\mathbf{R}_{\beta}$ are the centres of the localization regions of the functions $\phi_{\alpha}(\mathbf{r}$ and $\phi_{\beta}(\mathbf{r}$, which enforces localisation.[68]

The NGWFs are defined using periodic sinc, psinc functions, as a basis set, as shown as follows:

$$\phi_{\alpha}(\mathbf{r})=\sum_{k}D_{k}(\mathbf{r})C_{k,\alpha} \tag{1.91}$$

In this formula, $D_k(\mathbf{r})$ is a psinc function (a very localised spike) centred at point $\mathbf{r}_k$ and $C_{k,\alpha}$ is the expansion coefficient.[69] This provides us with a large basis set (plane-wave) level accuracy.

To calculate the single-point energy of a system, ONETEP uses two nested SCF loops. First of all, an initial guess for the NGWFs is given, and the kernel is iteratively converged, using the Li-Nunes-Vanderbilt (LNV) algorithm, until $K^{\alpha\beta}$ and $H_{\alpha\beta}$ commute. Because the LNV method uses a sparse matrix, this step scales linearly. Once this is converged, then the energy is minimised with respect to the NGWFs (ϕ) using the conjugate gradient method, which also scales linearly with respect to system size.

For geometry optimisations, ONETEP uses the L-BFGS algorithm shown in figure 1.21 on page 43, and the solvent method is an implicit method based on the work of Fattebert.[70] This model aims to remove all error from outlying charge. One of the differences in this model is that it forms a smooth transition for the barrier around the cavity, as opposed to a solid. This means that instead of the discrete function in equation (1.78) on page 38 the change in dielectric constant is continuous.

1.4.2 Molecular dynamics programs

1.4.2.1 Forcite

Forcite is a molecular dynamics tool available through the Materials Studio software environment.[71]

To run a Forcite calculation on a structure loaded in Materials Visualiser, the calculation task (either energy or geometry optimisation), the desired quality level for the task, and the force field must be selected. Forcite also has an analysis feature that can calculate and plot information such as temperature, pressure, volume, stress, cell parameters, and mean squared displacement, along with many others.

Forcite can be used with several different force fields, including the proprietary force field COMPASS II, a force field that has been used in several previous studies of nitrocellulose, such as in papers by Nardai and Bohn[72]. The results of this are discussed in section 2.4 when the application of these methods on nitrocellulose is further explored.

1.4.2.2 LAMMPS

Large-scale Atomic/Molecular Massively Parallel Simulator, LAMMPS, is a classical molecular dynamics code used for materials modelling.[73] It can be used to model ensembles of particles in liquid, solid or gas phase, in 2D or 3D systems with anything from a few particles up to billions.

It can be used on a single computer, but for its full potential should be used on parallel computers, which must support the MPI message-passing library. It is written in C++,

and is open source, all versions since it was first released in 2004 can be downloaded from the LAMMPS website.[74]

When running a calculation using LAMMPS, an input script is needed that contains

- The force field parameters and system interactions.
- Details of the system. This can be a continuation from a previous calculation, or an new input file, possible generated by Moltemplate.
- Specifics of the calculation that is being run - the ensemble, type of calculation, timesteps, properties to calculate and many more.

An example script used for this project can be seen in Appendix B.

LAMMPS has been widely used for simulations of polymers like cellulose and nitrocellulose. In particular, the OPLS (Optimised Potentials for Liquid Simulations) force field is of interest.[9,13] This has been used previously on nitrocellulose by Richards and is investigated in this thesis.[75]

1.4.3 Atomic visualisation programs and other tools

1.4.3.1 VMD

VMD (Visual Molecular Dynamics) is a molecular visualisation software, that can display many different structures simultaneously, with a variety of different representation styles.[76]

The program was developed with the specific need to display dynamic data, like molecular trajectories, with a graphical user interface and visualization console, and a text-based as well as a mouse-based user interface. It also includes a selection of analysis tools, like an RMSD calculator.

1.4.3.2 Avogadro

The selection of molecular builders available is very wide. Some only work on specific operating systems, some need paying for, and some are free. However, when it comes to needing to build molecules the options are more limited.

Avogadro is a molecular visualisation software, but it is also a free, open-source, cross-platform, three-dimensional, molecular builder, a list of qualities that are not often all found together.[77,78] It has an atom-by-atom molecular builder that makes it simpler to build small molecules by hand, and it also has a fragment library for

common molecules. It also has extensions that can be used to write simple input files for several computational codes, including NWChem.

1.4.3.3 Materials Visualiser

Materials Visualiser is the basis of the Materials Studio software environment, it can be used to build, manipulate and visualise molecules, and to interact with them through any of the other tools available through Materials Studio.[79] It contains atom-by-atom, fragment and specialist crystal and polymer builders, which have all been used in the course of this project so far.

1.4.3.4 Moltemplate

Large crystal models were built using Moltemplate, a molecule-building tool for LAMMPS.[80] Initially designed for building custom coarse-grained molecular models, it can also be used to prepare structures for all-atoms simulations as well.

One disadvantage of Moltemplate is its randomisation functionality. The input files used for the structures in this project are in Appendix A. The structures are built stepwise, from units into chains, into planes and then into full crystals in separate input structure files. All bonds must be defined, either within the unit or along the chain, including through the PBCs.

The order of units within a chain can be randomised using a random function. However, a new chain is not generated each time the 'chain' input file is called. This means that to get an $nxnxn$ crystal structure with a fully randomised combination of initial units, either n^2 chain files must be written or all n^3 units and their connectivity must be defined in one single input file.

Due to this, all of the structures used in this thesis are 1 single repeated unit or alternate between two units.

1.4.3.5 Multiwfn

Multiwfn is an analysis program that can be used to visualise wavefunctions.[81] It is open-source, free and user-friendly.

Chapter 2

Cellulose and Nitrocellulose, Background and Literature

Nitrocellulose has a wide range of uses. Although best known for its use in explosives, it is also used for propellants, coating films, and printing inks.[82]

Nitrocellulose is formed from the nitration of cellulose, through the following esterification reaction, which is reversible and highly exothermic.[83]

$$\mathrm{R{-}OH + HONO_2 \rightleftharpoons R{-}ONO_2 + H_2O} \tag{2.1}$$

Found as the main structural component of cell walls in plants and algae, cellulose is one the most commonly occurring natural materials.[84]

Despite the importance of nitrocellulose, it is not fully understood at the molecular scale. Given its unstable nature, this lack of insight has led to some of the worst chemical accidents of the 21st century, such as the explosion at the Port of Tianjin on 12th August, 2015.[85] This explosion, and several others over the last few years were linked to the nitrocellulose being in powder form, and a more accurate knowledge of the mechanisms that cause the ignition reactions could lead to improved safety measures to prevent incidents like these.[86]

The aim of this thesis is to present some methods by which these mechanisms can be investigated computationally. First, it focuses on using molecular dynamics to build large-scale nitrocellulose systems containing thousands of atoms, and testing how accurately they are modelling natural nitrocellulose. It then looks at how DFT can be used to study decomposition reactions of the nitrate groups in the context of chains longer than a trimer, and also in large-scale systems, thereby testing how the interchain interactions also affect the reaction energies.

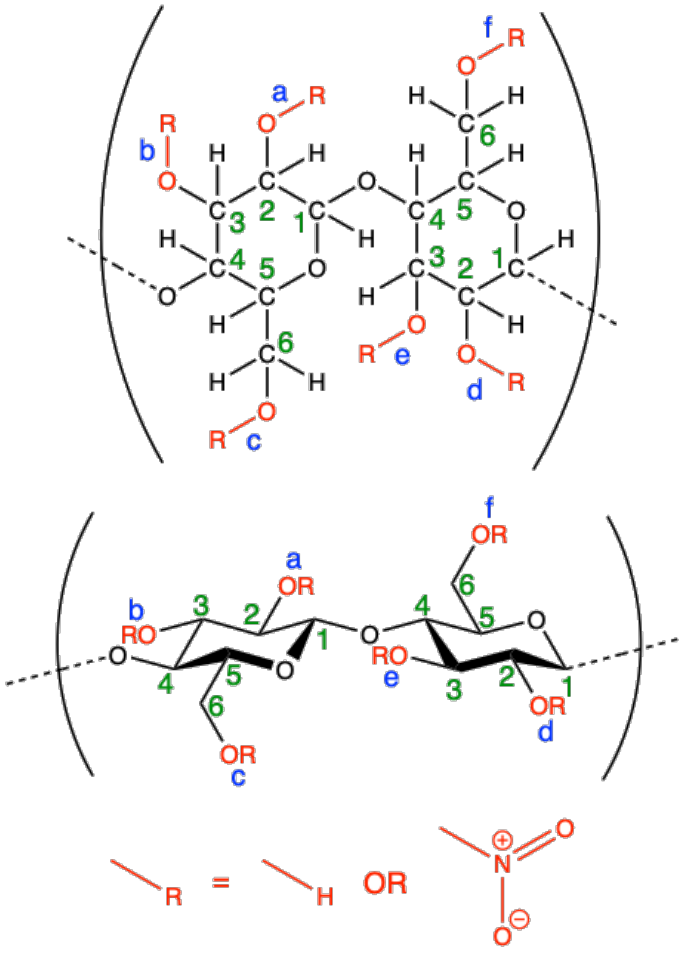


FIGURE 2.1: Chemical structure for cellulose and nitrocellulose, a disaccharide that is known as cellobiose, shown in skeletal structure and chair conformation. The green numbers are how the carbons are labelled for structural labelling. The blue letters are how the positions of the R groups will be referred to in this thesis. For cellulose all R groups are hydrogen, and when these are replaced by nitrate groups the cellulose becomes nitrocellulose.

2.1 Cellulose

2.1.1 Structure and history of cellulose

Throughout history, cellulose has been extensively used all around the world. Today it is used in the production of paper, textiles, food additives and propellants.[84]

Found in cotton and rayon, cellulose is a naturally occurring structure. Like starch and callose, it is formed of chains of glucose rings with the formula $C_6H_{11}O_5 - (C_6H_{10}O_5)_{x-2} - C_6H_{11}O_6$. In the case of cellulose, the glucose rings are connected with $\beta - 1,4$-glycosidic bonds. A repeat unit is shown in figure 2.1, when all the **-R** groups are hydrogens.[87]

The first known case of plants being purpose-grown for their fibres is hemp in ancient China (4500 BCE). Hemp was used for rope and cord due to its durability, strength and versatility, all properties that are a result of hemp's cellulose content. Papyrus was used by Ancient Egyptians in 3000 BCE to make early paper, due to its availability but also due to its fibrous nature, which again is thanks to its cellulose content.[88]

But despite its importance and history, cellulose was not officially chemically identified until 1837 and the idea that it is a polymer was not fully accepted until the 1920s. The discovery of useful cellulose derivatives, in particular nitrocellulose, was one of the driving forces for the new research. More information about the discovery of cellulose is collected in the timeline in table 2.1.

2.1.2 Crystal structures of cellulose

Cellulose is found in many different crystalline forms, with chains built up to structures that can contain up to 300,000 residues.[87] The two most commonly studied structures are known as cellulose Iα and cellulose Iβ.

When creating simulated structures of cellulose one factor that can significantly affect the resulting model is the size of the simulation, and the use of periodic boundary conditions, as these have large impacts on the computational time, but also on the structure created. This means that as computer capacity and speed have increased over the years, better models are able to be developed, with fewer assumptions needing to be made to make the simulations feasible.

For example, in 2007 French and Johnson modelled cellulose using nineteen chains, each eight residues long, in cellulose Iβ form, as this was a significant size for the available computational resources at the time.[89] The issue with these results was an unnatural twisting of the chains, not found when compared to experimental structures. Just the next year, they were able to almost double the size of the models, from 19 chains to 37, which changed the ratio of surface molecules in comparison to total molecules. They also now included periodic boundary conditions and were able to use QM calculations for comparisons, although at this stage they were only able to use QM for minimization on single repeat units. They were able to use their conclusions from this work to support that the interchain hydrogen bonds are extremely influential over the structure predicted, and this evidence has led to work on improving the hydrogen bonding in different force fields like MM4.

When it comes to the modelling of cellulose, there are conflicting views on what sort of cellulose model best portrays natural cellulose crystals, either that it should be fully crystalline, or that it should contain a combination of crystalline regions and amorphous regions.[90] The cellulose structure that is chosen should depend on what aspect of cellulose is being studied. If the depolymerisation of cellulose is being

1819	Braconnot first noticed a relationship between the properties of cellulose and glucose
1833	Braconnot finds xyloidine by dissolving plant substances in nitric acid, the first 'man-made' cellulose derivative, and a precursor to nitrocellulose
1837-1842	Payen extracts the fibre from different parts of a plant (tissue, seed, root tips etc.) and recognised these are all the same chemical substance
1839	The name 'Cellulose' is first used by the French Academy
1858-1877	**von Nägeli uses X-ray diffraction to study fibre bundles and says the visible fibres were further subdivided into submicroscopic crystalline structures**
∼ 1900	*Cellulose was believed to be a colloidal substance*
1913	Denham and Woodhouse find that when adding NaOH and dimethyl sulphate a trimethyl derivate is formed
1914	**Nishikawa suggests that the crystallinity of cellulose is discontinuous**
1914	*Tollens first suggests that cellulose might be a chain-like structure*
1920	**Herzeg and Jancke use X-ray analysis on different sources of cellulose (cotton, rayon, wood and more) and found that they have identical crystal structures. This becomes known as native cellulose.**
1923	Irvine and Hirst find the 2,3,6-trimethyl glucose is created after full hydrolysis of cellulose. The conclusion from this and the results of Dennham and Woodhouse is that the repeat unit of cellulose has 3 hydroxyl groups, 2 secondary and 1 primary at positions 2, 3 and 6.
1925	*Chain theory receives support after the polymerisation of formaldehyde and other substances is proven.*
mid-1920s	*Association theory goes against the colloidal theory. This says that cellulose is anhydroglucose with the empirical formula $C_6H_{10}O_5$, and the association between these molecules causes the high molecular weight that can be calculated*
1927	Zemplen, Haworth, Hirst et al. establish the position of the glycosidic bond
1932	Further testing of acidic hydrolysis of cellulose finds a small amount of 2,3,4,6-tetramethyl glucose has also been formed, leading to the conclusion that the end groups must have -OH groups on position 4
1938	**An accepted monoclinic unit cell for native cellulose, proposed by Meyer and Misch, is decided on**
1939	Richtmyer and Hudson establish that the units of cellulose are connected by β linkage, as opposed to α
1953	*Staudinger gets a Nobel prize for work into polymers including cellulose*
1955	**Balashov and Preston report that the width of a microfibril is different depending on the type of cellulose**

TABLE 2.1: A timeline to the discovery and understanding of cellulose, using details collected in the article 'Cellulose: a random walk along its historical path'.[88] Most events are discoveries about the chemical structure of cellulose, *events in italics about the polymer structure*, and **events in bold about the crystal structure**.

studied, amorphous cellulose is important to look at, as it is more 'open', the flexibility of the chain allows more access to the glycosidic bonds which are the bonds that break in this process.[91]

Some studies such as that by Meader *et al.* in 1978, model cellulose as chains, using several different helix parameters, and build systems from these.[92] One of the best models they found was a 5_2 helical conformation, where the spatial periodicity is 2.54 nm, with 5 glucose rings per period.

Most other studies create structures using a unit cell and use this to construct a crystal structure. Many computational chemistry studies of cellulose use microfibril structures as this is the most commonly occurring natural form. Microfibrils are fibres that occur naturally in wood pulp, and they can be up to 3 nm x 3 nm in cross-section, and up to several μm in length.[93] These are often simulated using a limited number of 'infinitely' long chains, using periodic boundary conditions. It has been found that the degree of polymerisation affects the crystal structures of the simulated results, as shown in papers by Bregado *et al.* and Srinivas *et al.* [19,94]

These previous studies into modelling cellulose structures will be used in this project as a basis for the simulations run. For example, the protocol discussed in section 1.2.7.2 and shown in table 1.2 on page 18, which was used for creating amorphous structures from crystalline cells, was used as a practice before the protocols described in chapter 3 were run.[19] It also demonstrated the importance of changing the conditions gradually, which was found to be a very important issue in the development of the protocol in chapter 3.

Some computational studies into the degradation of cellulose have also been performed, which, along with research into the degradation of nitrate esters and nitrocellulose described in the next section, will be used to build a model for the degradation reactions for the nitrocellulose created in this thesis. In particular prior studies into cellulose tend to focus on pyrolysis reactions, which involves the breaking down of a polymer chain into smaller parts, often referred to as "active cellulose".[95]

One study on this reaction was by Zheng *et al.* in 2016, using molecular dynamics simulations, and the ReaxFF forcefield.[96] The advantage of using MD and a reactive forcefield is that this increases the size of calculation that is possible - 7572 atoms were in the system being tested. The system was tested in different temperature ranges (500 - 1400 K to see the evolution of the products, and 1500 - 2800 K to focus on the inorganic gas behaviours). Molecular dynamics simulations then evolved the systems over time, up to 1400 ps, to see how they changed. They able to show that shorter simulations at a higher temperature were able to approximate longer simulations at lower temperatures, an observation that could be used to reduce compuational time in the future. Several pathways for the pyrolysis reaction were found, creating products

such as hydroxyl-acetone, 2-hydroxy propyl aldehyde, and glycolaldehyde, which agreed with the products seen in experiments.

The mechanisms calculated here support DFT research performed by Zhang *et al.* in 2015 and Wang *et al.* in 2020.[90,97] Zhang *et al.* used DFT calculations in GAUSSIAN 09 to calculate the enthalpy and Gibbs free energy of multiple different pathways and use these to conclude which mechanisms are the lowest energy, so most likely to happen. The energy barrier to the reaction they found to be most likely was 146 kJ mol^{-1}, which is actually similar to the activation energies found for the decomposition of nitrate esters in the next section, but most of the other mechanisms were significantly higher energy, meaning that investigation of the hydrolysis and thermolysis reactions were prioritised in this project.

2.2 Nitrocellulose

2.2.1 Structure and history of nitrocellulose

In 1833 Braconnot first discovered that an explosive could be created by dissolving plant matter in nitric acid, finding a substance he called 'xyloidine', a form of nitrocellulose. Later, in 1846, a German chemist called Christian Schönbein used a cotton apron to mop up a spillage of hot nitric and sulfuric acids and realised he had created a reactive derivative of cellulose, which he called gun cotton.[98,99] He realised the potential of the substance as a propellant, due to its high energy properties, and its smokeless burning and patented his method of creating nitrocellulose.[100] Being smokeless was a major advantage, previous gunpowder created large clouds of smoke, both giving away your position and obstructing any view you might have of your enemy.

From the 1860s nitrocellulose was being used in gunpowder, but it was still poorly understood, leading to a large number of accidental explosions. In fact, it is said that more casualties may have occurred through accidents in the factories creating nitrocellulose products, than through actual use in combat.[101] 1889 saw Frederick Abel, an English chemist and explosives expert, and James Dewar, a Scottish scientist, patenting a mix of guncotton and nitroglycerine known as cordite, which was more stable and reliable than pure nitrocellulose, which was soon being used by the British Army and Navy as their standard propellant.[102,103]

2.2.2 Uses of nitrocellulose

We have just discussed that since its first discovery, there has been significant interest in the explosive properties of nitrocellulose. However, it is used for many purposes in our modern world.

One of the other most important uses of nitrocellulose could possibly be attributed to the reducing number of wild elephants in the mid-1870s and the popularity of billiards. Humans had used most of the easily accessible resources that the earth had to offer which could be used to make billiard balls, such as ivory, so an artificial substitute was needed.[101] A man named John Wesley Hyatt discovered celluloid, a material created by mixing nitrocellulose with camphor and alcohol, under pressure. The solid celluloid was able to form a lacquer film due to the plasticity caused by the camphor.

Soon celluloid was also being used in thermoplastics for cinema film, due to its flexibility. However, while it helped advance the film industry, it also became known for its tendency to combust, meaning celluloid films were soon replaced.

The ability to form a hard polished lacquer has lead to many other uses of nitrocellulose. Celluloid has been used as substitutes for glass, ivory, amber, coral, pearl, enamel, rubber and wood.[101] Pyroxylin (another name for nitrocellulose) paints became favoured for painting cars. These days, one of the most common 'paint' uses for nitrocellulose is in nail varnish and other cosmetic uses.[104]

The fibrous nature of nitrocellulose, as well as its solubility in water, also made it very important in the early manufacture of artificial silks.[105] Nitrocellulose silk (known as Chardonnet silk) is created by nitrating cotton and then pushing the solution through very small holes to create thin strands, exploiting the insolubility and fibrous nature of nitrocellulose.[106] The resulting thread is then denitrated back into cellulose, to lower its flammability.

Even the explosive nature of nitrocellulose can be harnessed for purposes other than war - nitrocellulose dissolved into nitroglycerine gives it a gelatinous consistency between jelly and syrup depending on the proportions, creating a explosive called Ballistite, patented by Alfred Nobel.[107] Nitrocellulose-based explosives like this have been used in mining, the creation of tunnels, and even the clearing of forested areas for agriculture.

One of the biggest current concerns about nitrocellulose is its effect on the environment. It is insoluble in water which means that when it gets into the environment (either through the water used in the manufacturing process or through activities such as live-fire military training) it is very hard to get rid of.[108,109] This means that there are a lot of studies on the breakdown reactions and how they can be sped up.

2.2.3 Nitrogen content

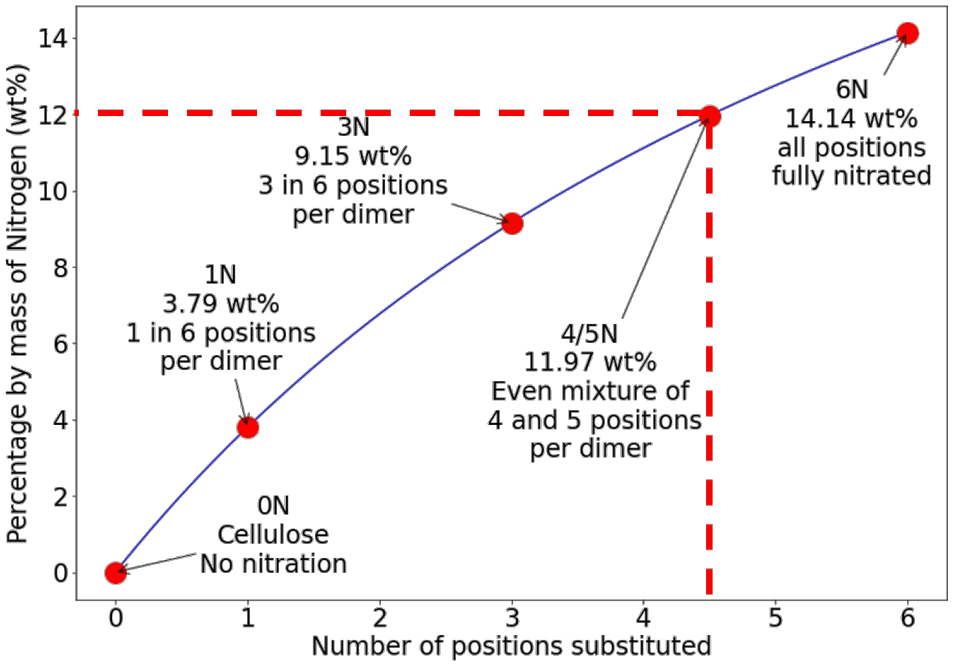


FIGURE 2.2: Nitration levels for nitrocellulose structures with different levels of substitution, where the number of positions substituted is the number of -R group positions in Figure 2.1 that are changed to nitrate groups. The 5 levels of substitution being simulated in this project are highlighted on this graph. The dotted line highlights 12 wt% nitrogen, the accepted limit of explosiveness.

The nitrogen content of nitrocellulose has a strong influence on its properties, such as the viscosity, and most importantly, the stability - the more nitrated the nitrocellulose, the less stable it is.[110] In this work, the "level of substitution" will be used to refer to the number of positions of the **dimer** structure that have been nitrated, out of the possible 6. This means that the nitrate content can be calculated using the following formula:

$$\text{percentage by weight} = \frac{\text{SUB} \times 14.007}{318.234 + \text{SUB} \times 46.005 + (6 - \text{SUB}) \times 1.008} \times 100 \qquad (2.2)$$

where SUB is the level of substitution (where SUB is ≤ 6, and SUB $= 6$ means all of the R groups in Figure 2.1 are nitrated). Figure 2.2 shows the relationship between the amount of substitution and the nitrogen wt% for significant nitrate levels. The maximum theoretical amount of nitration is all 6 positions of the dimer unit, a structure that contains 14.14 wt% of nitrogen.

Structures that are above 12 wt% nitrogen are classified as highly nitrated, are unstable to heat, and are mostly used in the manufacture of explosives.[111,112] Structures below 12 wt% nitrogen content are known as moderately nitrated, they are still flammable

and also have increased solubility in alcohol and ethers.[113] However, nitrocellulose quickly becomes highly unstable, meaning that experimentally, the highest level of nitration currently reported is 13.9 wt%.[114,115]

In the study of nitrocellulose, there are two main time periods that are focused on: the "safe service time period" - defined as the time before the nitrocellulose auto-ignites; and the "reliable service time" - defined as the time before the nitrocellulose has degraded sufficiently that the bulk properties are no longer as required, within tolerance limits.[116]

2.2.4 Decomposition reactions of nitrocellulose

There are two key ways in which nitrate esters like nitrocellulose can break down via three pathways:[117,118]

1. Acidic hydrolysis - the nitrate ester reacts with an acid, to create an -OH group, and HNO_3, nitric acid.

$$RONO_2 + H^+ \rightleftharpoons RONO_2H^+ \tag{2.3}$$

$$RONO_2H^+ \longrightarrow ROH + {NO_2}^+ \tag{2.4}$$

$${NO_2}^+ + OH^- \longrightarrow HNO_3 \tag{2.5}$$

2. Base hydrolysis - the nitrate ester reacts with a base, to create an -OH group and also creates NO_3^- and NO_2^- ions.

$$RONO_2 + OH^- \longrightarrow ROH + {NO_3}^- \tag{2.6}$$

$$RCH_2ONO_2 + OH^- \longrightarrow RC(O)H + H_2O + {NO_2}^- \tag{2.7}$$

3. Thermolysis - the bond breaks homolytically, to create two radicals.

$$RONO_2 \longrightarrow RO\cdot + NO_2\cdot \qquad \text{(initiation)} \tag{2.8}$$

$$RO\cdot + NO_2 \longrightarrow ROO\cdot + NO \qquad \text{(propagation)} \tag{2.9}$$

$$2\,ROO\cdot \longrightarrow RC{=}O + O_2 + ROH \qquad \text{(termination)} \tag{2.10}$$

Hydrolysis tends to occur at temperatures less than 100 °C in the presence of water, and thermolysis happens at higher temperatures, generally between 100 °C and 200 °C.

Once these initial reactions have happened, an autocatalytic chain reaction propagates, breaking down the polymer chains - when one of the products of decomposition is a catalyst for the decomposition reaction.[116]

In 2001 Bohn confirmed that both hydrolysis and thermolysis happen in the degradation of nitrate esters.[116] It was found that the activation energy of the thermolysis reaction is 160 - 170 kJ mol^{-1}, and the activation energy of hydrolysis is 100 kJ mol^{-1}. They then found the overall activation energy for the decomposition of nitrocellulose as a whole, and it was 146 $\pm$ 5 kJ mol^{-1} and this is between the two energies of the specific reactions, it was concluded that the nitrocellulose must be degrading by a mixture of the two. These energy values are plotted in figure 2.3.

Studies over the years have produced many values for the activation energy of degradation of nitrocellulose. The wide range of values here is due to the fact that there are many aspects to the reaction, and different methods of detection can be used to highlight the effects of different parts of the reaction.[82] Having an accurate understanding of the reaction energies is of great importance. Knowledge of these can lead to rate constants being calculated, and the rate constant can lead to the calculation of the time periods mentioned at the end of section 2.2.3, and also to insight into how storage conditions could be adjusted to increase the “reliable service time”, before the nitrocellulose has degraded sufficiently that it is not able to perform to the expected requirements.

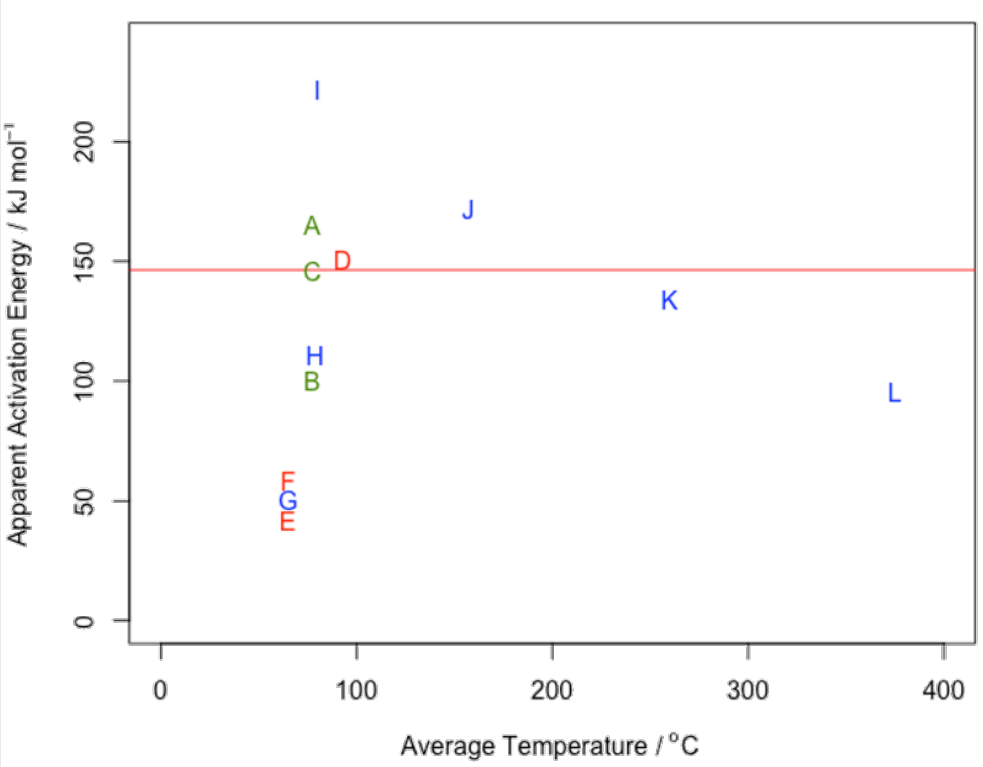
Many different analytical techniques have been used on nitrocellulose, each giving information about a different property of the chemical. A review of these techniques was written by Fernandez de la Ossa *et al.* (2011).[83]

The energy of bond dissociation for a nitrate ester is 146 kJ mol^{-1}, so it was suggested by Kimura in 1989 that the activation energy for reactions involving the breaking of this bond will be similar.[119] Figure 2.3 shows some recorded activation energies, and the bond dissociation is plotted as a horizontal line.

In the same paper, they reported that they had found the activation energy for the decomposition of nitrate esters to be 146 - 154 kJ mol^{-1}, using the Taliani test , which involves measuring the pressure change as a propellant is heated in a fixed volume. The paper also tried finding a value for nitrocellulose by analysing the luminescence but found the activation energies they got through this method to be a lot lower, due to impurities, particularly the oxygen in the air.

In 1997 Brill and Gongwer wrote a review paper, looking at a range of different reactions from various experiments, over a range of temperatures.[118] The type of reaction depends on the temperature of the surroundings and affects the activation energy, as seen in figure 2.3.

The experimental study that is the main subject of the paper includes data for several different types of reactions, performed in different experimental conditions.[118] The highest energy reaction seems to be the depolymerisation, which was performed by gel permeation chromatography over a temperature range of 70-90 °C.



Symbol	Experimental detail	From paper by
A	Homolytic cleavage in nitrate esters	Bohn 2001
B	Hydrolysis in nitrate esters	
C	Overall degradation of nitrate esters	
—	Bond dissociation energy of nitrate ester	
D	Homolysis; using Taliani test	Kimura 1989
E	Homolysis; using chemiluminescence for NC in air	
F	Homolysis; using chemiluminescence for NC in N_2 gas	
G	ROO formation energies for nitrate esters	
H	Hydrolysis energies for nitrate esters	
I	Depolymerisation energies for nitrate esters	Brill and Gongwer 1997
J	First-order autocatalysis energies for nitrate esters	
K	Flash heating energies for nitrate esters	
L	Combustion energies for nitrate esters	

FIGURE 2.3: This graph collects some activation energies given in 3 papers by Bohn 2001 [116] in green, Kimura 1989 [119] in red and Brill and Gongwer 1997 [118] in blue.

This depolymerisation reaction was studied by Pfeil *et al.* in 1985 and was found to be an autocatalytic reaction. [120] By quenching the reaction at various points, and measuring the nitrate content through measuring the infra-red absorbance of the nitrate groups, the rate was found to be able to be approximated by a first-order

reaction.

The least studied of the three types of reaction described at the start of this section is thermolysis, the homolytic breakdown of the reaction, which leads to a chain reaction creating RO· (alkoxy radicals) and ROO· (alkyl peroxy radicals). One particular study, in 1989, by Kimura did investigate the activation energies for thermolysis, through a chemiluminescence method.[119] The termination step of the reaction is highly exothermic and so accompanied by an emission of light, which can be measured, and by measuring how the amount of light being emitted is affected by the temperature, the kinetics of the reaction can be found. They found out that the activation energy for the thermolytic decomposition of nitrocellulose was significantly affected by the atmosphere it was in - oxygen in particular was found to decrease the activation energy and increase the luminosity.

2.3 How could we use computational chemistry to study nitrocellulose?

The aim of this thesis is to use molecular dynamics simulations to create large-scale systems of nitrocellulose, and then use DFT to investigate the decomposition reactions in this bulk situation, as opposed to just in a solvent approximating a solution. Although there does not seem to be any work done previously with this combination of focuses, there are many previous studies that can be used as guidelines.

For example, there is some literature into the creation of nitrocellulose structures used MD simulations, although a lot of these focus on mixtures of nitrocellulose and a second chemical, as nitrocellulose mixtures are commonly found in propellants.[75] More studies of nitrocellulose are discussed in the following section.

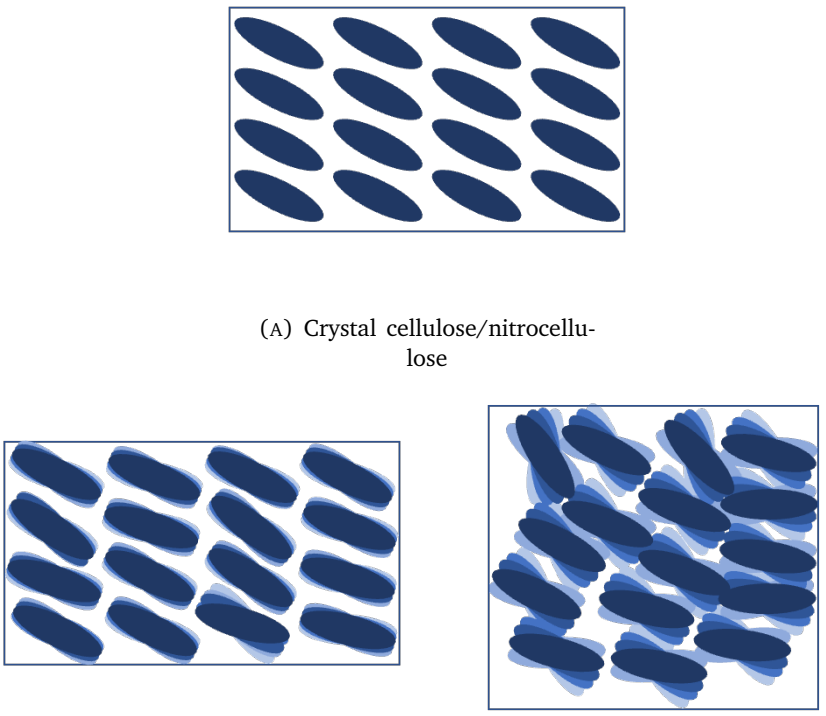
However, the simulation of crystalline and amorphous cellulose has been much more widely published,[89,90] as discussed in section 2.1.2. Some of these studies were used as a basis for this projects creation of large-scale, non-crystalline nitrocellulose structures.[19]

When it comes to the study of the degradation reactions, DFT is more useful. There again is limited research published into these reactions, although nitrate esters have been studied, concepts for which could be applied to nitrocellulose calculations, as both contain the same functional groups. These will be discussed in the section 2.5.

This project also wants to run the degradation reactions in the context of bulk systems created using MD. DFT often would not be considered for calculations of this size due to it being cubically scaling, however, the use of ONETEP in this project will make running DFT on systems of this size possible.

2.4 Previous Investigations of Crystal Structures of Cellulose and Nitrocellulose

The helix conformation of cellulose found by Meader *et al.* in 1978, previously mentioned, and the unit cell found based on this, has since been used in several molecular dynamics studies. It is also used for nitrocellulose, by nitrating the unit cell.[92]



(A) Crystal cellulose/nitrocellulose

(B) Cellulose

(C) Nitrocellulose

FIGURE 2.4: Comparative sketches of 3 systems inspired by the work of Nardai and Bohn[72] - (A) A crystal structure of cellulose/nitrocellulose - this is an end on view, looking along the chains, which are all regularly spaced and straight. (B) Cellulose after equilibration - the chains are not as regularly spaced, and twisting along the chains has begun to occur. (C) Nitrocellulose after equilibration - all of the chains are more twisted, and the arrangement of the chains is also twisting. The shape of the simulation cell has also changed.

In 2017, Nardai and Bohn used this unit cell to show how nitrating the crystal increases flexibility.[72] Figure 2.4 shows a sketch summary of their results - the cellulose structure stayed mostly intact, and the nitrocellulose distorted out of its original position, after being subjected to the same molecular dynamic simulations.

In most molecular dynamics studies of nitrocellulose, there is a specific application in mind, and this affects how nitrocellulose is examined. Often it is considered part of a mixture with other chemicals. For example, in a paper from 2008, Ma *et al.*

nitrocellulose was considered as both a single base propellant (pure nitrocellulose), and a mixed propellant of nitrocellulose and nitroglycerine, and the work compares the energies and properties such as the detonation energy of the different systems[121] .

Tables 2.2 on the facing page and 2.3 on page 68 compile the details of a variety of different structures for cellulose and nitrocellulose, computationally generated and experimental.

2.4.1 Forcefields used for nitrocellulose and cellulose MD studies

Many molecular dynamics simulations of nitrocellulose and cellulose have been reported in the literature, using a variety of different forcefields. However, the requirements for forcefields will be different, depending on the focus of the study.

Cellulose is a carbohydrate, and there are many forcefields specifically parameterised for these. One of the most commonly used forcefields for cellulose is GLYCAM06, a biomolecular forcefield parameterised against data from quantum mechanical calculations.[128] It has many advantages, such as it can more accurately represent the ring conformations of cellulose than OPLS and GROMOS.[11] One example of its use is in 2016, when Poma *et al.* used GLYCAM-06 to create a coarse-grained model for the transformation of cellulose Iα to Iβ, which allows us to study how these two allomorphs can coexist, and build an understanding of how amorphous systems might be formed.[129]

However, a forcefield designed for carboyhydrates that is ideal for use on cellulose is not going to be suitable for nitrocellulose - the nitrate groups are not going to necessarily be parameterised.

One forcefield that has often been used in studies of nitrocellulose, generally in the context of propellants, is the COMPASS forcefield. A particular example is in 2008 when Ma *et al.* used Materials Studio to study nitrocellulose, nitroglycerine and systems that contained a mixture of these two and calculate detonation parameters.[121] Unfortunately, COMPASS is not currently implemented in LAMMPS, which this project uses, so was not able to be used for this study.

A forcefield that has been used for cellulose, and is also parameterised for nitrate groups is OPLS-AA, although molecules with nitrate groups such as nitrocellulose still have larger errors than molecules without.[130] Parameters from it have also been used for studied of nitrocellulose.[75] OPLS-AA has also been used successfully to study the breaking up of bundles of cellulose,[131] and its dissolution in solutions.[132] OPLS-AA has already been discussed in section 1.2.2 on page 9, and is the forcefield that was chosen to be used in this project.

TABLE 2.2: Densities of Cellulose
T/S/E means Theoretical (T) / Experimental (E) / Simulated (S)
[a] microfibril structures are less dense than regular bulk crystal structures, as they consist of a fibre of chains going through the simulation cell, and density is calculated over the total simulation cell, including the empty space around the fibre.
[b] Theoretical densities are densities calculated using the unit cell dimensions and the mass of the units present in the unit cell.

Ref.	Density value	T/E/S	Structure type
19	1.18 - 1.26 g cm^{-3} [a]	S	Microfibrils
93	1.501 g $cm^{-}3$	S	I_β
	1.656 g cm^{-3}	E	I_β
110	1.62 - 1.64 g cm^{-3}	E	I_β
	1.545 - 1.585 g cm^{-3}	E	Cotton(mix of structures)
122	1.582 g cm^{-3}	T [b]	I_α
	1.599 g cm^{-3}	T [b]	I_β
123	1.5861 g cm^{-3}	S	I_α
	1.5596 g cm^{-3}	E	I_α
	1.6125 g cm^{-3}	S	I_β
	1.6161 g cm^{-3}	E	I_β
124	1.3762 g cm^{-3}	S	Amorphous structures

TABLE 2.3: Densities of Nitrocellulose
T/S/E means Theoretical (T) / Experimental (E) / Simulated (S).

Ref.	Density value	T/E/S	Structure type	Other details
85	1.618 - 1.680 g cm^{-3}	E		0 - 13.41 wt%
75	1.600 g cm^{-3}	E		298 K and 100 kPa
	1.619 $\pm$ 0.01 g cm^{-3}	S	Crystalline, orthorhombic unit cell	298 K and 100 kPa. Custom force field, using parameters from GAFF and OPLS
92	1.48 g cm^{-3}	S	Unit cell created from 5_2 helix NC chains	Model created from x-ray data, no simulation run
125	1.58 g cm^{-3}	-	Partially crystalline	"Commonly accepted value" for NC cast from solution
	1.58 g cm^{-3}	E	Cotton	11.6 wt% to 13.1 wt% nitrogen content at 15-20 °C
126	1.67 g cm^{-3}	E		13.3 wt%
127	1.35-1.40 g cm^{-3}	E	Crystalline	20 °C

2.5 Previous Investigations of Decomposition of Nitrate Groups

There have been many previous studies into the decomposition of nitrate groups. However, most are experimental, with only a few looking at the reactions computationally using Density Functional Theory.

These studies are often on the -ONO_2 in the context of nitrate esters, as opposed to nitrocellulose, and when it is nitrocellulose, rarely on more than a monomer unit of nitrocellulose.[72,133]

One paper, which has already been discussed, that investigates the decomposition of nitrate esters experimentally is by Kimura.[119] Previous studies had reported activation energies for the homolytic cleavage of nitrate ester groups to be 171-196 kJ mol^{-1}. However, they predicted that the activation energy might be approximately 146 kJ mol^{-1}, which is the bond dissociation energy for a nitrate ester, which is much lower. They tested the reaction by measuring its luminescence, and recorded energies of 146 - 154 kJ mol^{-1}, which is closer to the expected energy, and also the energies shown earlier in figure 2.3.

In 2012 Shukla and Hill published several studies using DFT to study nitrocellulose. All the calculations they ran using Gaussian09. For the first two, geometry optimisations were performed on all the required structures using B3LYP and 6-311G(d,p) as the functional and basis set respectively. This was done in gas and in bulk water solution modelled using the Polarizable Continuum Model (PCM) approach, which uses a continuum which is polarizable, as opposed to the conductor-like continuum used in COSMO. Single point energies were then calculated for the optimized structures using the electron-correlated MP2 method using the cc-pVTZ basis set.

The first paper looked at the energies required to denitrate a monomer of nitrocellulose, by finding energies for structures with different combinations of nitrated positions, including intermediate states.[134] By comparing the energies in different combinations, it is possible to work out which sequence of denitrations is most likely to happen.

In the second paper, they looked at the difference between the energies required for successive denitrations of a monomer and how this is affected by changing the -OH on carbon 1 between the α and β position (in a chain of nitrocellulose, this -OH is replaced by the glycosidic bond and is in the β position).[109]

In the third paper, they compared the activation energies for monomers, dimers and trimers.[133] Here the geometry optimization calculations were run as before, but the single point calculations used the M06-2X meta-hybrid functional, as it was more

suited to the larger systems being tested here. They found that the activation energies for positions C3 and C6 were decreased, and for C2 was increased.

The results from all three of these studies are presented in table 2.4.

TABLE 2.4: Results from the three Shukla papers, using the numbering convention shown in figure 2.1 on page 54. [109,133,134]

Structure type	Most Favourable Nitration Order	Also Possible Order	Source
β-anomer	C3 $\rightarrow$ C2 $\rightarrow$ C6	C6 $\rightarrow$ C2 $\rightarrow$ C3	109
α-anomer	C6 $\rightarrow$ C2 $\rightarrow$ C3	C2 $\rightarrow$ C6 $\rightarrow$ C3	134
Dimer and Trimer	C3 $\rightarrow$ C6 $\rightarrow$ C2	-	133

An experimental study in 1991 by Hiskey *et al.* studied the thermal decomposition of many different nitrate esters, and the results showed that the secondary nitrates tested, cyclohexanol nitrate and 2-propanol nitrate, had higher activation energies - 180 kJ mol^{-1} and 191 kJ mol^{-1} respectively - than the primary nitrate energies (169 kJ mol^{-1} for 1-pentanol nitrate for example). [117] This is because the rate-determining step is usually homolytic cleavage of the bond, via S_N2 reaction (although the paper did find that tertiary nitrates undergo E1 reactions, meaning they have significantly lower activation energies). This supports some of the computational results in table 2.4, as position C6 is a primary nitrate, and C3 and C2 are secondary, which would imply C6 would be the most favourable to denitrate. In the third paper they found that in dimers and trimers as opposed to monomers, it is in fact C3 that is the most favourable to denitrate, but if you look at the structure in figure 2.1 you will see that the C3 position in the first (left hand) unit will be a lot less sterically hindered.

2.6 Summary

Cellulose and nitrocellulose are two chemicals vital to the modern world. Since the early 1800s scientists have been studying them, discovering new uses and properties, and over recent years increasingly this has been done with computational techniques as well as experimental. Most of these studies fall into one of two categories:

1. Studies into the large-scale crystal structure
2. Studies into chemical structure and reaction mechanics

The crystal structure of cellulose has very thoroughly been studied using both experimental and computational methods. Experimentally many distinctive crystal structures have been identified, with different properties depending on how the cellulose is packed. Computational studies of cellulose systems often focus on how working out how to model these crystalline structures accurately.

The crystal structure of nitrocellulose has been less comprehensively studied, with the majority of papers being based on experimental results, and only a few written about simulations of nitrocellulose. It has been found by Nardai and Bohn that nitrocellulose will become more disordered than cellulose when subjected to the same simulation conditions.[72]

Knowledge of the reaction mechanisms of nitrocellulose is built based on comparisons with reactions of smaller nitrate esters. Again, there is a lot of literature based on experimental studies, but there are a number of computational studies, using DFT. Trying to find the activation energy for the degradation of the nitrate ester functional group of nitrocellulose can be very important for understanding reaction kinetics, which has practical uses in terms of dealing with reactive highly nitrated nitrocellulose.

One study by Shukla and Hill concluded that to calculate accurate activation energies for the degradation of nitrocellulose the smallest molecule that is analogous is a trimer model.[133] However, there do not seem to be any papers that follow up on this conclusion and use DFT on larger structures of nitrocellulose.

Chapter 3

Constructing Model Structures of Paracrystalline Cellulose and Nitrocellulose

3.1 Introduction

In this chapter, a protocol for creating realistic structures of nitrocellulose using molecular dynamics will be presented. It particularly focuses on the crystallinity of the systems being created. It will also provide a detailed analysis of the geometric and dynamical parameters used to quantify the degree of crystallinity for the structures created here, with nitration levels varying from 0 - 14.14 wt% nitrogen content. Paracrystalline cellulose could not be created using the protocol designed here, however, it was found that the more nitrated a nitrocellulose system the more the structure tends to paracrystallinity. This is due to a decrease in the number of hydrogen bonds present, and an increase in the size of the functional groups pushing the chains apart, and weakening the interactions between the chains of the structure. The structures created are representative of realistic systems, which will in the future be able to be used to build further understanding of long-term storage of nitrocellulose.

In 2023 an article titled “Using molecular dynamics to simulate realistic structures of nitrocellulose of different nitration levels” was published in PCCP. This paper was based on work performed for this thesis, and this chapter expands on it.

3.2 Creation of Initial Structures

The input files included in Appendix A were created based on the unit cell provided from crystallographic data by Nishiyama *et al.* of cellulose Iβ.[136] Structures were then built using the chain, plane and block structure files, using Moltemplate. The periodic boundary conditions were applied by using partial units, and the bonds across the border were specified.

The unit structure in the appendix is a "master" unit. This means that it has both nitrated and non-nitrated functional groups for positions **a** through to **f**. These are denoted by comments including `Cell` or `NCell`. Once it has been decided whether a particular position is nitrated or not, the relevant lines are kept and unnecessary ones deleted- eg. if position **a** is to be nitrated, all lines ending "`# aCell`" should be deleted, and "`# aNCell`" be kept, and vice versa for when position **a** is to be left as a hydroxy-group.

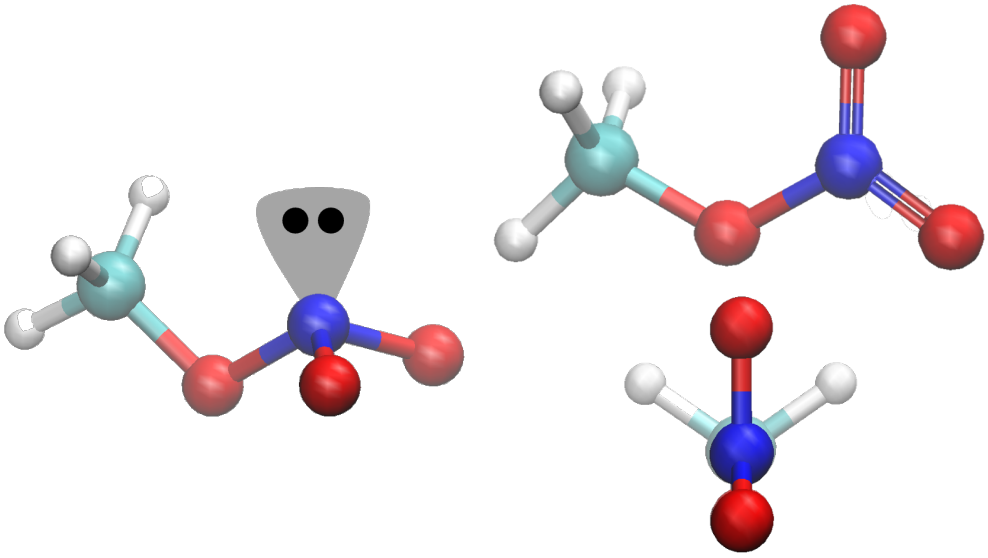
3.3 Choice and Verification of Force Field

Before running molecular dynamics simulations on nitrocellulose a suitable force field was needed.

Previously discussed in section 2.4.1, the OPLS (Optimized Potentials for Liquid Simulations) All-Atom force field was chosen for this research, as this is a well-defined force field within LAMMPS.[13] The bonding parameters of this force field are either based on the values in the AMBER-all atom force field, or found through comparison to ab initio molecular calculations using RHF/6-31G*//RHF/6-31G* on over 50 molecules, and the non-bonded parameters are based on Monte Carlo statistical mechanics simulations.[9] OPLS-AA has been used as a basis for studies on nitrocellulose.[12,75]

A parameter of importance for the accurate representation of the -ONO_2 group is the improper dihedral (figure 1.11 on page 20).[7] When tests were run with the OPLS-AA file provided in the GitHub repository of LAMMPS it was found that this was not being correctly defined, as the -ONO_2 group is a flattened trigonal pyramid shape, as shown in figure 3.1a. This means that before the force field could be used for nitrocellulose, the parameters to describe this had to be found.

The OPLS-AA parameters for these angles were found using LigParGen, a web server that can be used to generate OPLS-AA force field parameters for a structure inputted through SMILES or PDB, and added to the LAMMPS implementation of the force field. The atoms that needed parameterising were those in the -ONO_2 groups.[137] These were found by using LigParGen to give parameters for a molecule of HNO_3. These



(A) -ONO_2 group modelled on methyl nitrate with the original force field.

(B) -ONO_2 group modelled on methyl nitrate with the modified force field

FIGURE 3.1: 3.1a shows how the -ONO_2 group was arranged with the original force field as if there is a lone pair on the nitrogen. 3.1b is the -ONO_2 group after the force field was updated using parameters for HNO_3 from LigParGen.[137] The geometry is now planar, as there are double bonds.

parameters were added to the force field file, and are included in Appendix section A.1. Once these had been changed the -ONO_2 group behaved as shown in figure 3.1b, meaning that the issue had been corrected.

3.4 Decompression Protocol for Creating Paracrystalline Structures

Once the crystal structures had been created using Moltemplate, the structures need to be simulated at room temperature and pressure. The initial structures created are unnaturally dense, as atoms are added when nitrating, but the distance between the chains was not changed. This means that if the volume is not changed, the initial structure will be extremely high pressure, and will not cope with sudden extreme changes, so they must be changed gradually. The protocol is illustrated in figure 3.2 on the following page and discussed further in the following bullet points.

- An energy minimization was performed on the structure, using the conjugate gradient method
- The initial pressure was found. Due to the method used to make the nitrocellulose - adding nitrate groups into a cellulose structure without making a change to the size of the simulation cell - the density of the structures is going to

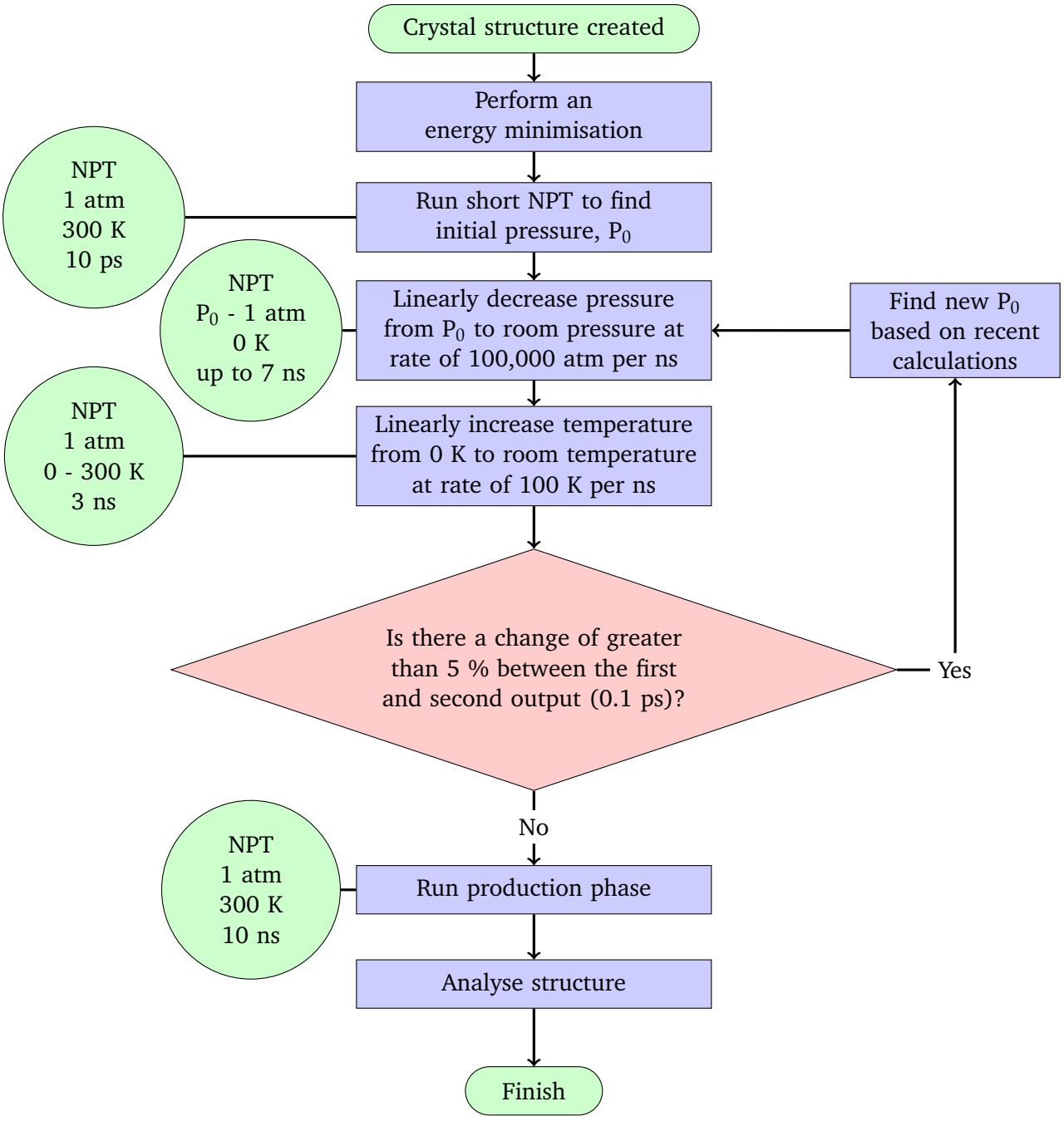


FIGURE 3.2: Flow chart demonstrating the protocol. Green circles show the details of calculations being run, and red diamonds represent points at which the data should be processed to analyse the progress of the structure and decisions made.

be unnaturally high, increasing the initial pressure calculated. These initial pressures vary based on the nitration level and are shown in table 3.1.

- The pressure was then brought down to 1 atm at a rate of 100,000 atm per ns using NPT simulations, which seems fast but was chosen to be compatible with all of the timescales needed, based on the initial pressures shown in table 3.1
- 10 ns of NPT were run on the structures, and then an analysis of the crystallinity was run.
- In certain cases, either where paracrystalline structures had not been formed, or further simulation was chosen to be performed, annealing cycles were run to see if further heating and cooling would help break the crystallinity.[123]

TABLE 3.1: Initial pressure values (P_0) found in the first step of the protocol shown in figure 3.2, and the corresponding time to be brought to room pressure at a rate of 100,000 atm per ns.

Structure	0N	1Na	1Nb	1Nc	1Nd	1Ne	1Nf
P_0 / atm	100	41,500	42,100	41,600	41,800	43,100	37,000
Time / ns	0.001	0.415	0.421	0.416	0.418	0.431	0.371

Structure	3Ncfa	3Ncfb	3Ncfd	3Ncfe	4/5N	6N
P_0 / atm	234,000	199,000	200,000	214,000	403,000	703,000
Time / ns	2.34	1.99	2.00	2.14	4.03	7.03

All LAMMPS calculations were run using 0.5 fs timestep, Nosé-Hoover thermostat and barostat, with the temperature relaxation time set to 100 timesteps (50 ps) and the pressure relaxation time set to 1000 timesteps (500 ps). Periodic boundary conditions were used in all 3 dimensions, to approximate a bulk system.

3.5 Effect of Different Nitration Levels and Combinations in Nitrocellulose

Several structures were created, with different levels and combinations of nitration. The 13 structures created are shown in figure 3.3 on the following page.

Before analysing how crystalline the structures produced are, it must be ensured that the simulation has equilibrated. Looking at the RMSD, every structure reached an equilibrium within the first 2 ns of the 10 ns production phase.

The RMSD was also used to validate that the periodic boundary conditions were being applied appropriately. Figure 3.4 on page 79 shows some RMSD graphs for the 4/5N structure during the 10 ns production phase. In particular, these graphs plot the RMSD

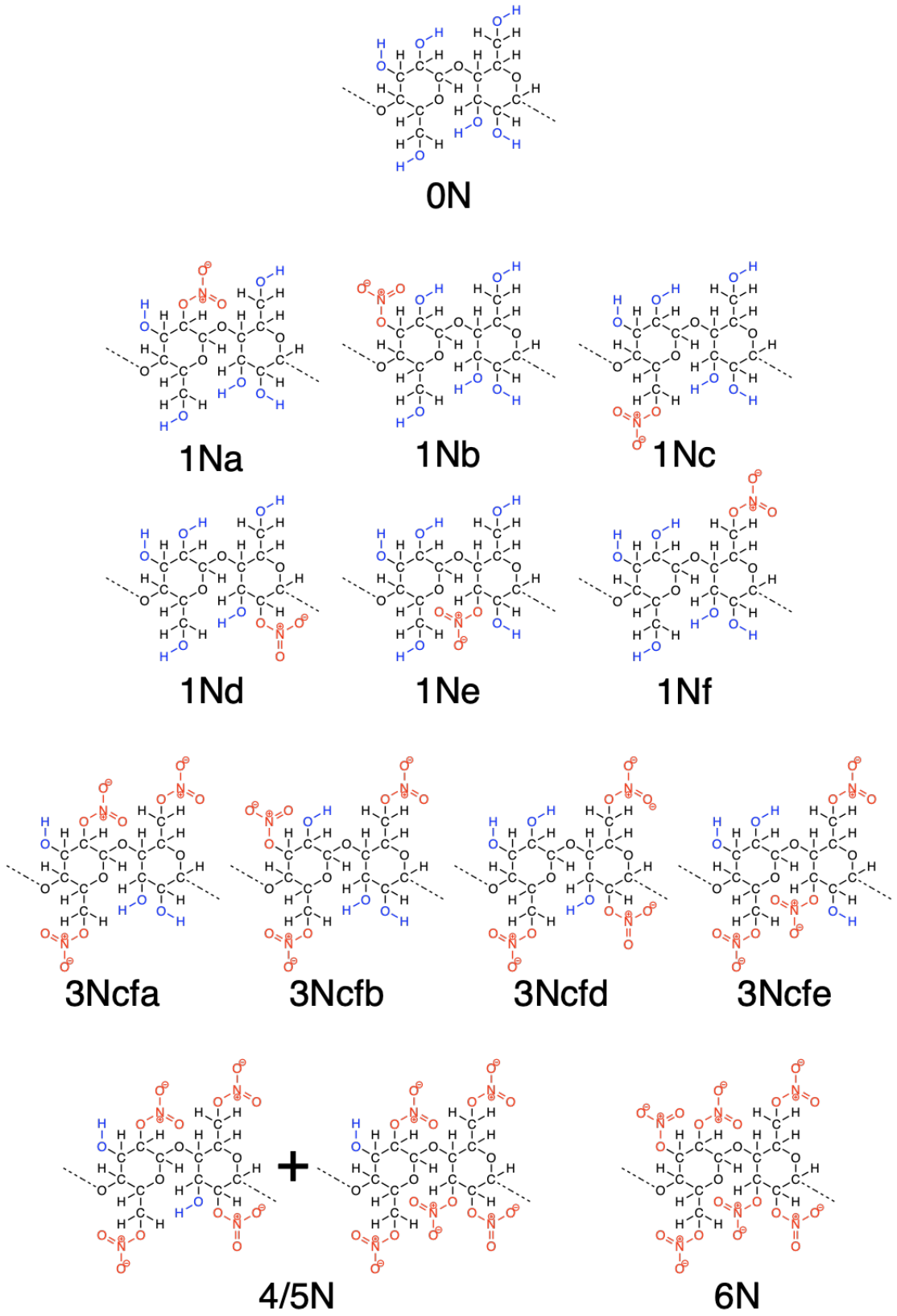


FIGURE 3.3: 13 structures were created. -OH groups are shown in blue and -ONO_2 groups are shown in red.

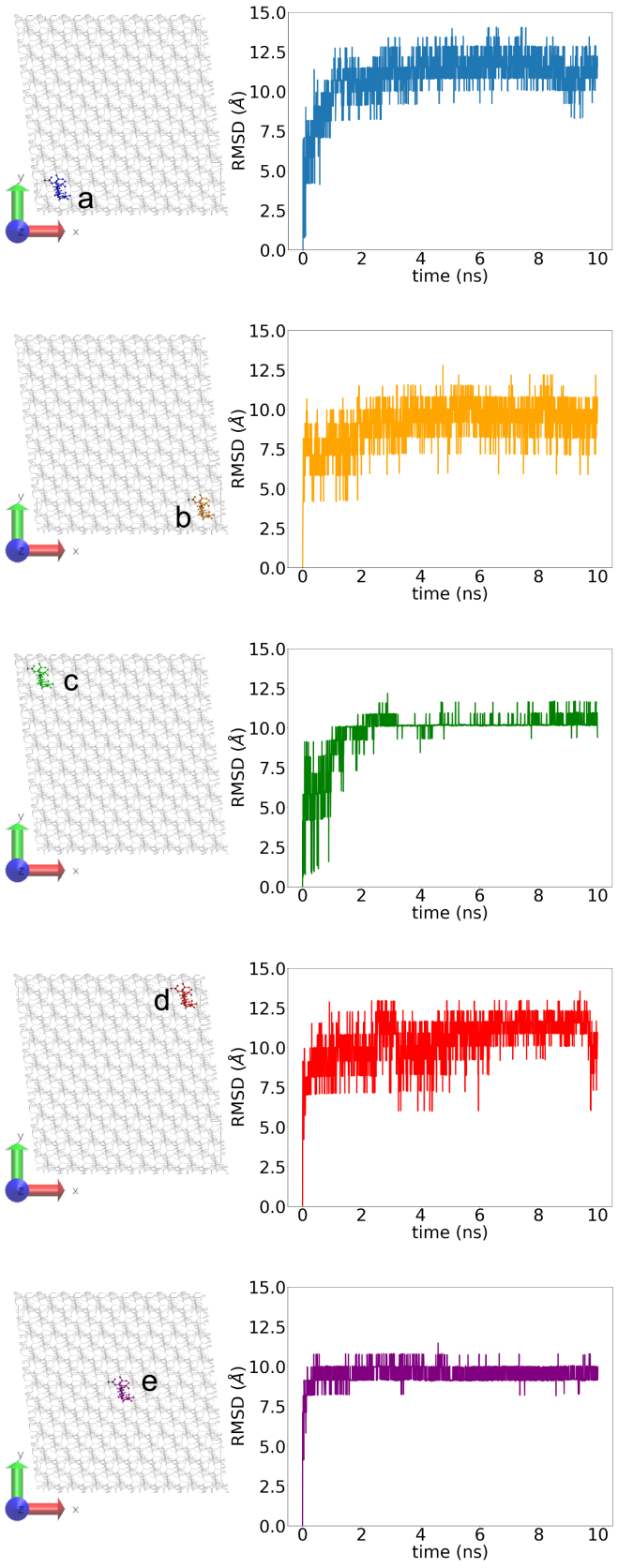


FIGURE 3.4: This figure shows the RMSD graphs for 5 separate chains in the 4/5N system. The relative positions in the initial structure are shown left, and the RMSD relative to their starting position on the right

for five specific chains within the structure. There is clearly less variation in structure e, which is in the middle of the block, implying that there is a difference between the treatment of the centre of the system as opposed to the edges where the periodic boundary conditions are applied. However, all of the RMSD shown here do level out at similar values, so the periodic boundary conditions have been applied correctly. All of the other structures showed similar results regardless of their nitration, meaning the periodic boundary conditions have been successfully applied.

3.5.1 Densities

After the simulations had been run as discussed on the 13 different structures described above, the average densities were collected in figure 3.5.

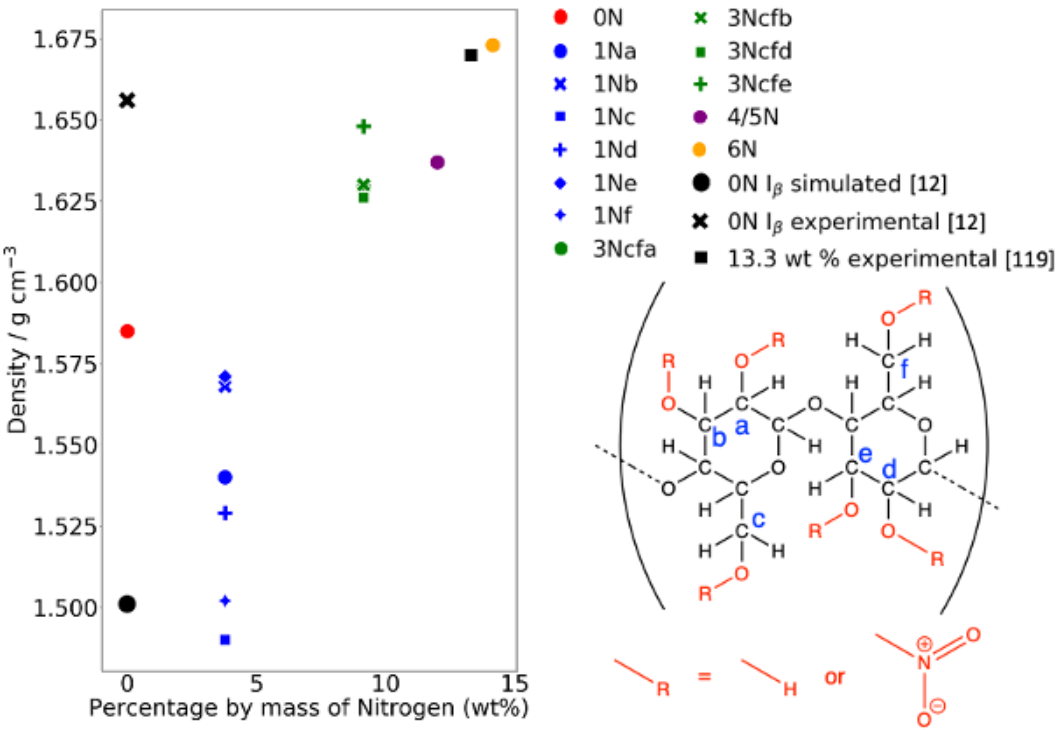


FIGURE 3.5: Average densities of all of the structures generated calculated over the full 10 ns of the NPT production phase plotted relative to their nitrogen content. Also shown in black are three literature values with specific nitrogen content values for comparison. All of the structures simulated are within a realistic range of densities and follow the general trend stated in reference [85] that as the nitration level increases, so does the density.[85] The simulated density value quoted here was found by simulations run with GROMACS and a different force field and protocol, which is why it has a different density from the structure simulated in this work.[93]

Measuring the density of the structure does not give us information about what is going on within the cell in terms of crystallinity. However, as density is a property that is relatively easy to calculate for experimental systems and simulated systems alike, it

can be used to support the validity of the simulated models, in comparison to experimental data.

The density of the cellulose simulated here can be seen to be somewhere between that of experimental cellulose and cellulose simulated in previous work.[93] The densities for the other structures follow the expected trend of increasing density as the nitration level increases, as is expected based on the literature.[85] The fully nitrated structure (14.14 wt%) fits fairly well with the density given in reference [126] for 13.3 wt%.[126]

For the 1N structures, the density values show the densities have approximately paired up, with the positions at corresponding positions on the two rings of the dimer giving similar densities.

In the simulation of the 3Ncfe, there was a phase change during the 10 ns production phase, from paracrystalline to crystalline. This is discussed later in the chapter in section 3.5.5 on page 91.

3.5.2 Lattice parameters

Another method of measuring the phase change of a simulated model, used by Kulasinski *et al.* in 2014 was to plot the lattice parameters and show how they change with time, proving a phase change.[93]

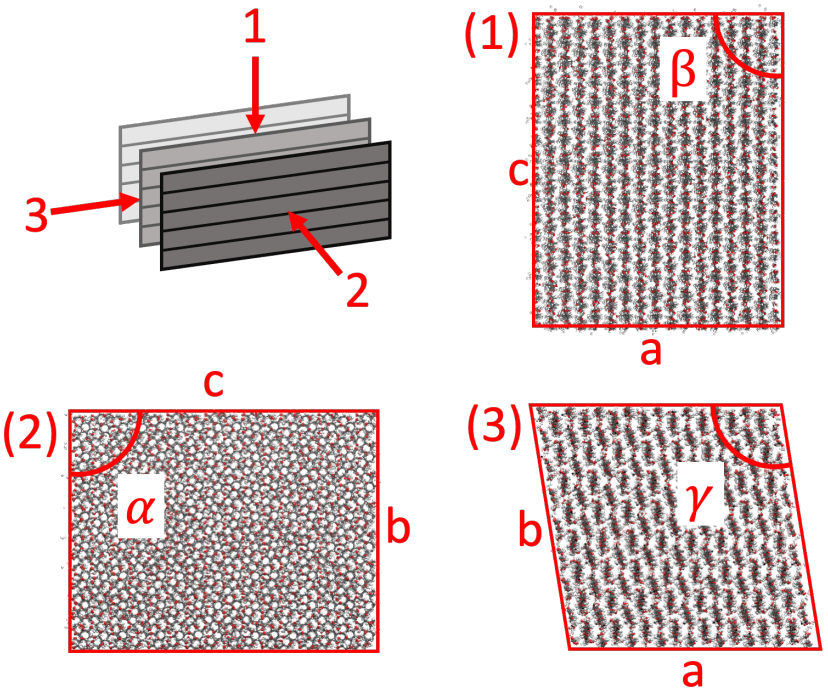
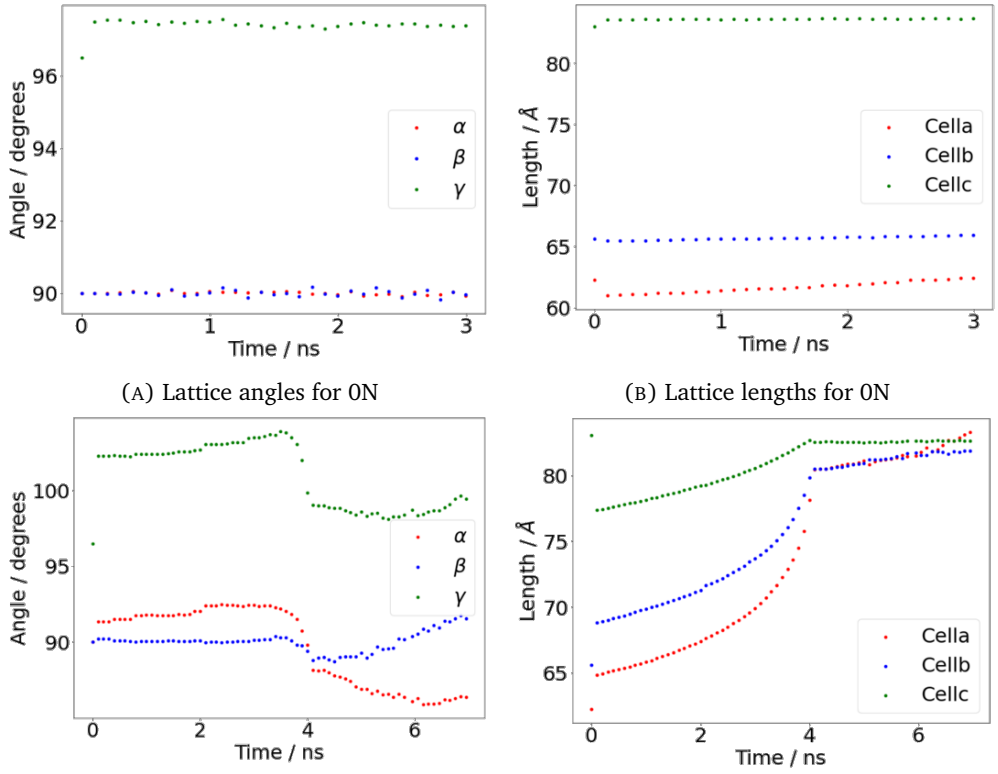


FIGURE 3.6: The lattice parameters, demonstrated on the simulation cell of cellulose.



(A) Lattice angles for 0N

(B) Lattice lengths for 0N

(C) Lattice angles for 4/5N

(D) Lattice lengths for 4/5N

FIGURE 3.7: Graphs showing how the lattice parameters (defined as shown in figure 3.6) during the simulation for 0N (cellulose) and 4/5N (nitrocellulose). It emphasises the fact that while the cellulose stays mostly the same during the equilibration, there is a clear phase change in the nitrocellulose structure.

Figure 3.7 shows how these lattice parameters change over the course of the equilibration process. It can easily be seen that the structure of cellulose is not changing much, whereas the 4/5N nitrocellulose has a very clear phase change at approximately 4 ns.

Normally the density also can be compared in this way. However, due to the artificially high initial densities caused by the method chosen for the creation of the initial structures, the density cannot be used in this case.

3.5.3 Visualisations and RDF

For every structure, the RDF was calculated. As discussed in the introduction, section 1.2.8.5 on page 20, calculating the RDF produces a characteristic graph based on the relative positions of the atoms. This means an RDF can give us information about the crystallinity of a structure. Figure 3.8 on the next page shows an RDF calculated between the carbon positions represented by the green stars in figure 2.1 on page 54. This carbon is chosen as it is part of the glycosidic bond, meaning it can give important information about the shape of the chains.

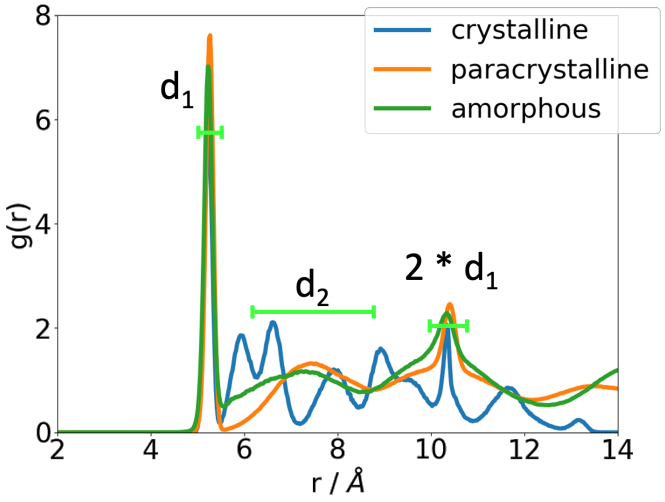
The structure used for the crystalline RDF is the initial cellulose structure, created from unit cells of cellulose Iβ, meaning that it is the most crystalline. The relative distances between all of the atoms are at discrete values, as the structure is strictly ordered, and so there are several very clear peaks. As the structure becomes paracrystalline most of these peaks begin to smooth out, as the positions average out.

The other two lines are plotted from nitrocellulose structures created in this project, the 'paracrystalline' line the RDF of 6N nitrocellulose after the protocol had been performed on it, and the 'amorphous' line is for a cellulose structure when it is heated to 1000 K, at which point the crystallinity was completely broken. The threshold between paracrystalline and amorphous is not clearly defined. For this project, it was thought that the periodic boundary conditions prevented the structures from becoming fully amorphous, as the chain still needs to go in one side and out the other of the simulation cell, so there is a limited amount of twisting and combining is allowed. The cellulose heated to 1000 K is an exception to this, as the simulation cell contracted sufficiently that a lot more flexibility in the position of the chains was allowed, making it on the border between this definition of paracrystalline/amorphous. Cellulose heated to this great a temperature was the only time that this was seen.

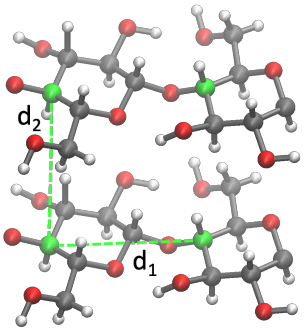
There are still two very clear peaks in the paracrystalline RDF shown in figure 3.8a on the next page, at around 5 Å, and 10 Å. This is to be expected, as they represent the distances along the chain, which are not going to be subject to many variations as the periodic boundary conditions will prevent the folding of chains.

This means that the "smoother" the RDF graph, the more paracrystalline a structure is. The smoothness can be considered as a combination of two factors - a smooth line will have fewer, wider peaks. Figure 3.9 on the following page shows these two measures for each of the structures.

The cellulose structure was still crystalline, even after the protocol had been performed on it, as can be seen in the visualisations in figure 3.10 on page 85. The RDF shown in figure 3.10c on page 85 has 11 narrow peaks with a prominence of $g(\mathbf{r}) > 0.05$, making it the most crystalline of any of the structures (figure 3.9). Further testing was



(A) RDF plot



(B) Cellulose model with carbon atoms of interest highlighted in green

FIGURE 3.8: 3.8a shows the RDF between the fourth carbon of each monomer within cellulose structures. The d labels here show regions of peaks representing specific distances, illustrated in 3.8b, with the d_1 and d_2 demonstrated on two dimer sections of cellulose. The carbons between which the RDF is being measured are shown in green (illustrated by a green * on the structure shown in figure 2.1). Other carbon atoms are shown in grey, oxygen atoms in red and hydrogen atoms in white.

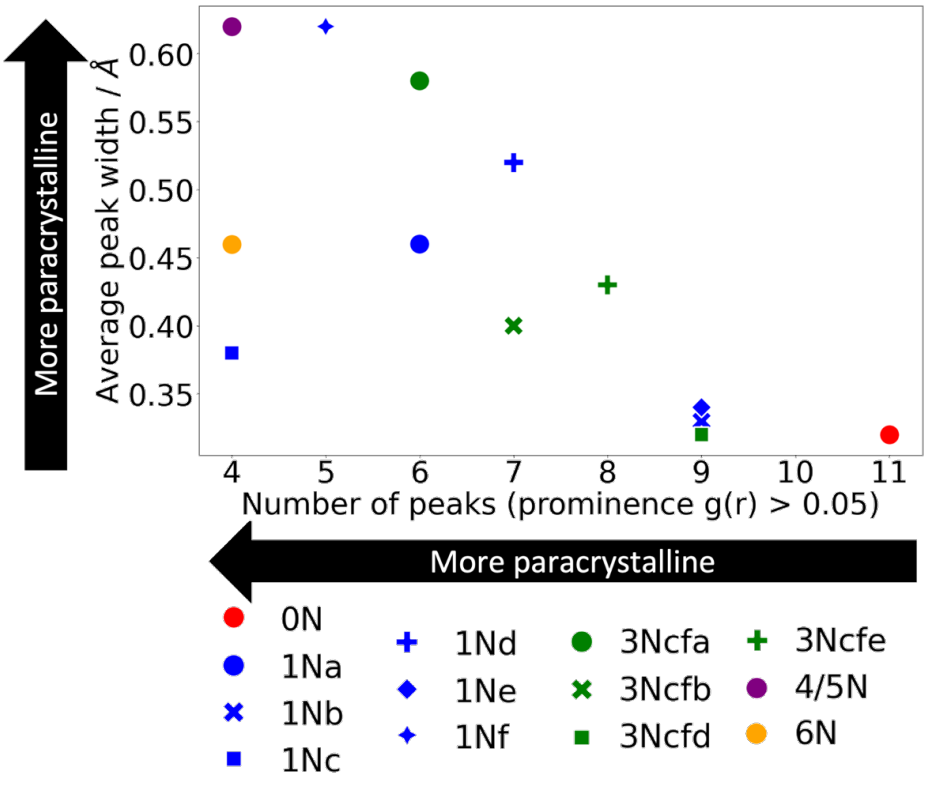
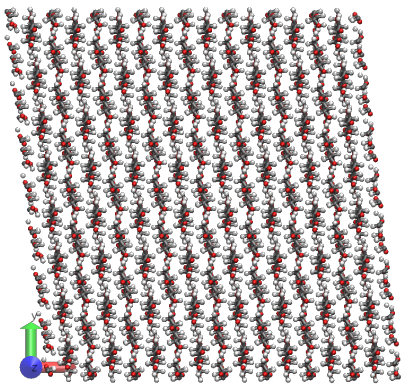
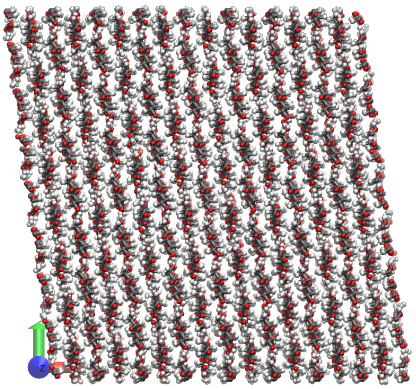


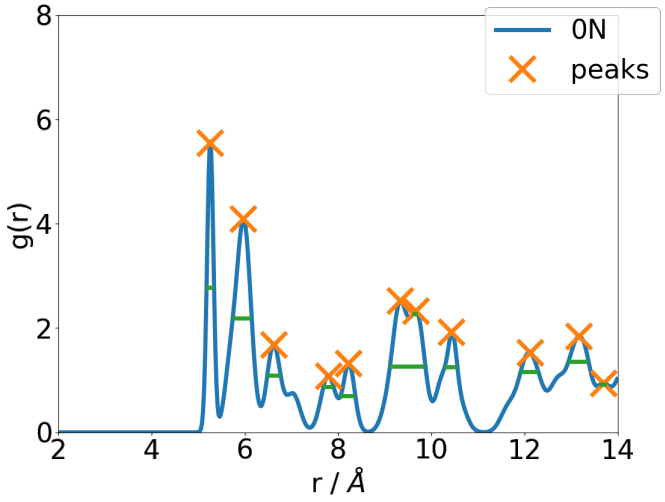
FIGURE 3.9: The number of peaks in the RDF over the 10 ns production phase, at over g(**r**) = 0.05 intensity. The widths of the peaks are measured from halfway up the peak prominence.



(A) Initial crystal structure of cellulose.



(B) Final crystal structure of cellulose, after 10 ns of production.



(C) RDF over 10 ns of production, with peaks of intensity g(**r**) $<$ 0.05 marked, and their widths at half the prominence of the peak.

FIGURE 3.10: Cellulose remains crystalline after the protocol has been run

run on the cellulose to see if a paracrystalline structure could be created, described in section 3.6 on page 93.

Figure 3.11 on page 87 shows the RDF graphs for all 6 singly nitrated structures and the visualisations of some of the structures created, and the RDF over the 10 ns NPT production phase. Looking at the RDF graphs in figure 3.11 on page 87 it can be seen that the position of the single nitrogen can affect the crystallinity of the structure, and all of the structures created are at least somewhat paracrystalline.

Here, as with the densities, a slight pairing based on the nitration position can be seen, especially in the visualisations. It can be seen that 1Nb and 1Ne look the most compact and crystalline, and in figure 3.9 they are further right and lower than the other 1N

structures, 1Na and 1Nd are less regular, and 1Nc and 1Nf are the most irregular and paracrystalline, closer to the top left of figure 3.9 on page 84, with much more twisting of the chains. The fact that 1Nc and 1Nf are the two most paracrystalline structures supports the suggestion that the reason these structures lose crystallinity during the simulation time, unlike cellulose, is that the added nitrate groups are affecting the inter-chain interactions - the off-ring nitrate positions can disrupt this to a greater extent.

The triple-nitrated structures were created based on the observations made from the single-nitrated structures, with the two off-ring positions, c and f, nitrated, as they would have the greatest chance of producing paracrystalline structures. Four structures were made, each one with a third position nitrated, one of the four remaining positions.

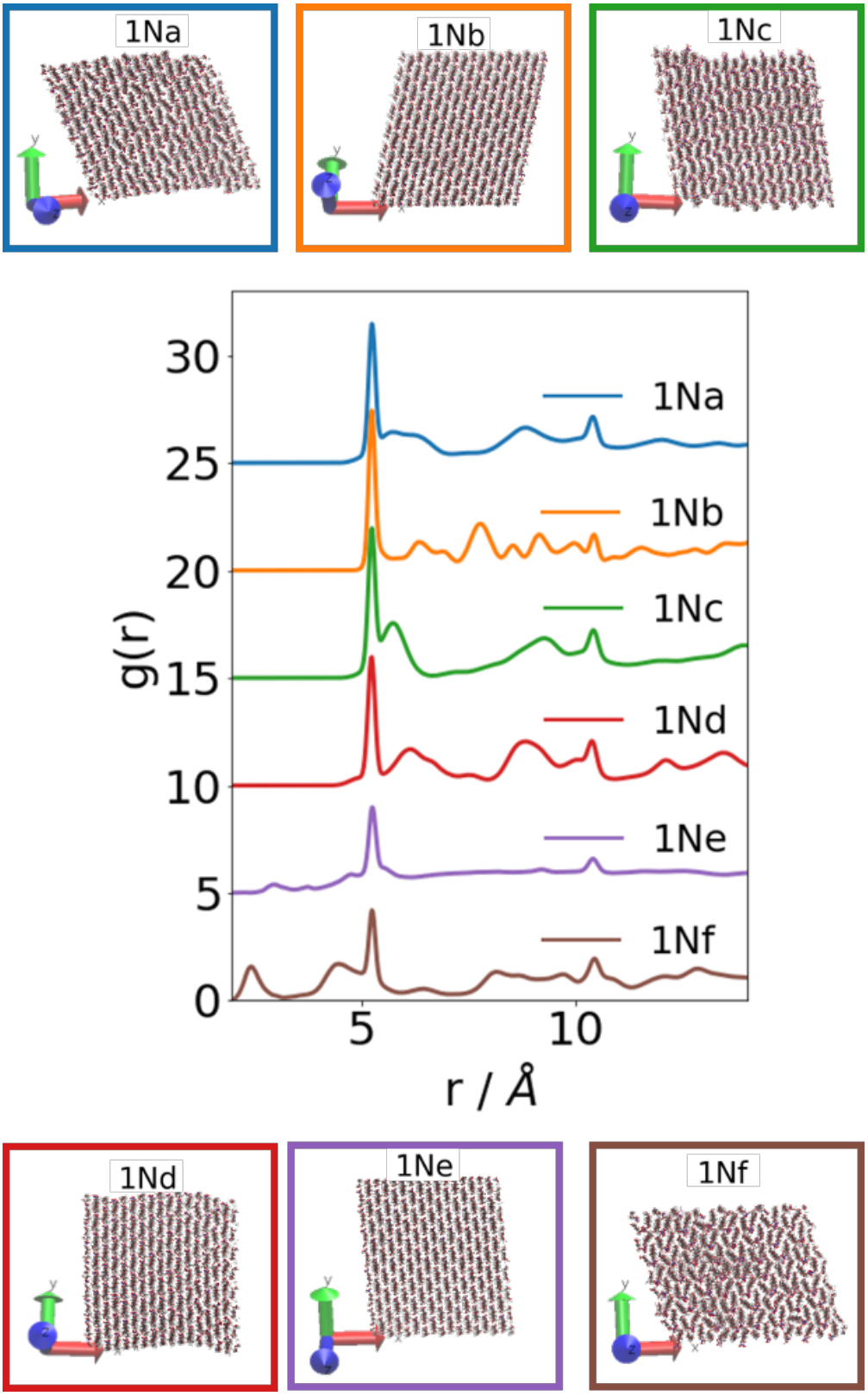


FIGURE 3.11: RDF graph for all 1N single-nitrated structures, with successive curves offset by 5 units along the y-axis, and the corresponding structures, rotated for the best view from the end on with the chains going into the plane of the page.

All of the RDF graphs and visualisations (figure 3.12) for the triple-nitrated structures exhibited less paracrystallinity than expected, compared to the single-nitrated structures. Particularly given that 1Nc and 1Nf both were paracrystalline, a structure containing both of these positions nitrated would be expected to be more paracrystalline.

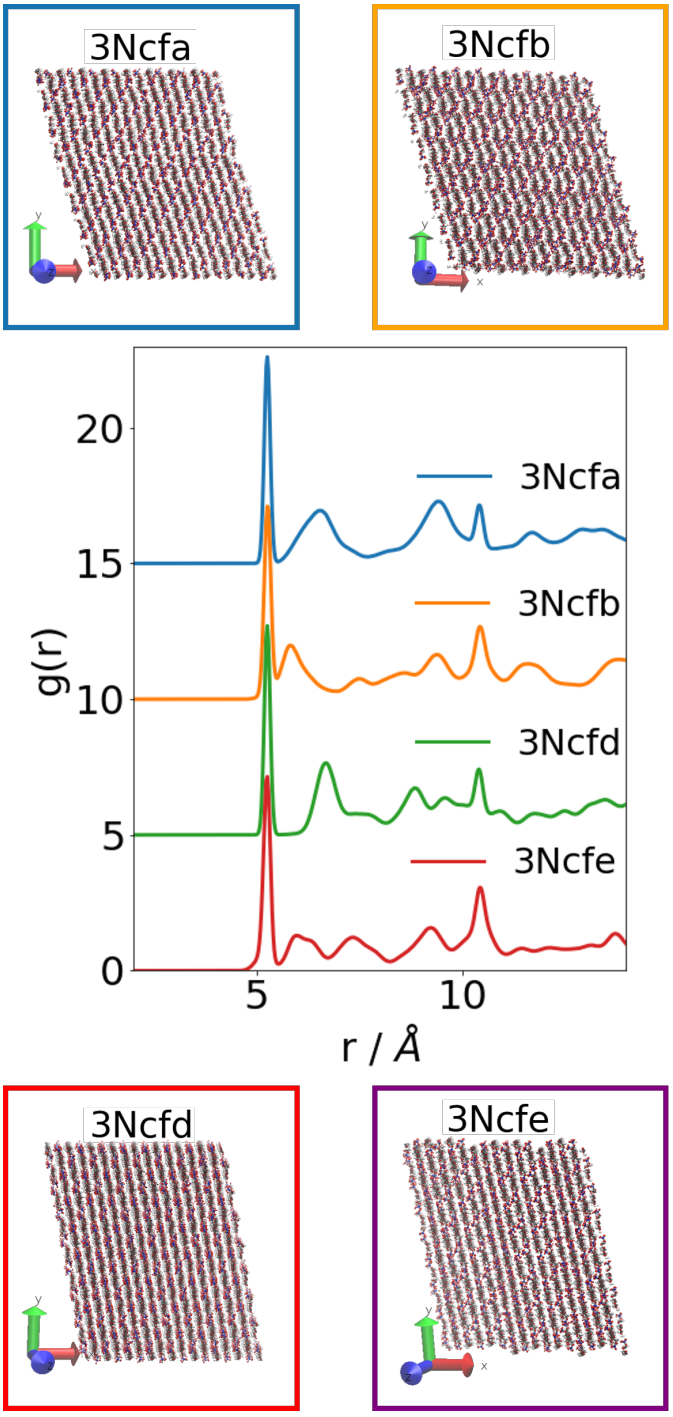
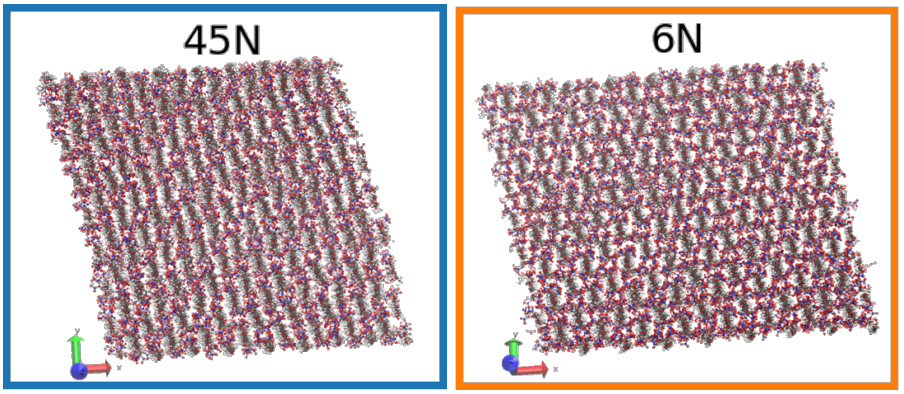
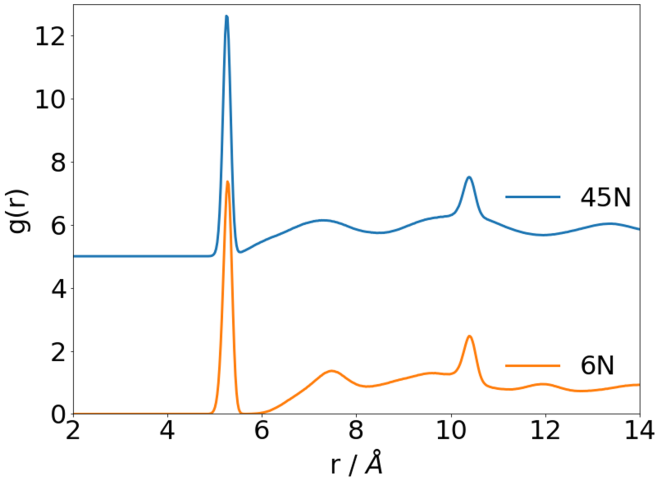


FIGURE 3.12: RDF of structures for four of the triple-nitrated structures. Successive curves are offset by 5 units along the y-axis. The RDF show that the structures are a little more crystalline than some of the 1N structures, although the visualisations still seem quite ordered.



(A) RDF plot for 4/5N (B) RDF plot for 6N



(C) Cellulose model with carbon atoms of interest highlighted in green

FIGURE 3.13: RDF (with 0.05 peaks marked) and visualisations for the structures obtained during the 10 ns production phase of simulations of 4/5N (3.13a and 3.13c) and 6N (3.13b and 3.13c), showing that paracrystalline structures have been obtained.

The 12 wt% structure of nitrocellulose was made through alternating dimers that are 4 nitrated (at positions acdf) and 5 nitrated (at positions acdef), and the 6N structure was created by nitrating every possible position. Looking at the visualisations and RDF in figure 3.13 both of the structures created for these nitration levels are paracrystalline.

Another potential way in which the RDF data could be used to try to quantify the crystallinity of the structure is to study the position of the second peak. In figure 3.8a on page 84 the area labelled (d_2) represents the distances between the chains, so when comparing structures of the same level of nitration, the ***r*** of the second peak can give information on how far apart the chains are. This cannot be used for comparison between differently nitrated structures, as the steric hindrance will also have a large effect on this.[19]

The intensity of the second peaks can also give a little information on the crystallinity

of the structure - in figure 3.10c on page 85, the second peak of the crystalline cellulose is a lot higher ($g(\mathbf{r})=4.09$) than the second peak of the 4/5N structure ($g(\mathbf{r})=1.14$) in figure 3.13a, where the distance between the chains is much less discrete.

In general, therefore, it was found that the structures produced by the protocol were to some extent paracrystalline for all the structures except for cellulose. The most paracrystalline structures were the more highly nitrated structures, 4/5N and 6N. By comparing the single-nitrated structures, the structures pair up again as the densities do in figure 3.5. The positions that created the most paracrystalline structures were the off-ring positions, which suggests that the nitrate groups are getting in between the chains and causing them to break.

3.5.4 Hydrogen bonds and other analysis

When considering the hydrogen bonding of a system it becomes more important to consider the positions that have not been nitrated, as the -OH groups are where the hydrogen bonds form. It is impractical to compare the hydrogen bonding of structures with different levels of nitration. The number of hydroxyl oxygen atoms affects both the number of donors and the number of acceptors, so the number of bonds cannot be scaled proportionally to the number of -OH groups.

However, structures of the same nitration levels can be compared. For example, figure 3.14 on the next page shows the number of hydrogen bonds in the single-nitrated structures. The relative number of hydrogen bonds show a similar trend to the densities and RDF graphs - assuming the more crystalline the structure, the more hydrogen bonds it contains, the structures nitrated at position c and f are the most paracrystalline, and positions b and e the least.

Over the other structures, the trend is that the more nitrated the structure, the fewer hydrogen bonds there are, as expected. This does suggest another reason why the nitrated structures break crystallinity more easily than the cellulose - not only do the nitrate groups require more space, so push the chains apart more, but there are fewer hydrogen bonds able to form, meaning that there is less rigidity.

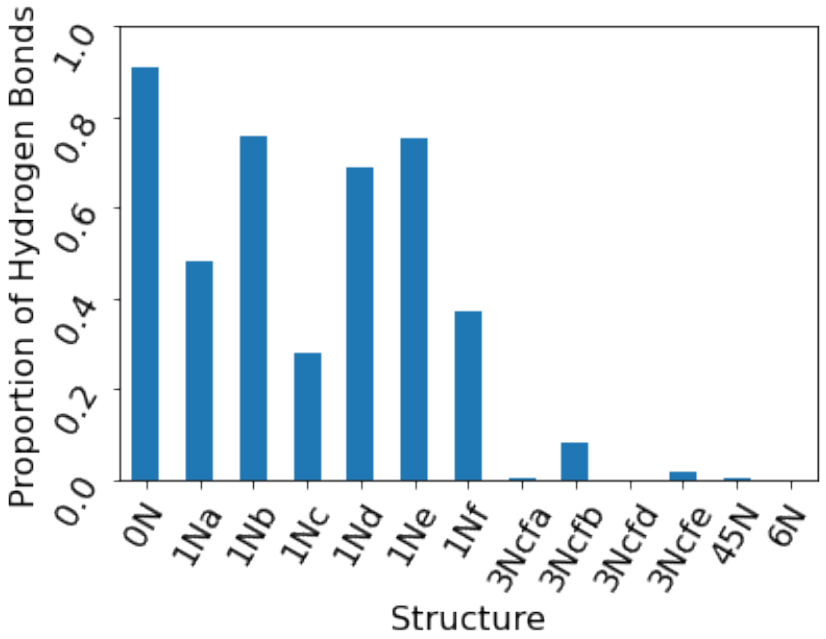


FIGURE 3.14: Average number of hydrogen bonds in each structure over the 10 ns production phase, calculated with a 3 Å cutoff for the donor-acceptor distance, and 150° angle for the donor-hydrogen-acceptor. Scaled to be proportional to the number of OH groups present, to allow for comparison between structures of different nitration levels.

3.5.5 Investigation of the 3Ncfe structure

Looking at the average densities plotted in figure 3.5 on page 80, it can be seen that 3Ncfe had an outlying average density. When plotting the densities of the structures as the simulation progressed (figure 3.15a) it became clear that something different was happening to the 3Ncfe structure during the production phase - there was a phase change at 5.6 ns.

This phase change was confirmed by comparing the RDF graphs before the change of density at 5.6 ns, figure 3.15b shows a significant change in the crystallinity, and the visualisations shown in figure 3.15c and 3.15d support that the structure is becoming more crystalline.

It was interesting that this happened for only one structure. A few tests were able to be run on the 3Ncfe structure to see if changing the initial structure slightly could make a difference. Five tests were run, each created by shifting the initial position of one of the hydrogens of the unit cell. These are shown in table 3.2.

Of the five tests, two of the simulations failed. The error in both cases was an issue caused by atoms behaving unnaturally, and either moving too far, into a space that is being calculated by a different computer processor, or being too close to one another. The settings were chosen to minimise this issue - the list of neighbouring atoms, which

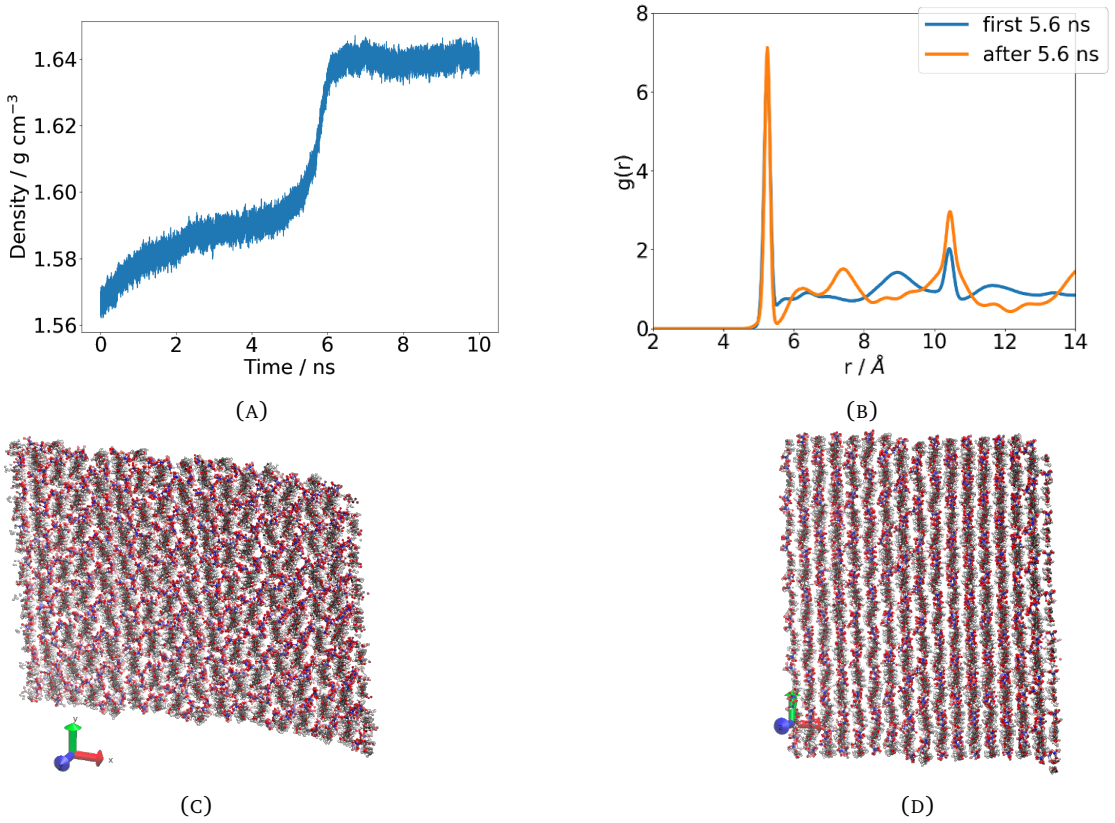


FIGURE 3.15: When the simulation of 3Ncfe structure was run, the density during the 10 ns production phase was found to take a major shift at around 5.6 ns (3.15a). When the RDF of the structure before and after this shift are plotted (3.15b), it can be seen that there is a significant difference in the crystallinity of the structure. This can also be seen visually (3.15c), (3.15d). Further tests were run on this structure, discussed in table 3.2

is used to assign atoms to computer processors, is being rewritten at every step, so the atoms have less time to move out of range. This means that in the two calculations that failed the problem is the initial geometry, which has been changed for the purpose of these tests.

Of the remaining three, all showed a slight change of density, but this did not cause a significant change to the crystallinity, according to the RDF and visualisations. This implies that the phase change is caused by the exact geometry that was used to create the structure.

TABLE 3.2: Results of tests run on 3Ncfe. Every shift is (0.001, 0.0, 0.0) and coordinate units are Å.

Test	**Original**	**Test 1**	**Test 2**	**Test 3**	**Test 4**	**Test 5**
Structure	Original geometry	Shift **1st** H of each unit	Shift **2nd** H of each unit	Shift **3rd** H of each unit	Shift **4th** H of each unit	Shift **5th** H of each unit
Ran to completion?	Yes	No	Yes	No	Yes	Yes
Density Change?	Yes	X	Yes	X	Yes	Yes
RDF Change?	Yes	X	No	X	No	No
Visual Change?	Yes	X	No	X	No	No

Annealing tests, as described in the next section, could have been run on this structure to see if paracrystalline structures of 3Ncfe could be produced. However, the main reason for the further exploration here was to try and understand **why** this structure was behaving differently from the other triple-nitrated structures. The final conclusion was that it seems to be the precise initial structure that causes the shift change, and a very small change in the initial structure prevents it. Further investigation, including annealing cycles, could be pursued into this.

3.6 Further Annealing Testing

The cellulose structure was further tested with annealing cycles. An extension to the protocol in figure 3.2 on page 76, is now shown in figure 3.16 on the following page. Annealing cycles were run for cellulose, as paracrystalline cellulose had not been produced, and on the 4/5N nitrocellulose, to test if the paracrystallinity can be increased.

By looking at RDF graphs from the first annealing cycle of 4/5N, when the system has been heated up to 700 K the system has become even more paracrystalline, arguably to the point of being amorphous, shown by the blue line in figure 3.17 on page 95. However, in the next part of the annealing cycle, the temperature is lowered again to

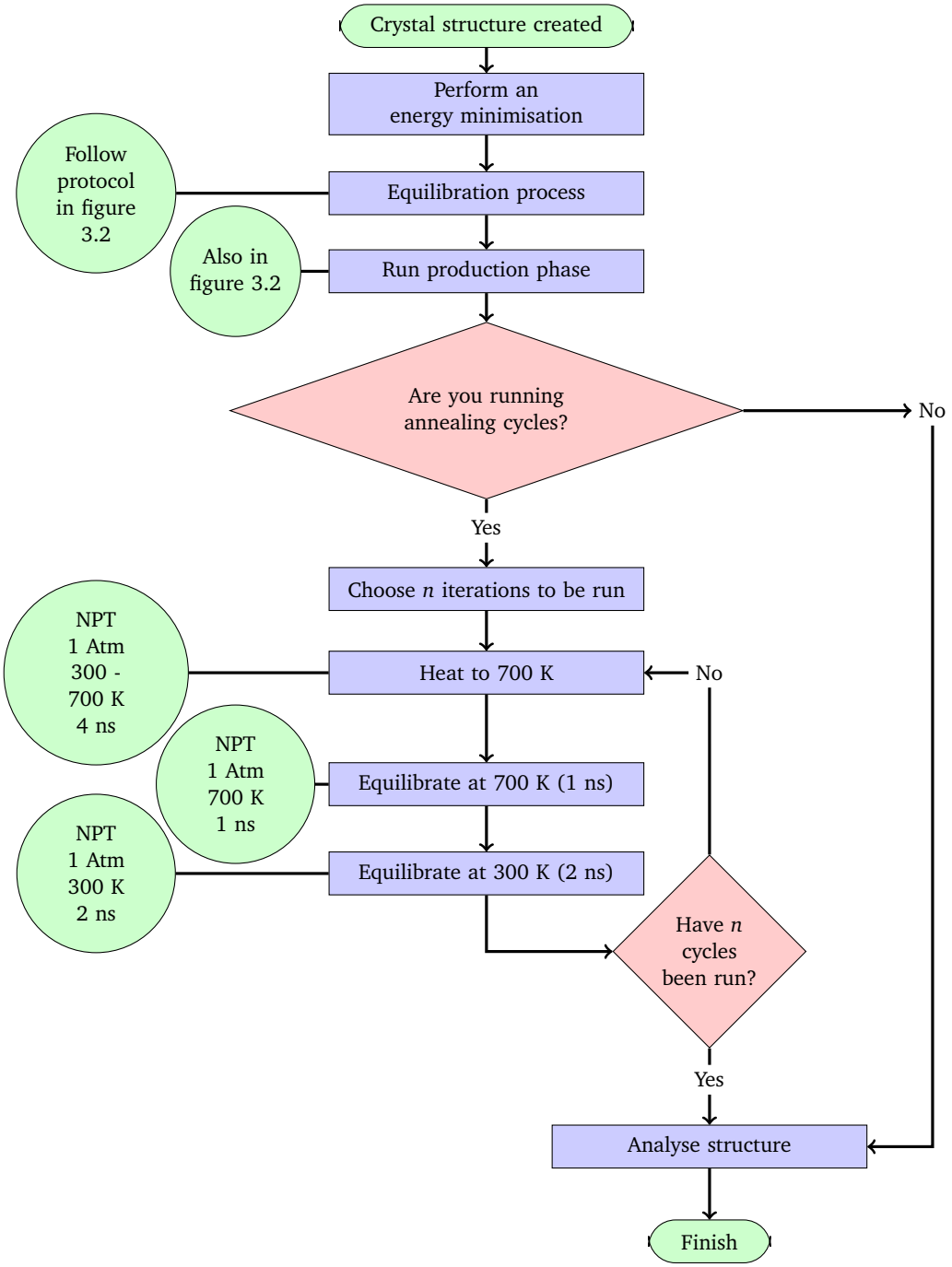


FIGURE 3.16: Flow chart demonstrating the protocol. Green circles show the details of calculations being run, and red diamonds represent points at which the data should be processed to analyse the progress of the structure and decisions made.

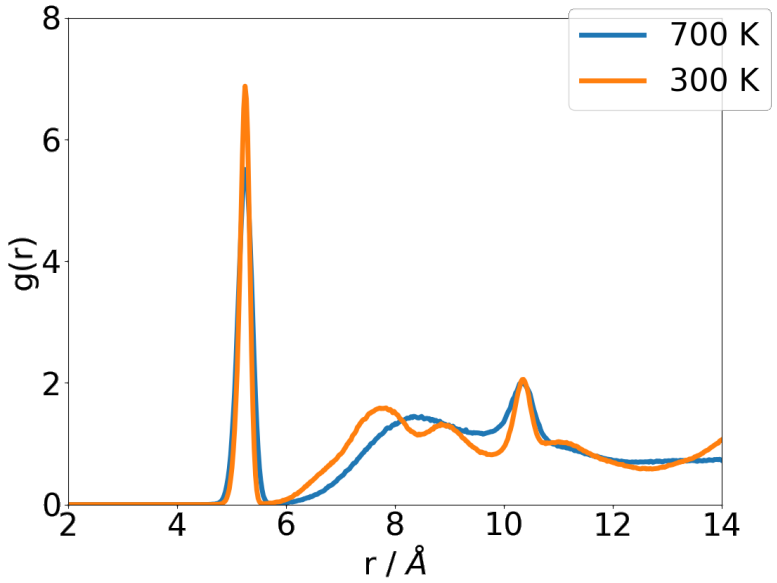


FIGURE 3.17: RDF for the 4/5N nitrocellulose during 1 ns at 700 K and during 1 ns at 300 K, during the first cycle.

300 K, and some crystallinity is restored, as seen in the orange line. This was found to be the case for the four annealing cycles that were run.

Eight annealing cycles were run on cellulose, shown in figure 3.18 on the following page. It shows how the temperature is being changed, and how this affects the density of the structure. Four RDF graphs are shown - the first RDF plot (i) is for the cellulose structure during the production phase, previously shown in figure 3.10b on page 85, which is very crystalline. RDF plot (ii) shows a more paracrystalline structure when the temperature has been brought up to 700 K. Unfortunately when the temperature is immediately brought to 300 K after this, shown in RDF (iv), the structure returns to being very crystalline (only slightly more paracrystalline than the initial structure). Finally comparing RDF from the 300 K periods from the first cycle (iv) and the last (iii), these two overlay exactly, showing that running further annealing cycles is not improving the paracrystallinity.

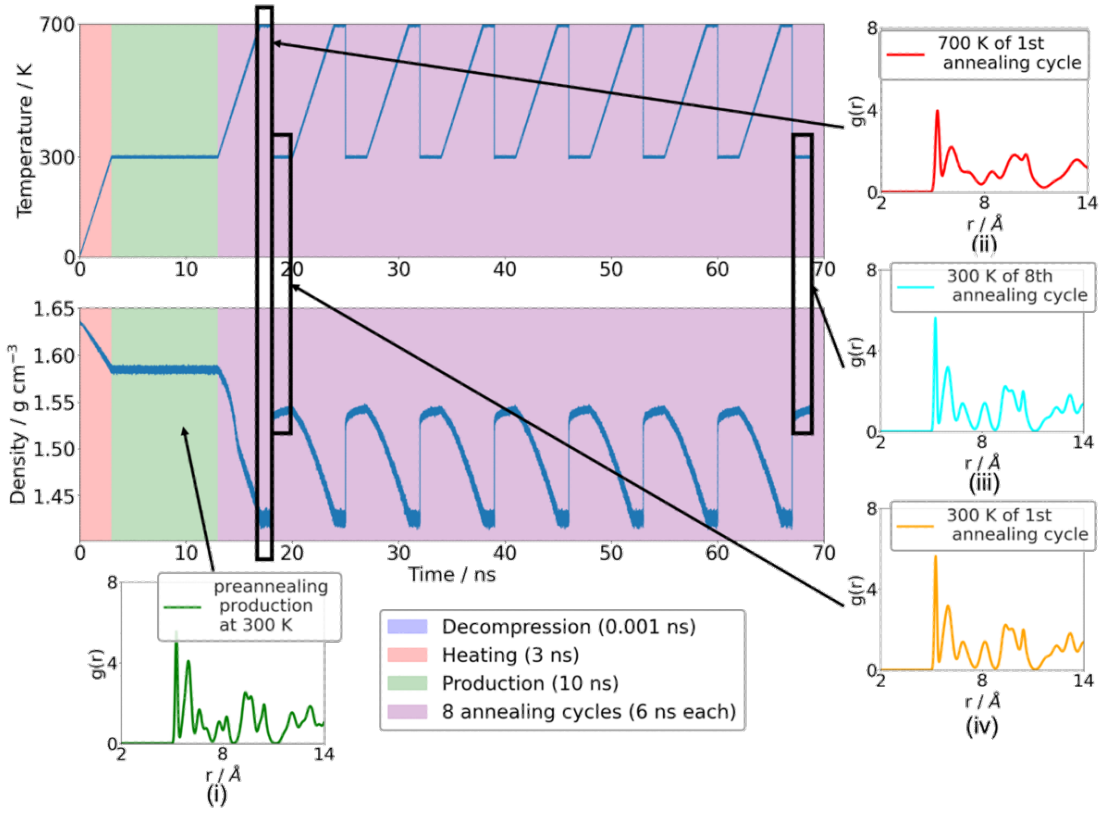


FIGURE 3.18: During the entire simulation, the pressure is kept constant and the temperature is varied as shown on the upper line of the plot. The change in density is inversely proportional to the change in temperature.

3.7 Conclusions

In this chapter a protocol for creating systems that can be used to model paracrystalline nitrocellulose of varying degrees of nitration, to reflect realistic systems, has been presented.

It was unable to create paracrystalline cellulose, even after using annealing cycles - suggesting that the addition of nitrate groups helps to break the crystallinity.

The single and triply-nitrated structures created became more paracrystalline than the initial crystal structure, although the effect of adding more groups was not as great as expected - the triple-nitrated structures were a similar level of paracrystallinity to the single-nitrated. It was found that the nitration of the off-ring positions of the nitrocellulose caused a more significant difference than the on-ring positions, further supporting the suggestion that interference of the nitrate groups with the interchain interactions is what causes this ability to become paracrystalline.

However, by following the protocol developed here, it was possible to create fully paracrystalline structures for the 12 wt% and fully nitrated nitrocellulose. Although fully nitrated nitrocellulose is too reactive to be created experimentally, the 12 wt% is

representative of the nitrocellulose that is often used in propellants and explosives and therefore could be used to study other properties of the substance, such as viscosities, and effects of solvents.

There are two reasons that increasing the number of nitrate groups may increase the paracrystallinity - firstly the nitrate groups are sterically bulky, causing more space to be required between the chains. Secondly, adding nitrate groups removes hydrogen bonds, which are responsible for maintaining the crystallinity, as was confirmed with a detailed analysis of hydrogen bonding.

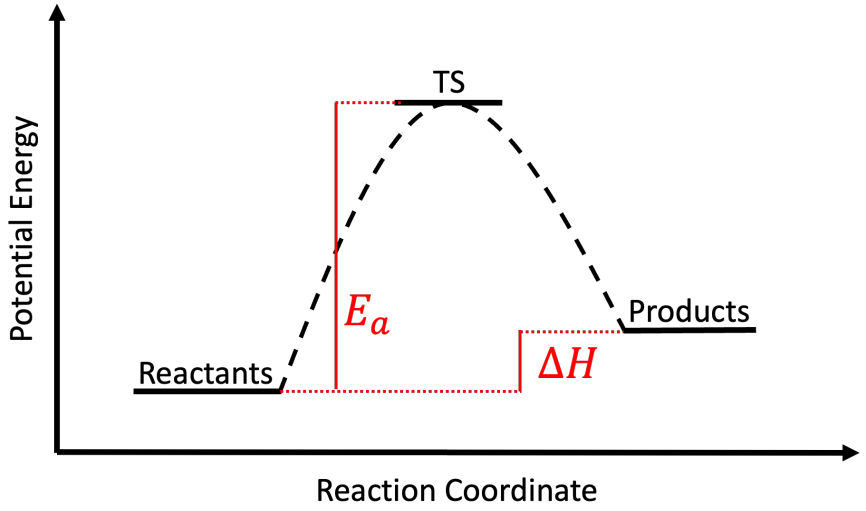
Chapter 4

Calculations on Small Nitrate-Based Systems

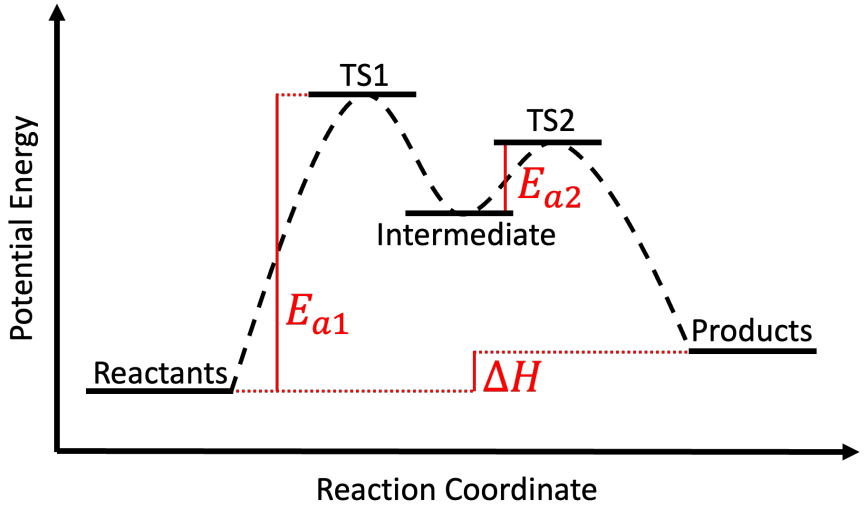
4.1 Introduction

The next part of this study focuses on quantum calculations, and using these to investigate the denitration reactions of nitrocellulose. This chapter contains several different short investigations into how DFT can be used to study the properties of model systems containing nitrate groups, such as nitrate esters and monomers of nitrocellulose, and study their fundamental reactions. It includes some details on the preliminary tests that were run to optimise parameters, and then some tests looking at possible mechanisms for hydrolysis of nitrate esters. The thermolysis reaction is harder to study as it requires the spin to be controlled, this is also investigated in this chapter. Finally, a couple of tests are run on monomers and dimers of nitrocellulose, looking at how to find the optimum position for the nitrate groups to be in, and looking at how the denitration of different positions can affect the reaction energy.

When looking at the kinetics of a reaction, there are different energy values that can be calculated, shown in figure 4.1a on the following page. One of the most important energies to be calculated is the activation energy (E_a, shown on the figure), as from the activation energy, information about the reaction kinetics can be calculated. The activation energy is the energy required to overcome the activation barrier. However, finding this value requires the energy of the transition state, which is not always easy to find. When studying a standard reaction where you know the reactants and the products, it is comparatively simple to find the reaction energy (ΔH). This is found by calculating the difference in the energies of the products and the reactants, and requires no information about the transition state.



(A) Single step reaction



(B) 2-step reaction with an intermediate state.

FIGURE 4.1: Sketch reaction profile diagrams for two reactions, and the energy values that can be calculated.

$$\Delta H = \sum E_{products} - \sum E_{reactants} \tag{4.1}$$

The reaction energy can be used to tell us how exothermic or endothermic a reaction is. The activation energy, however, E_a, can give us more information about the rate of reactions, through the Arrhenius equation as discussed in section 1.3.12. This requires a transition state to be found.

Sometimes the reactions being studied might have an intermediate state, a stable local minimum, as seen in figure 4.1b. The reaction state with the largest activation energy is going to be the rate-limiting transition, meaning that it is the activation energy for the overall reaction.

4.2 Testing Different Calculation Parameters

The effect that different exchange-correlation functionals, basis sets and even the chosen software can have on a calculation was tested using the reaction homolytic dissociation reaction:

$$R - ONO_2 + H^+ \longrightarrow R{-}OH^+ \cdot + NO_2\cdot \tag{4.2}$$

Two values were also calculated for the heterolytic reaction for comparison

$$R - ONO_2 + H^+ \longrightarrow R{-}OH + {NO_2}^+ \tag{4.3}$$

In this case the R- group being used was CH_3CH_2-.

The reaction energy was found, by calculating three geometry relaxations - R-ONO_2, and each of the two products separately - this was to ensure that both of them stayed in their radical state. A calculation was not run for H^+ as it is a single proton so there is no "geometry" to be optimised and it has no electrons, so DFT cannot be performed on it.

Calculations were run using Gaussian, ONETEP and NWChem. For each combination of functional / basis set and program, the calculations were run in a vacuum and also in a solvent with a dielectric constant $\epsilon = 3$, and plotted relative to each other, shown in figure 4.2. $\epsilon = 3$ was chosen as it represents the dielectric constant of a bulk substance of interest for the project.

The main functional being tested was PBE, as this is available in NWChem, ONETEP and Gaussian. One other functional was tested, M06-2x, to see how a hybrid functional would change the results calculated. ONETEP does not currently have hybrid functionals, so this functional could not be used for calculations that are comparable between ONETEP and NWChem. The calculations using NWChem and PBE were then run on basis sets of increasing size, to see how comparable they would be to the ONETEP calculations.

All of the energies calculated here are very high, an issue that was further looked into in Chapter 5. The problem is caused by the products and reactants not being in the geometries that would result from the general mechanism of an S_N2 reaction.

There is an obvious anomaly in this data - the NWChem (STO-3G, PBE) calculation. This value is likely to be different because it uses a very small basis set.

When the data is reframed to remove that data point, as in figure 4.3, the next two points that stick out are the two heterolytic reactions, which means that the reaction energy of the hydrolysis reaction is higher than the thermolysis reaction. The ONETEP

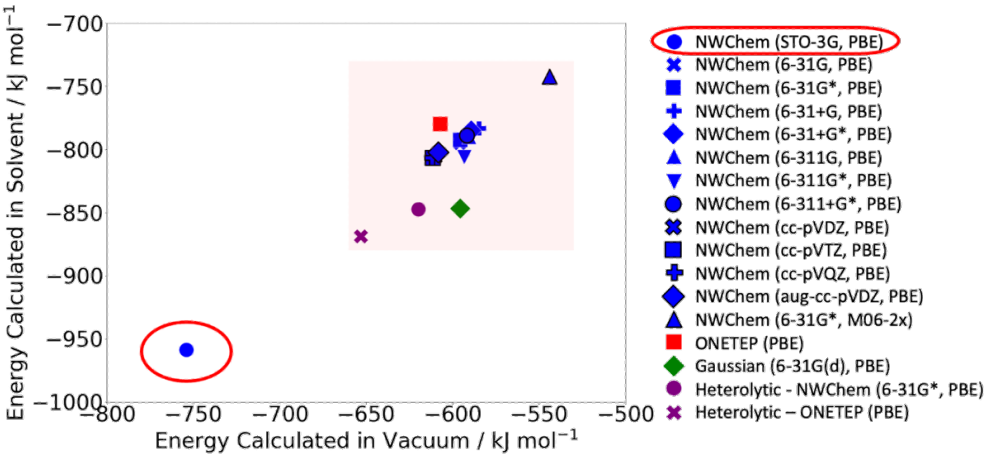


FIGURE 4.2: Initial graph for different calculated values. The circled calculation is discussed in the main text, and the red square shows the region that is plotted in the next figure (figure 4.3).

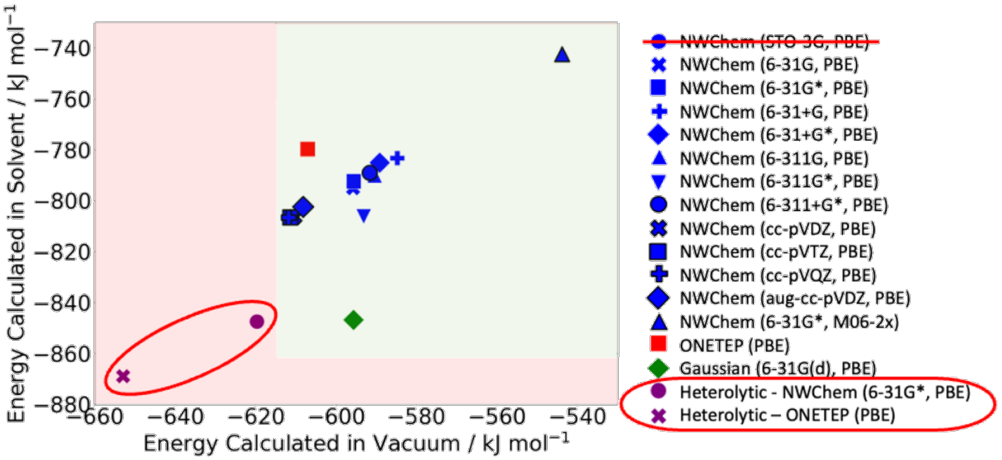


FIGURE 4.3: Region highlighted in red from figure 4.2. The circled calculations are discussed in the main text, and the green square shows the region that is plotted in the next figure (figure 4.4 on the facing page).

calculated reaction energy is a lot further out than the NWChem value, this is due to the NWChem calculations being with a smaller basis set meaning that the calculation is much less accurate.

Once the heterolytic reaction energies have been taken out of the frame, we begin to have a better fit, seen in figure 4.4. Two more values now stick out - firstly the NWChem calculation using the M06-2x functional, and secondly the Gaussian calculated value. The Gaussian value has a similar energy to the other structures in the solvated energy that is different. This means that the difference is probably caused by the solvent method being used. The M06-2x functional has a different energy for both the calculated values, but that is also expected, as this is the only hybrid functional being tested in this case.

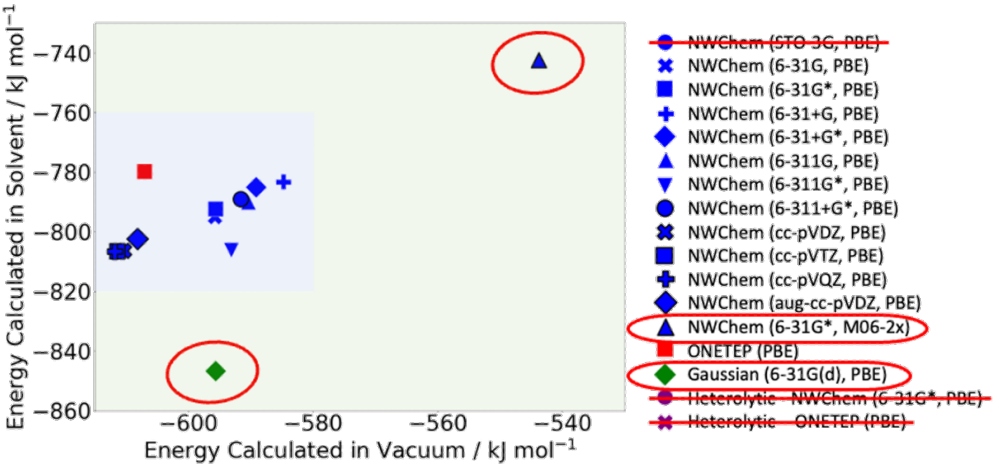


FIGURE 4.4: Region highlighted in green in figure 4.3 on the preceding page. The circled calculations are discussed in the main text, and the blue square shows the region that is plotted in the next figure (figure 4.5).

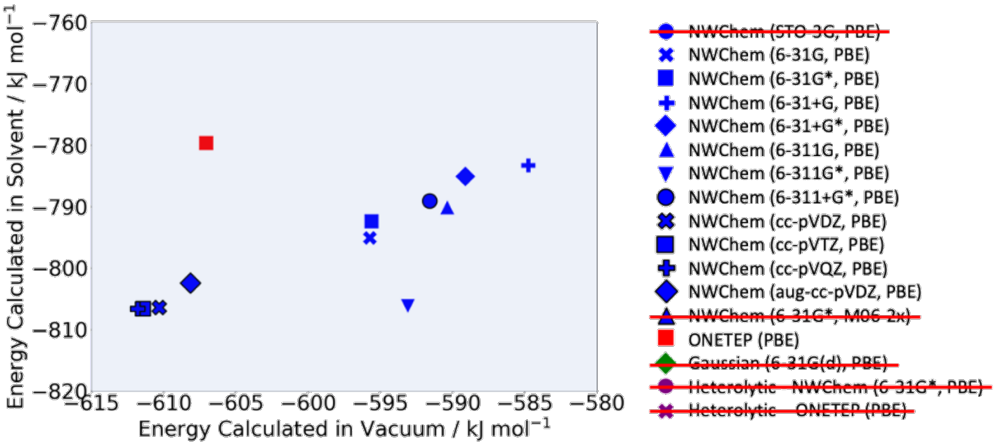


FIGURE 4.5: Region highlighted in blue from 4.4.

Finally, figure 4.5 shows that the results that remain are mostly NWChem PBE calculations. The bottom left of the framed data are the correlation-consistent basis sets. The augmented basis set, aug-cc-pVDZ, has lower calculated energies, and the other three, cc-pVDZ, cc-pVTZ and cc-pVQZ have similar solvent energies, but slightly increasing energies calculated in a vacuum, as the zeta value is increased.

The energies calculated in ONETEP are comparable to the correlation-consistent basis sets, as the largest basis sets being tested here, for the vacuum calculation. ONETEP did calculate a higher energy for the system in a solvent, but it uses a different solvent method, so that is expected.

The basis set chosen for future calculations in NWChem was 6-31+G*, as it is both polar and diffuse, but it is more efficient than the other polar and diffuse basis sets, and also the correlation-consistent basis sets, as can be seen in table 4.1.

4.3 Calibration of Number of Processors in NWChem

NWChem achieves parallelisation through MPI, which means separate cores, or processors, are able to run individual processes independently, and pass information back and forth. There is a limit to how much speed-up can be achieved through parallelisation, as there is an overhead caused by the need for communication between the cores. With this in mind, each of the six main functions in NWChem used in the process described above of investigating the reactions were tested using simple molecules. The ideal number of processors found through this method would then be used in future calculations.

These initial calibration tests were run with a nitrate molecule without a carbon chain, methyl nitrate. All the calculations performed here used the basis set 6-31+G*, and the functional B3LYP.

The chosen optimum number of processors are tabulated in table 4.2, and one example of the calibration curves is shown in figure 4.6.

4.4 Hydrolysis Reaction using NWChem

The hydrolysis of nitrate esters (previously shown on page 61) are as follows:

$$RONO_2 + H^+ \rightleftharpoons RONO_2H^+ \tag{4.4}$$

$$RONO_2H^+ \longrightarrow ROH + {NO_2}^+ \tag{4.5}$$

$${NO_2}^+ + OH^- \longrightarrow HNO_3 \tag{4.6}$$

TABLE 4.1: Total wall time for the vacuum and solvent calculations, for both the reactants and the products, for all the basis sets being tested in NWChem. The basis set chosen for further calculations is highlighted in bold

Basis Set	Total Calculation Time / s
STO-3G	60.8
6-31G	96.8
6-31G*	111.5
6-31+G	155.2
6-31+G*	**142.2**
6-311G	139.7
6-311G*	191.7
6-311+G*	302.6
cc-pVDZ	230.3
cc-pVTZ	1595.5
cc-pVQZ	28374.5
aug-cc-pVDZ	696.6

TABLE 4.2: Tabulated number of processors to give optimum speed-up.

Calculation type	Number of Processors
Geometry Optimisation	20
Single Point Energy Calculation	10
NEB	20
Saddle point search	40
Frequency testing	40

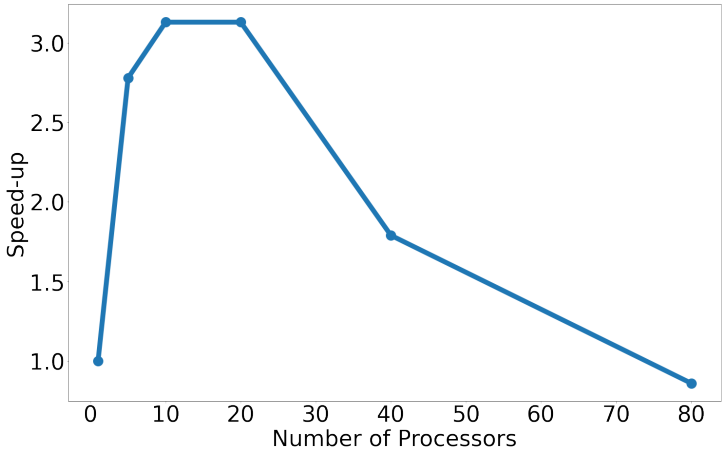


FIGURE 4.6: Calibration of single point energy calculations, using the total time taken to calculate the energy for 8 different starting molecules: CH_3OH, CH_3NO_2, CH_3NOOH^+, $CH_3O\cdot$, H_2O, H_2O^+, NO_2 and $NO_2\cdot$. Even with 8 molecules, this is a fairly short time scale, with the longest time being 38 seconds, and the fastest taking 12 seconds. 10 processors were chosen for future calculations. These molecules were tested as they were the extent of the calculations being run at the time. This test should have been repeated with larger molecules when the size of the systems being tested was increased.

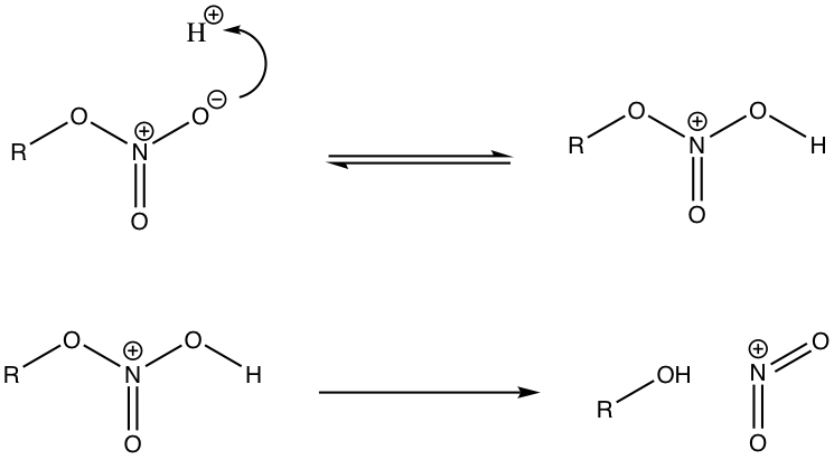
For this aspect of the project, the hydrolysis reaction was investigated on small nitrate esters, described in figure 4.7. This is a two-step reaction, each of which must have separate calculations.

4.4.1 Transition state searching

Both NWChem and ONETEP NEB calculations struggled to converge on molecules of even this size, taking several thousands of iterations for relatively small molecules, and still not converging, meaning it would not be feasible for larger systems. For both programs default convergence settings were used.[66,74]

Therefore to investigate the reactions in NWChem a process was developed, using the following steps:

1. Forming of Geometries - basic atom arrangements were put together using Avogadro.



R = $CH_3CH_2CH_2CH_2CH_2CH_2$

FIGURE 4.7: The reaction scheme for hydrolysis using R=$CH_3(CH_2)_5$, based on equations 4.4, 4.5 and 4.6.[118] The mechanism of how each of these steps proceeds is to be determined using DFT.

2. Geometry Optimisation - for every initial molecule made, a geometry optimisation was performed using NWChem.

3. Reactant/Product Geometries - for the reaction being tested the optimised geometries of the initial molecules were put together and centralised as needed.

4. NEB - using NWChem NEB calculations were run. These were not converging so an arbitrary number of iterations were run, to get an estimate of a transition state. In this case 20 was chosen

5. Visualisation of Results - The bead energies were plotted, and the geometries of maxima were taken as preliminary saddle points. The path was also visualised in VMD.

6. Saddle point search - saddle point searches were run on the maximum beads of the NEB.

7. Frequency tests - Frequency tests were performed on every saddle point found to prove that it was a saddle point, and to see if the imaginary frequency could give any clues as to the route of the reaction when visualised in VMD.

8. Solvent - In cases where transition states were found, calculations of the energy were re-run using solvent methods, to see how the energy was be affected.

4.4.2 Results for hydrolysis Stage 1

The first part of the reaction, the addition of a H^+ ion, was run in two separate ways in NWChem. Initially, this was run with a H^+ ion at an arbitrary distance from the

starting molecule. The geometries initially used were in different orientations in space, meaning that the product needed to be rotated before it could create the product specified. This is what is happening between the beads where the energy is remaining nearly constant in figure 4.8. The figure does also tell us that there is no real energy barrier to be overcome in the addition of H^+ to the molecule, as there is no maximum on the path.

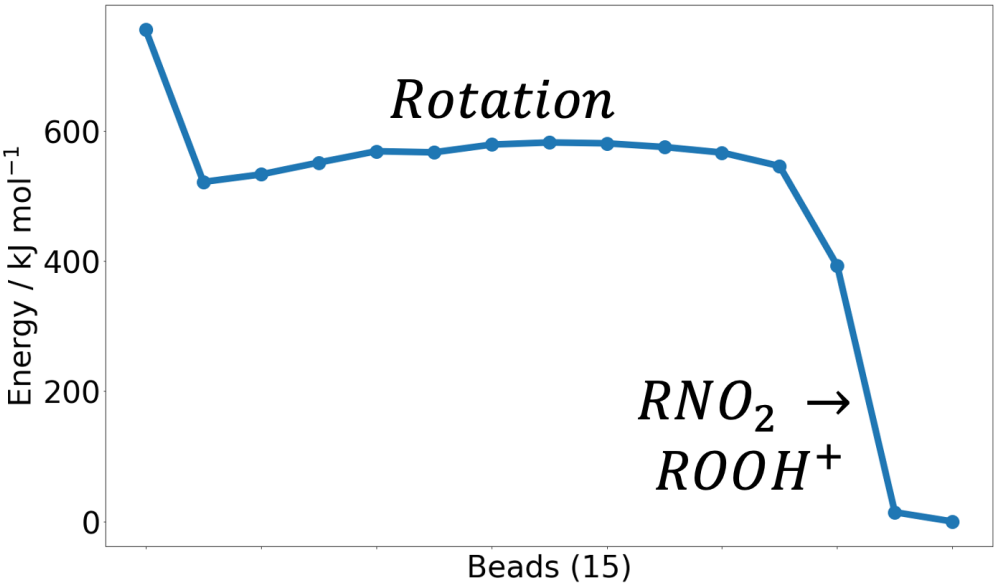


FIGURE 4.8: Bead energies for $CH_3(CH_2)_5ONO_2 + H^+ \rightarrow CH_3(CH_2)_5ONOOH^+$. NEB calculation, run in NWChem, with 15 beads, using basis set 6-31+G* and functional B3LYP, after 500 iterations

In an attempt to confirm whether this was the case, the same was repeated, using a hydronium ion and giving water as a product. The NEB geometry (bead energies for which are shown in figure 4.9) shows that the H^+ breaks away at the start, and without any real prompting.

Both of these confirm the fact that this part of the reaction is considered to be an equilibrium reaction, because of the lack of energy barrier.

This reaction coordinate diagram is unusual, due mostly to issues with the structures being used - the fact that there are so many beads being used on the rotation of the structure indicates that the reactants and products should be set up more carefully. The same reaction was tested again later in section 4.5.1, and a more classic reaction coordinate diagram was obtained.

4.4.3 Results for hydrolysis stage 2

Running NEB on the second step of the hydrolysis reaction gave an interesting profile. As shown in figure 4.10 there appear to be 3 beads that have higher energy than others.

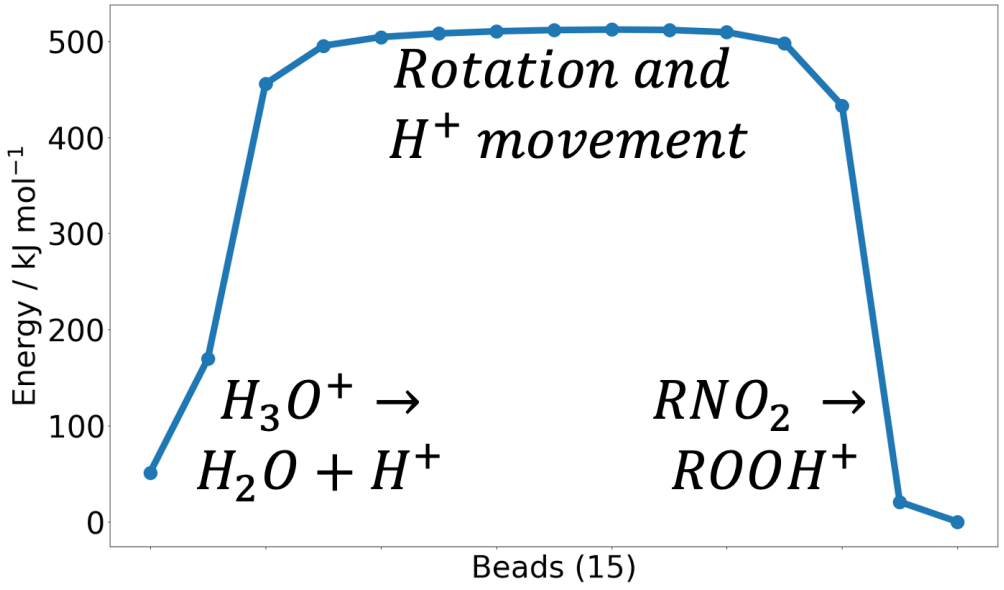


FIGURE 4.9: $CH_3(CH_2)_5ONO_2 + H_3O^+ \rightarrow CH_3(CH_2)_5ONOOH^+ + H_2O$. Bead energies for NEB calculation, run in NWChem, with 15 beads, using basis set 6-31+G* and functional B3LYP, after 500 iterations

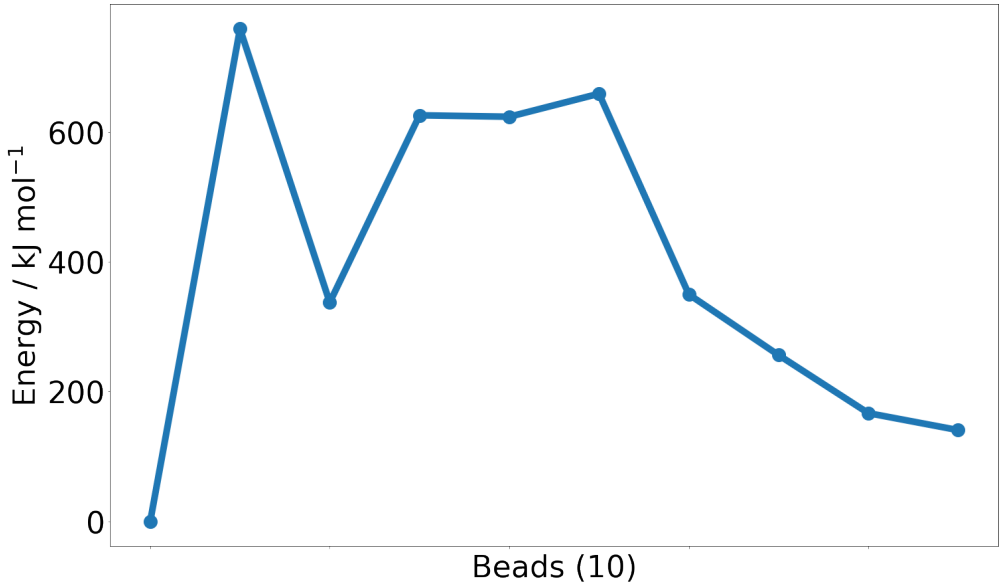


FIGURE 4.10: Bead energies for $CH_3(CH_2)_5ONOOH^+ \rightarrow CH_3(CH_2)_5OH + NO_2^+$. NEB calculation, run in NWChem, with 15 beads, using basis set 6-31+G* and functional B3LYP, after 500 iterations

Each of these, beads 2, 4 and 6, had saddle point searches run on them. Bead 6 gave a saddle point in which the backbone of the molecule had been rearranged, as seen in figure 4.11, so it was decided that this was not likely to be on the final MEP. Beads 2 and 4 had saddle points found, and using their frequencies, an idea of the reaction was found, shown in figure 4.12. This seems to represent a feasible mechanism.

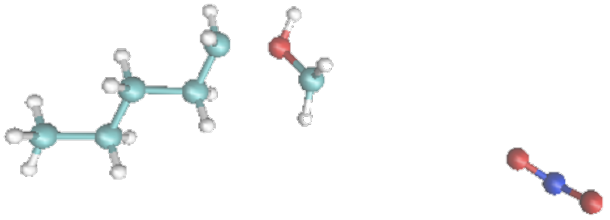


FIGURE 4.11: Saddle point found using NEB bead 6 as initial geometry. The backbone of the molecule has been changed, with the oxygen inserting itself into the chain, meaning it is unlikely to be part of the reaction pathway. The distance between the 5th carbon and the oxygen is 1.8 Å, meaning that although the bond is not visualised it is still there.

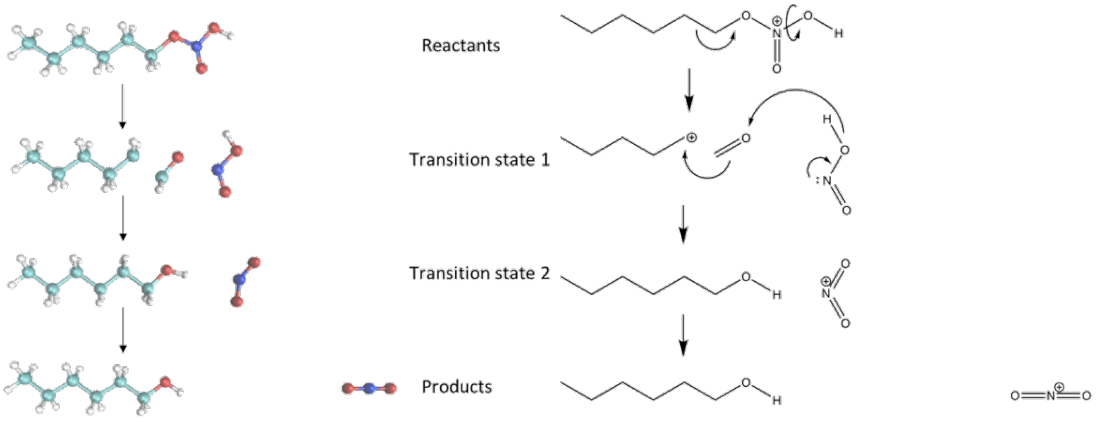


FIGURE 4.12: An idea of how the mechanism for the second part of the hydrolysis reaction could proceed. Figure 4.13 shows the bonding in the first transition state.

The bond distances shown in figure 4.13 show that the last C-C bond probably does not fully break as shown in the mechanism of figure 4.12, however, the near planar arrangement of the CH_2-O implies that there maybe some amount of π bonding character to the C-O bond, which helps in the design of a reaction mechanism where the charge is conserved.

4.4.4 Intermediates

For a reaction such as that found above, containing two transition states, there must be an intermediate state, as shown in the reaction profile in figure 4.1b on page 100. Therefore the next step was to find this intermediate. Several methods were used to investigate this.

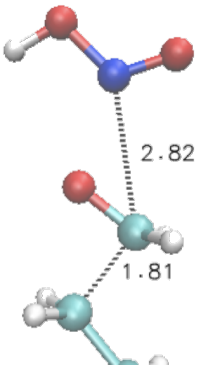


FIGURE 4.13: The bond distances at the end of the chain for the first transition state.

1. The first method attempted was to look back at the original NEB results, take the geometry at the minimum between the two maximums, and optimise it. This unfortunately gave a geometry with the chain rearranged, like the geometry shown in figure 4.14, so was not correct.

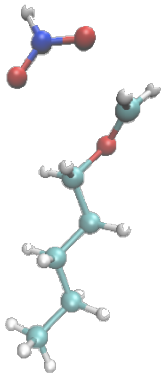


FIGURE 4.14: Example of a failed intermediate geometry, where the chain backbone has been reordered.

2. The next idea that was tested was to run a short NEB with the start and finish geometries as the transition states. This however also produced a reordering of the backbone as shown in figure 4.14.

3. Looking at the vibrations of the imaginary frequencies for transition state one, the frame that was furthest in the forward direction of the reaction was taken as a starting value for geometry optimisation, but once again the backbone was reordered.

4. The final attempt was to make a guess geometry - with the hydrogen halfway between where it is in the two transition states. This was optimised giving a geometry, seen in figure 4.15 that would make sense in the mechanism, although it is similar to transition state 2 with rotation around the N-OH bond. This had an energy that was around equal to the starting reactants, shown on the reaction profile in figure 4.17. As this was the only possible intermediate found that did

not involve rearrangement of the backbone, it was chosen as the most feasible intermediate and used for the next set of calculations.

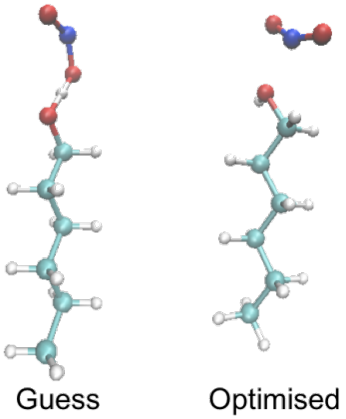


FIGURE 4.15: The initial guess for an intermediate, and the optimised geometry for it.

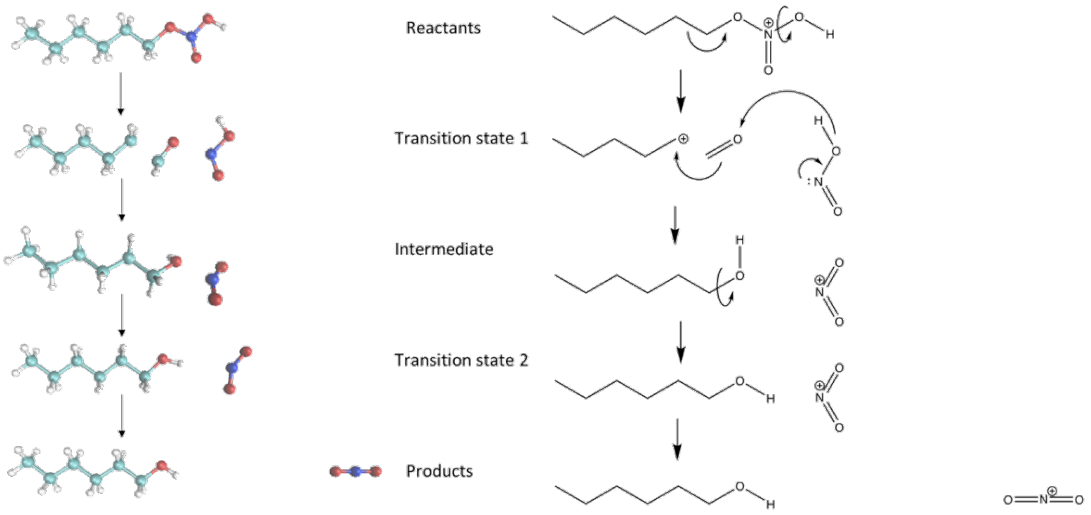


FIGURE 4.16: A possible mechanism for the second part of the reaction found using NWChem calculations.

4.4.5 Solvent

Using this final intermediate, and the other geometries previously found, a reaction coordinate diagram was plotted, and is shown in figure 4.17. Again, this reaction profile graph is problematic, as the products are extremely unfavourable compared to the optimised transition states and intermediate. This could be due to the wrong minimum energy geometry being found, or due to how the two products are being combined.

However, figure 4.18 shows that by performing the calculation in a solvent such as water the energy of the products becomes lower as the solvent stabilises the system.

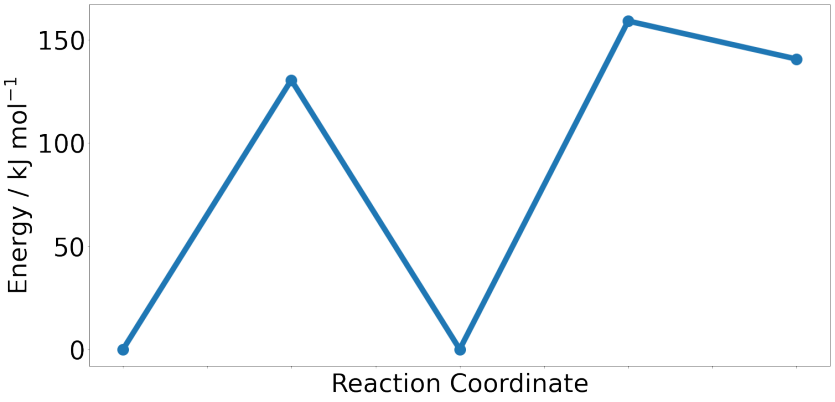


FIGURE 4.17: Energy profile plotting the energy of each known geometry in the hydrolysis reaction, using ab initio calculations, using basis set 6-31+G* and functional B3LYP

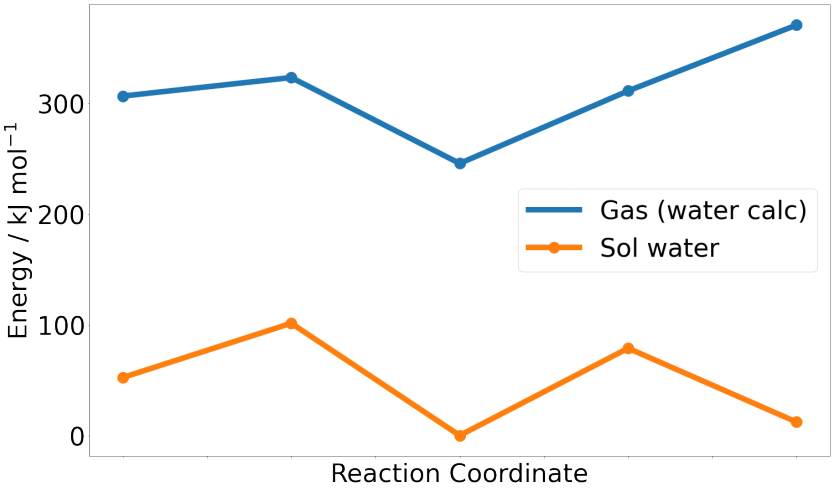


FIGURE 4.18: Energy profiles with solvent methods applied, using basis set 6-31+G* and functional B3LYP. The activation energy of this reaction would be 48.3 kJ mol^{-1}.

This is promising as according to the data collected by Brill and Gongwer,[118] the activation energy for hydrolysis of nitrocellulose is $\sim$ 110-115 kJ mol^{-1} (seen in figure 2.3 on page 63). It is plausible then that for a simpler nitrate ester, the activation energy could be 48.3 kJ mol^{-1}, as shown in the energy profile in figure 4.18.

4.5 Hydrolysis Reaction using ONETEP

The process was then repeated using ONETEP for the calculations. The NEB calculations were run with the initial geometries found in the NWChem process, to save time. Hybrid functionals were not available in the ONETEP code at the time of this work, so a functional called revPBE was used.[138] This was chosen as it is a functional that is in both NWChem and ONETEP, so as to make it more comparable.

Unfortunately, there was not enough time to rerun all of the NWChem calculations with a comparable functional, as the previous calculations had been done with B3LYP.

4.5.1 Hydrolysis part 1

The first part of the hydrolysis reaction was tested in ONETEP, and the bead energies are shown in figure 4.19. This is similar to the results found in NWChem, there is more features in the profile, but the energy barrier seems to be the same energy, 510 kJ mol^{-1} in NWChem and 580 kJ mol^{-1} in ONETEP, a difference that might be accounted for by the use of different functionals.

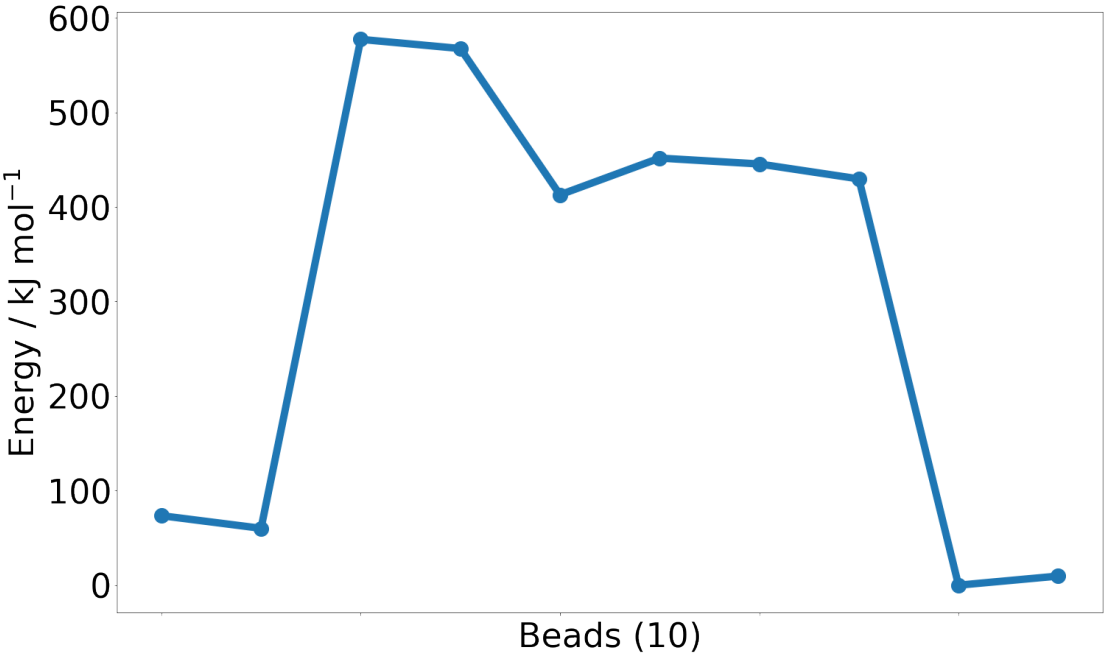


FIGURE 4.19: Bead energies for ONETEP NEB calculation of the first part of the hydrolysis reaction after 59 iterations, using revPBE.

Through visualising the path, it appears that the transition state is once again the H^+ ion dissociating with the H_3O^+ ion, as found before.

4.5.2 Hydrolysis part 2

The initial NEB for the second part of the hydrolysis reaction was run on the same starting and finishing values as was used in figure 4.10, and the bead energies were calculated as shown in 4.20, which lead to an estimate mechanism shown in 4.21.

This supports the mechanism suggested from the NWChem results. The mechanism shown to get from b to c is not as good, the mechanism found in NWChem went

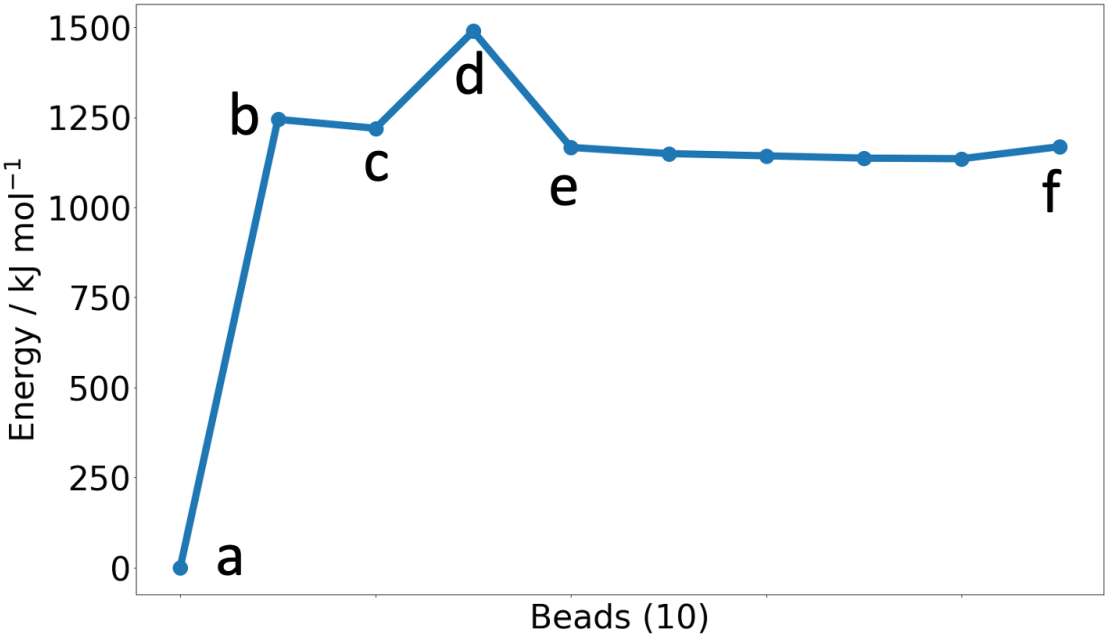


FIGURE 4.20: Bead energies for the NEB calculation for the second part of the hydrolysis reaction, in ONETEP, using revPBE, after 54 iterations. The letters refer to the geometry steps used in figure 4.21

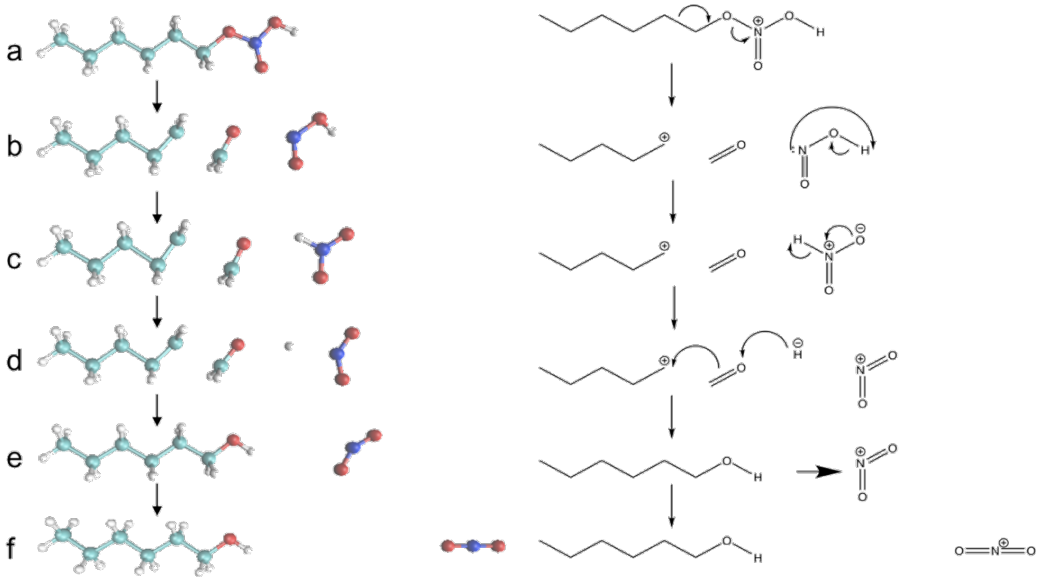


FIGURE 4.21: A possible mechanism for the second part of the hydrolysis reaction, based on the results from the NEB calculation in ONETEP

straight from something similar to b, to something similar to d. It might be that if the NEB ran to completion geometry c would not appear.

4.6 Summary of transition state searches for hydrolysis reactions

Although neither NWChem or ONETEP were able to converge the NEB calculation, other programs were not used as Gaussian does not offer NEB, and CASTEP has only implemented it recently in 2022, after this area of work was completed. [139] Other transition state searching methods, including LST-QST in NWChem and Gaussian were tested but did not succeed in getting any significant results.

Most of the issues in this section have been suggested to be due to the set-up of the products and reactants used here. This is partially due to the way in which the two reactants and two products are combined.

The next chapter focuses on the basic hydrolysis, and problems were encountered before it was realised that this reaction should be modelled as an S_N2 reaction, which means that the chirality of the groups should switch. In a similar way, the general reaction mechanism for acidic hydrolysis should be considered as there may be a similar issue with how it has been modelled here.

4.7 R-O-NO_2 Homolysis / Thermolysis Reaction

The second of the key reactions that was suggested to be looked into in this project is thermolysis, which can be modelled by a homolytic bond breaking in equations (2.8) - (2.10) (originally found on page 61), and already tested briefly in section 4.2 on page 101.

$$\mathrm{RONO_2} \longrightarrow \mathrm{RO}\cdot + \mathrm{NO_2}\cdot \qquad \text{(initiation)} \tag{4.7}$$

$$\mathrm{RO}\cdot + \mathrm{NO_2} \longrightarrow \mathrm{ROO}\cdot + \mathrm{NO} \qquad \text{(propagation)} \tag{4.8}$$

$$2\,\mathrm{ROO}\cdot \longrightarrow \mathrm{RC{=}O} + \mathrm{O_2} + \mathrm{ROH} \qquad \text{(termination)} \tag{4.9}$$

The difficulty in studying this reaction comes from the radicals. It is not possible to run a MEP searching algorithm such as NEB, because they can only take one spin state as an input. Because of this, there has not been much use of DFT to study homolytic reactions generally. In 2019, Hait *et al.* used Time Dependant DFT (TDDFT) to find the Coulson-Fisher point, the point of the bond dissociation at which the spin-polarised

state becomes more energetically stable.[140] They used this to study the dissociation of H_2 molecules. Very recently some more results have been published by the same people, Hait and Head-Gordon, applying the same methods to some other small molecules, such as H_3CCH_3 and H_3NBH_3, however, there are no published results for a nitrate ester.

4.7.1 Homolytic fission of R-O-NO_2

When a bond is homolytically broken, there are three different theoretical states that the electrons could be in, shown in figure 4.22.

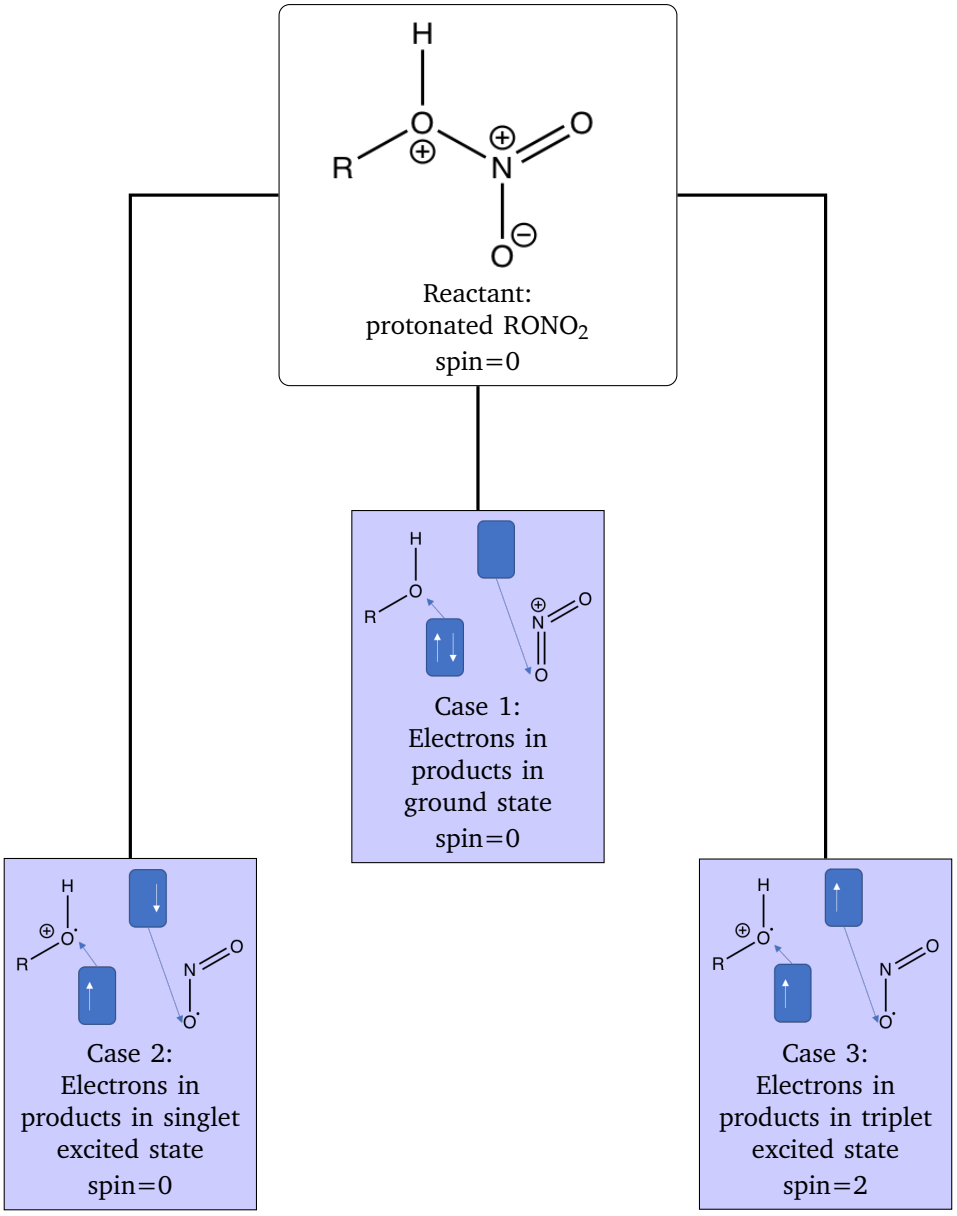


FIGURE 4.22: Different possible spin states for ROH^+NO_2

The first case, the singlet ground state, is where the electrons remain paired, and a radical has not been formed. Here the spin is 0, and this is the result that will be obtained when a restricted, RHF, calculation is run with spin equal to 0.

The third case, the triplet excited state, can be found by running restricted open shell, ROHF, calculations, with the spin set to 2.

For the second case broken symmetry calculations must be run, so make sure radicals are formed, but the overall spin remains 0.

4.7.2 Calculations

ONETEP contains a variant of DFT, Ensemble-DFT (EDFT). This improves on one of the weaknesses of regular DFT, which is its low effectiveness on systems containing a band gap, such as semiconductors and insulators. EDFT considers an ensemble of ground and excited states, by calculating over an ensemble of different excitation states.[142] This can be useful for studying this particular reaction, as it can then give the optimum amount of spin contamination for a system.

The homolytic fission of bonds has mostly been studied on H_2 diatomic molecules, but this investigation took the theory and applied it to nitrate esters. As the bond that is to be broken lengthens, there reaches a point where having spin in the system becomes energetically favourable - this is known as the Coulson-Fischer point.[140] This point was found for the reaction shown in equation (4.8).

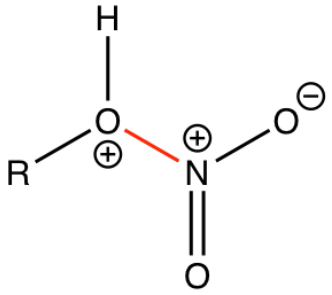


FIGURE 4.23: Structure of the model with the bond to be broken highlighted in red.

The initial geometry of the reactant was fully optimised using ONETEP. From this optimised geometry the length of the RHO^+-NO_2 bond was adjusted, and new structures generated with lengths every 0.1 Å, adjusting the initial bond length by -0.6 Å to +2.4 Å. Each of these new structures then had further calculations performed on them, with the positions of the O and N fixed to maintain that bond length. Three different optimisations were run for each structure - one with spin=0, one with spin=2, and one with spin relaxation (through EDFT) on.

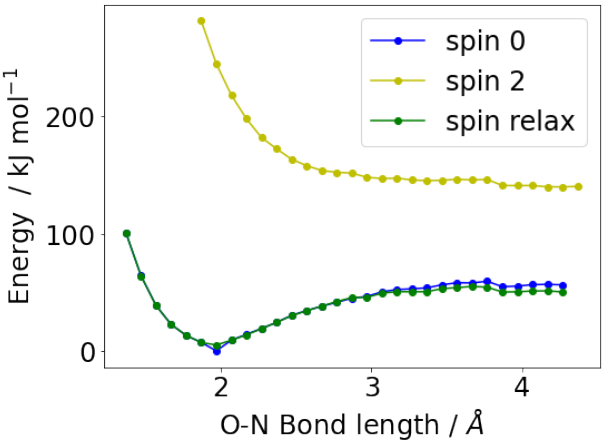


FIGURE 4.24: Energies calculated for geometry optimisations whilst adjusting the bond length, using different spin settings

The results can be seen in the graph in figure 4.24. It can be seen that the pure spin 2, triplet excited state, model is highly energetically unfavourable, and for the shortest bond lengths these were unable to be calculated. The spin relaxation calculations at the close bond length are giving approximately the same energy as the spin 0 calculations, meaning that the ground state is the most stable.

As the calculated energy increases, there begins to be a divergence as spin contamination begins to appear in the spin relaxation calculations. The ONETEP output file gives a variable called $\langle S \rangle$, which is a measure of the contamination. Plotting this, as in figure 4.25 confirms that the divergence is happening where the spin 0 and spin relaxation calculations begin to split at 2.77 Å. There is a visible inconsistency at the minimum energy structure, the spin 0 calculation of 2 Å. This is due to ONETEP optimising to a structure that is visibly not correct.

Another set of calculations was then run, to compare the energies of pulling the bond apart for a primary nitrate to a secondary nitrate. Shown in figure 4.26, the results show that when the nitrate is on a secondary carbon the divergence in energy is at approximately the same bond length - 2.77 Å, but the secondary carbon is at a higher energy after the spin has diverged. This means that the thermolysis reaction will definitely be affected by the nitrate position - the off-ring positions are primary nitrates, and the other positions are secondary.

Knowing the Coulson-Fischer point, unfortunately, does not tell us anything about the overall reaction energy, or the energy required to break this bond, but it does tell us the length at which the bond is likely to break and is an interesting first step to discovering more about the homolytic fission reaction of the bond.

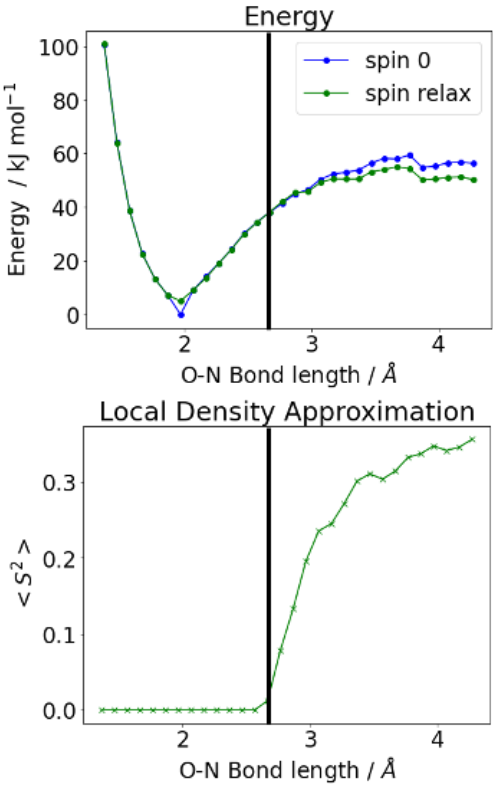


FIGURE 4.25: Results show the spin contamination beginning when the energies of the geometries optimised with spin relaxation begin to diverge from the geometries optimised with spin fixed to 0.

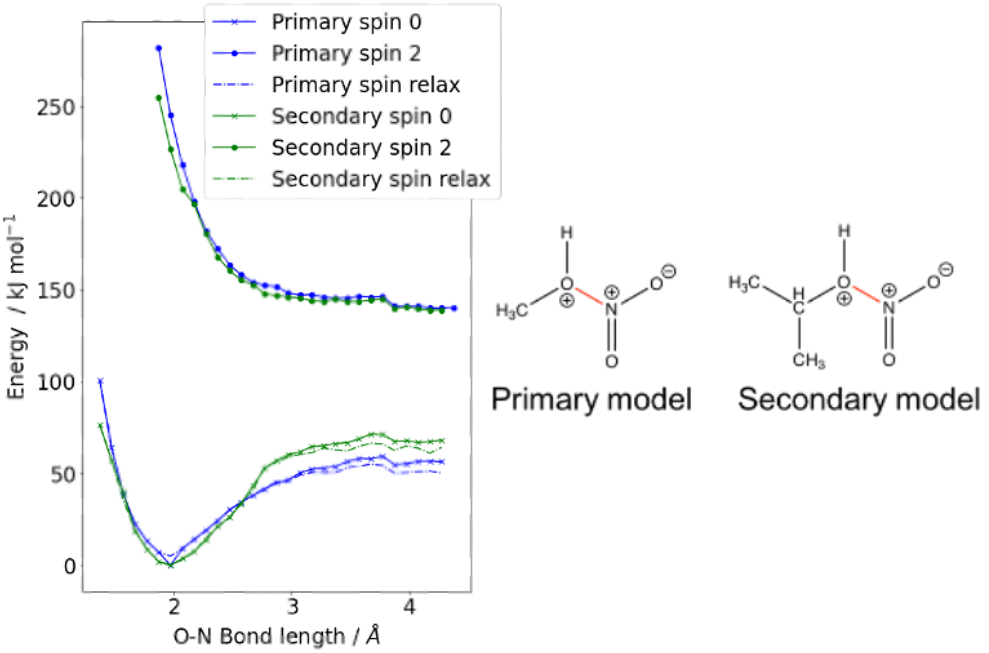


FIGURE 4.26: Energies calculated using different spin settings for a primary and secondary nitrate group.

4.8 Nitrocellulose Torsion

The DFT calculations so far described in this chapter are focused on finding relaxed geometries by finding minimum energy structures and trying to find transition states, by searching for saddle points. However, it can be also used to map a potential energy surface as a particular parameter is changed, by manually changing the structure and calculating the single-point energy.

In their 2017 paper, Nardai and Bohn calculated how the energy of a monomer of nitrocellulose changed as the dihedral angle of the nitrate group removed from the ring (carbon position 6) was varied.[72] They were using this as a way to verify the use of the COMPASS II forcefield on nitrocellulose.

For this project, a short investigation into the dihedral angles of a monomer of nitrocellulose was run to compare the accuracy of NWChem DFT calculations. The monomer structure was formed by taking a monomer out of the cellulose unit cell previously described.[136] The OH groups on C2 and C3 were then replaced by ONO_2 groups, set so that the group was in the same plane as the H on the same carbon, as shown in figure 4.27. The OH group on C6 was changed to an ONO_2 group, and multiple structures were created with the dihedral varied incrementally.

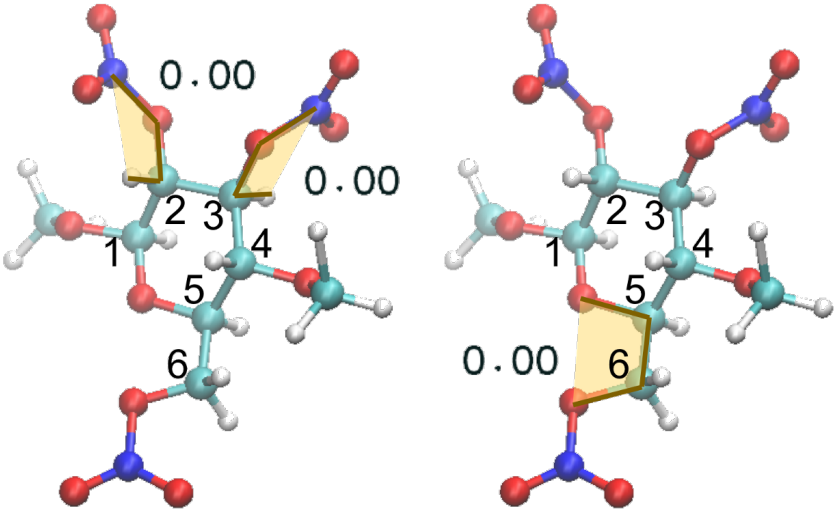


FIGURE 4.27: Torsion angles of ONO_2 groups, on the left, the torsions on C2 and C3 are fixed, and on the right, is the torsion at C6 which is to be changed incrementally.

Geometry optimisations were run on the structures are described as follows. All calculations were performed in NWChem, using 6-31G* basis set and BLYP functional. Energies were plotted against the dihedral angle, examples of which are shown in figure 4.28. From the results presented in the paper, it was expected that the plot would show two minima, at $\sim 60°$ and $\sim 300°$, and maxima at $\sim 140°$ and $\sim 225°$ and a global maximum at $0°$.[136]

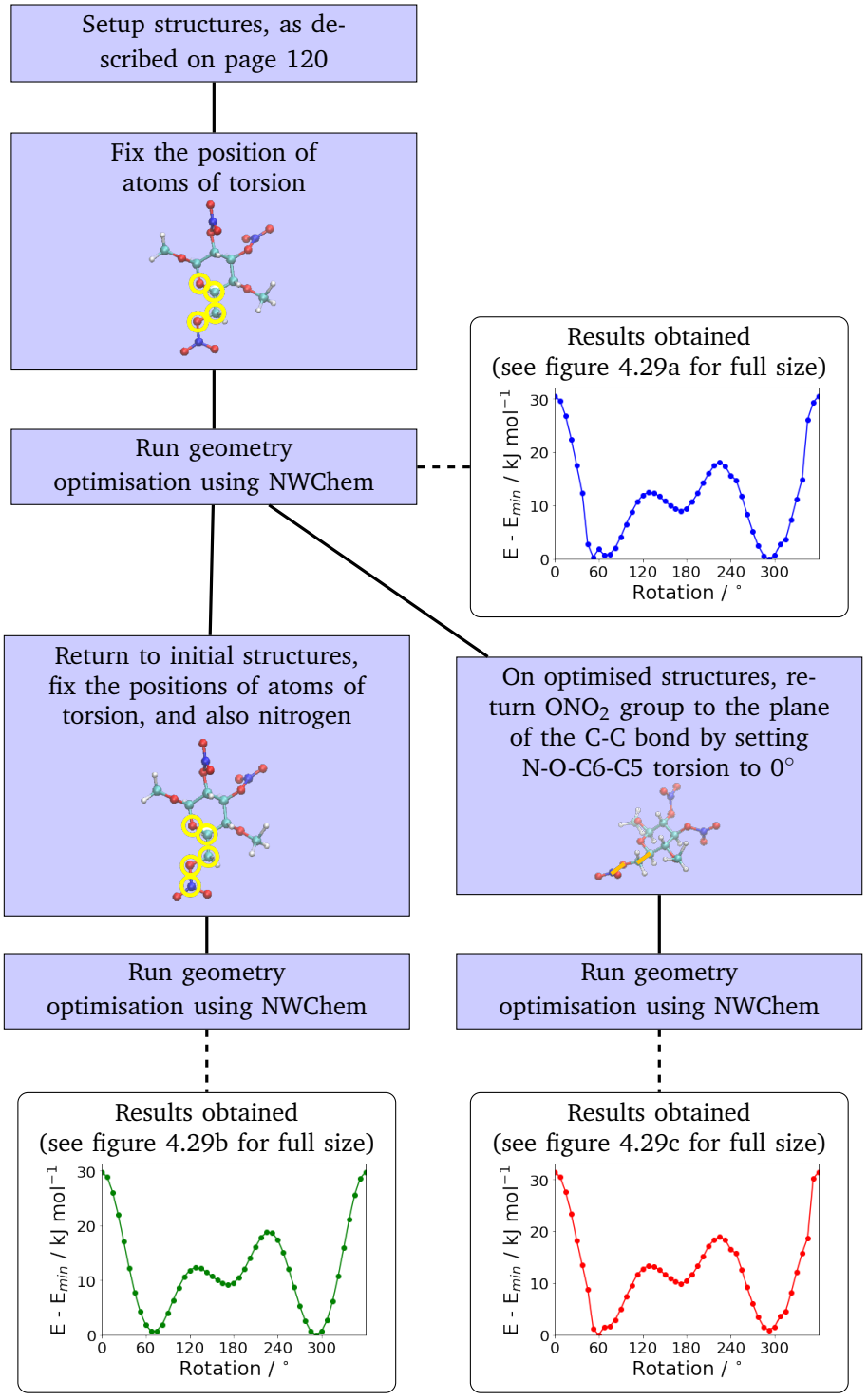
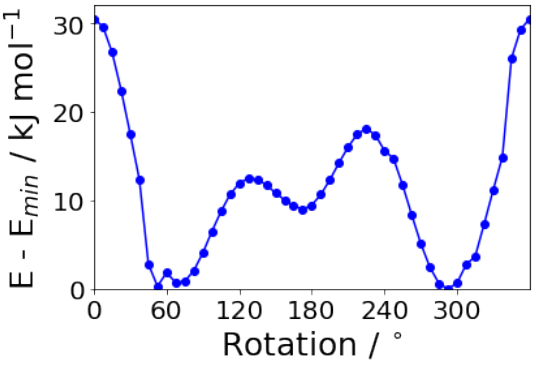


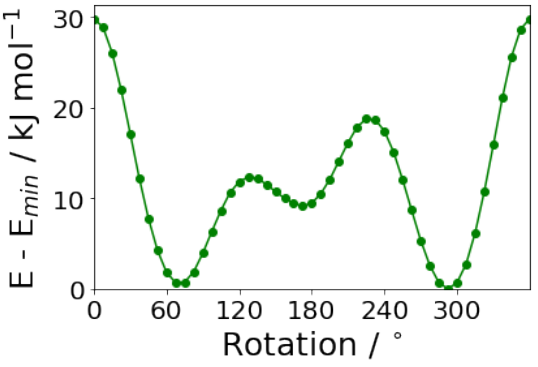
FIGURE 4.28: Trial and improvement of torsion testing method.

- NWChem geometry optimisations were run on the structure, with the position of the 4 atoms in the dihedral fixed so that the angle was maintained. This resulted in a plot that was generally shaped as expected, although there was a strange kink at $\sim 60^\circ$ and the energies between 300° and 360° were not as smooth as expected. This discontinuity was only approximately 2 kJ mol^{-1}, but it was still investigated - in the literature this energy profile was smooth. This lead to visual inspection of the molecules, which are shown in figure 4.30, showing that there is something unusual going on with the structures, and prompted further analysis.
- To try and prevent these discontinuities two different approaches were tested. The first of these returned to the original structures, and re-ran the geometry optimisations, holding the 4 atoms of the torsion fixed, and also the nitrogen atom. This produced the results shown in figure 4.29b. It can be seen that this gives the smoothest curve.
- The second method was to assume that the minimum found might not be the most favourable - the visualisation of the structures, shown in figure 4.30 on page 124, shows that for the structures from 37.5° to 67.5° the position of the $-ONO_2$ group was at a different angle from most other structures. To solve this the N-O-C6-C5 torsion was set to 0° to move the nitrate group back up. The geometries were then reoptimised, the results being shown in figure 4.29c. The results were not particularly improved by this.

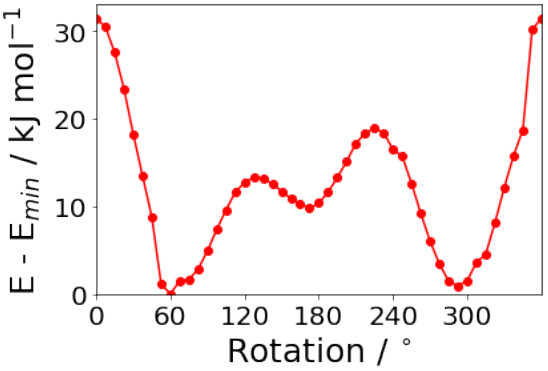
Looking at the final results in figure 4.29 on the facing page, fixing the position of the nitrogen atom as well as the 4 atoms in the dihedral removed the unusual structures that were being created.



(A) Initial results when 4 atoms are held fixed during geometry optimisation



(B) Results when 5 atoms are being held fixed during geometry optimisation



(C) Results when the structure was reoptimised after the N-O-C6-C5 torsion had been reset to 0^o

FIGURE 4.29: Energies of optimised geometries when the O-C5-C6-O torsion angle is held fixed using different methods

4.8.1 Torsion in unit cell being used

The unit cell being used for the large-scale structures in the previous chapter were based on crystallographic data of cellulose obtained from experimental structures. Adding nitrate groups might cause a significant effect on what torsion angles are the most energetically efficient, as the nitrate group is considerably larger than a hydroxyl group, so this analysis could help support whether the torsion in the unit cell being adapted fits a nitrate group.

The torsions of the C6 nitrate group were measured to see how well they agree with the DFT energies. The unit cell given was based on a dimer, so there are two dihedrals in it, one which was 306° and the other 299°. Comparing this to the DFT calculations run in this section, these are in the minimum energy regions, which supports the unit cell being realistic, as it is in a low energy configuration.

This small analysis does not take into account the interchain interactions, therefore providing information on what the torsion angle would be likely to be within the large-scale structures after the protocol has been performed on them. This would be

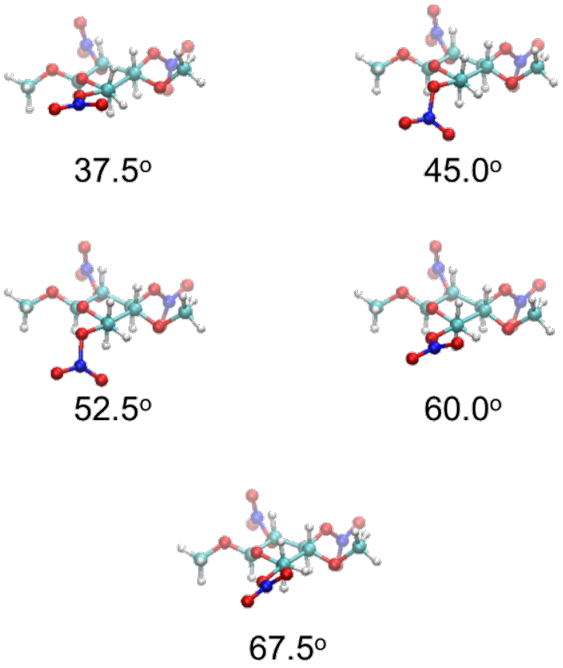


FIGURE 4.30: Optimised geometries from calculations with 4 atoms of the torsion fixed, from around the 60° geometry. Although the nitrate groups are fixed in place, these visualisations have been rotated so that the ring is in the same position, for visual comparison.

hard to do, as although the initial distances between the chains are the same, the energetically stable structures all have different densities and different distances between the chains. On the other hand, it does validate the addition of nitrate groups without making changes to the dihedral angle.

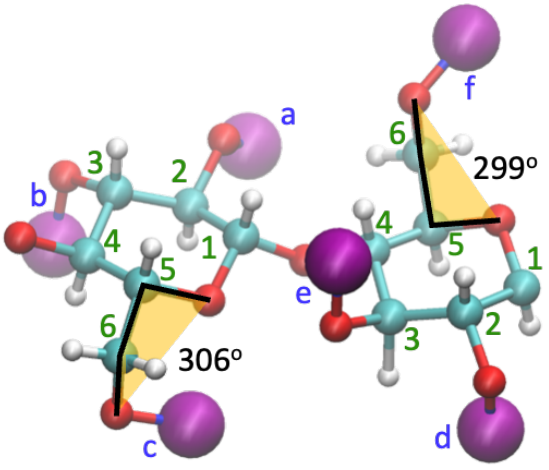


FIGURE 4.31: A dimer from the unit cell. The -NO_2 groups are beaded together and shown in purple.

4.9 Conclusions

This chapter has included several investigations into the properties of nitrate esters, monomers and dimers of nitrocellulose.

The hydrolysis of nitrate esters was tested, using NEB theory, and a suggested mechanism found, which agrees with both ONETEP and NWChem calculations. The differences between the ONETEP and NWChem results for stage 1 can be accounted for by the fact that different functionals have been used, as ONETEP did not have hybrid functionals at the time that these calculations were being run. However, for stage two, the energy difference between the beads of the ONETEP calculation are considerably higher, almost 1,500 kJ mol^{-1}. At the time that these calculations were run, NEB had only just been implemented in ONETEP, so it could have been an issue with that, or it could just have been that the ONETEP calculations only ran just over 50 iterations before the project needed to move on. When the energies of the structures for this method are calculated using a solvent method, the calculated activation energy is comparable to experimental literature values, such as those presented in figure 2.3 on page 63.[118]

Information about the energies of the thermolysis reaction was not able to be found using the methods used on the hydrolysis, as the radicals formed need spin settings to be specified, which is not possible with NEB calculations. A study into the properties of the homolytic breaking of the bond was run, by calculating the energy of different spin states at different bond lengths, something that has been done previously to study the homolytic fission of small molecules such as H_2, but not on larger molecules such as the ones used in this project.[140] This was able to give an idea of at what point it would be more favourable for the stretched bond to become two radicals but does not give any information about any energy barriers that might be required to break the bond. At this point, it was decided that this had strayed too far from the initial remit of the project, however, this is something that has potential for further study.

Finally, a short investigation was performed into the orientation of the nitrate groups, by finding the angle with the optimal energy. This was used to support, and increase confidence in the unit cell used for the nitrocellulose structures, as the unit cell being used has the nitrate groups in an optimal orientation.

Chapter 5

Hydrolysis of nitrocellulose in large-scale systems

Having developed the methods described in the previous chapters, the next step was to apply some of the methods developed for studying the degradation reactions from chapter 4 to larger models of nitrocellulose such as the paracrystalline structures created using the methods discussed in chapter 3.

5.1 Defining Structures for Use in this Chapter

In this chapter two main types of structure will be used. As the focus is on getting results that recreate realistic systems, both of these are going to be 4/5 N 12 wt% nitrated, as described at the start of Chapter 3. This is the structure created of alternating units with 4 or 5 out of the possible 6 positions nitrated. The labelling of the positions is demonstrated again in figure 5.1. For the initial calculations, the structure being used is the smallest unit possible of 4/5 N nitrocellulose - this is a short chain, 4 units long, the first two forming a 4-nitrated dimer, with positions **a, c, d** and **f** nitrated, and the final two forming a 5-nitrated dimer, with positions **a, c, d, e** and **f** nitrated, shown in figure 5.1. At the end of a natural chain, the end groups will be hydroxyl groups. However, the length of the chain we are studying is much shorter than would normally be considered, so for results more representative of large-scale nitrocellulose, a different functional group should be used. A methoxy group would have improved the precision of the results, as it represents the continuing chain better, but it still isn't enough to recreate the charge density of a glycosidic bond.[143] In this case, the "chain" structure is being used as a model for checking the methods, meaning that using hydroxyl will be sufficient.

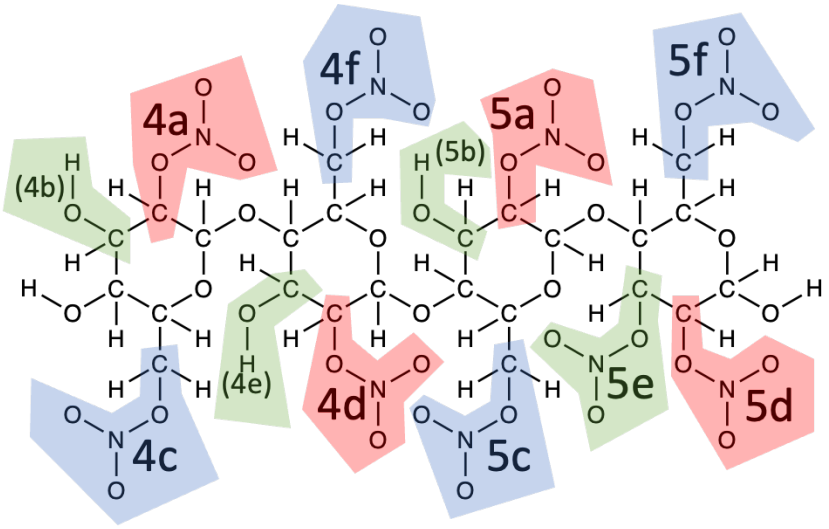


FIGURE 5.1: Skeletal structure of chain unit, containing a 4 nitrated dimer and a 5 nitrated dimer, and capped with hydroxyl groups. The possible nitrated positions are labelled, and 'like' positions are colour coded - for example all of the off-ring positions (**a** and **d** positions) are blue. The non-nitrated positions are highlighted with pale purple, and labelled in brackets.

The final structure that is studied in this thesis is a small block, shown in figure 5.2, of the type developed in Chapter 3. It is a 2 x 2 x 2 paracrystalline structure of 4/5 N nitrocellulose created using the protocol and containing 816 atoms. Most of the calculations were run on the first frame of the production phase, although some were run on other structures, these will be specified.

For the initial calculations, the products were created by removing the two oxygens attached to the nitrogen, changing the nitrogen atom into a hydrogen, and adjusting the positions, so that the -OH bond length was 0.97 Å, and the C-OH bond was 1.41 Å, the equilibrium bond lengths taken from the OPLS-AA force field parameters, as shown in figure 5.3. All the bond angles were maintained from the original structure. This was done using a Python script in Appendix B.

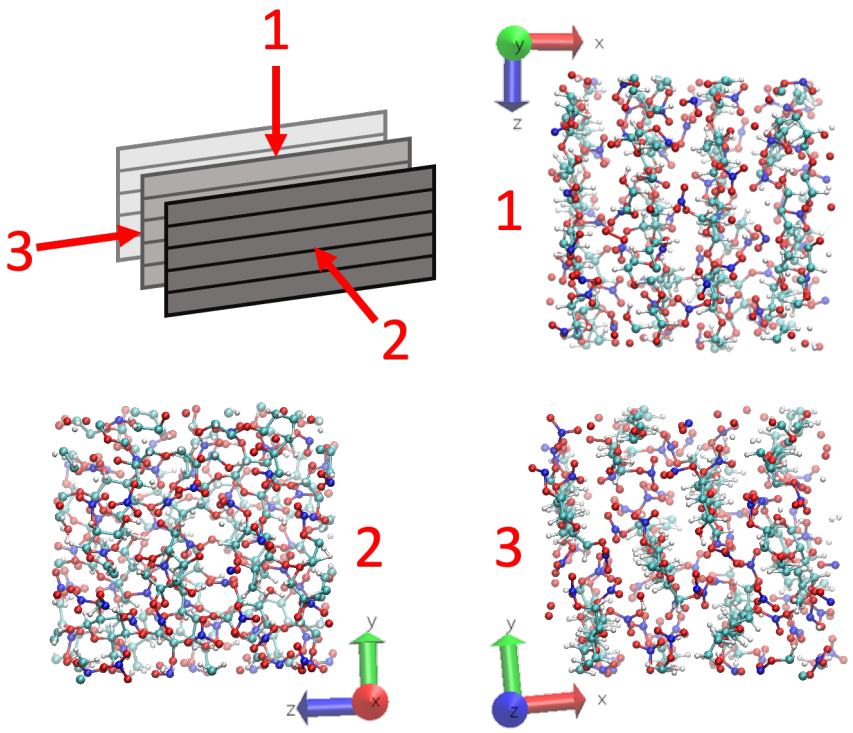


FIGURE 5.2: The structure of the first frame of the 2 x 2 x 2 block that is being used in this chapter.

5.2 Initial Calculations Setup

The aim of this investigation at this point was to apply the calculation of reaction energies techniques, tested in the previous chapter, to the chain and block structures. The values for the reaction energy were calculated for three different reaction pathways, by calculating the energies of each of the separate components shown in figure 5.4. This was repeated for each of the 9 types of nitrate positions on the chain unit, and also 10 frames from the block unit. The energies calculated for the alkaline hydrolysis are shown in figure 5.5.

The positions have been paired up to have similar nitration positions (**4c** and **5c** for example) next to each other for comparison. It is difficult to see from the graphs whether there is any correlation between the positions of the energies. When the orders of the energies are put into a table, as shown in table 5.1, there are a few things that can be seen. **5c** is the highest energy structure in all but two of the structures. The 4 off-ring positions (**c** and **f**) are in the top 4 energy reactions 65 % of the time, showing that there is some correlation between whether the nitrate is on or off-ring, but not a particularly strong one. This supports what we already know that the off-ring positions are more stable (the fact that in the paracrystalline single nitrated systems generated in chapter 3 are formed more easily shows this), as the most stable nitrate positions will require the most energy to be denitrated.

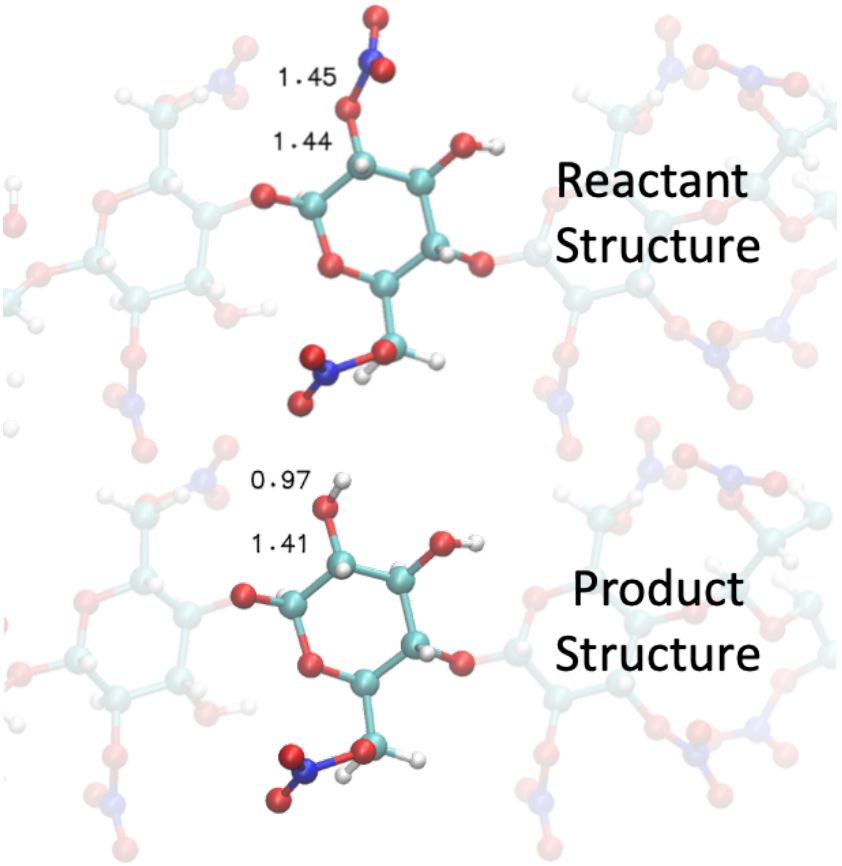


FIGURE 5.3: A direct comparison between a reactant and product structure for the denitration of position **4a** on the chain shown. The product is created by removing the two oxygens attached to the nitrogen, then changing the nitrogen to hydrogen, and adjusting the bond lengths.

TABLE 5.1: Order of reaction energies from lowest to highest, for each of the different structures tested. The positions are coloured based on their position as shown in figure 5.1, with the off-ring positions in blue .

Chain	5a	<	5e	<	4f	<	5d	<	5f	<	4d	<	4c	<	5c	<	4a
0 ns	5a	<	4d	<	5e	<	5d	<	5f	<	4a	<	4f	<	4c	<	5c
2 ns	4f	<	5e	<	4c	<	4d	<	4a	<	5a	<	5f	<	5d	<	5c
4 ns	5d	<	4d	<	4a	<	5e	<	5a	<	5f	<	4c	<	4f	<	5c
6 ns	4f	<	5f	<	4c	<	4a	<	4d	<	5e	<	5a	<	5d	<	5c
8 ns	5d	<	5e	<	4d	<	4a	<	4c	<	5c	<	4f	<	5f	<	5a
10 ns	5e	<	4d	<	5d	<	4a	<	5f	<	4c	<	4f	<	5a	<	5c
Average	4d	<	5d	<	4f	<	5e	<	5a	<	4a	<	4c	<	5f	<	5c

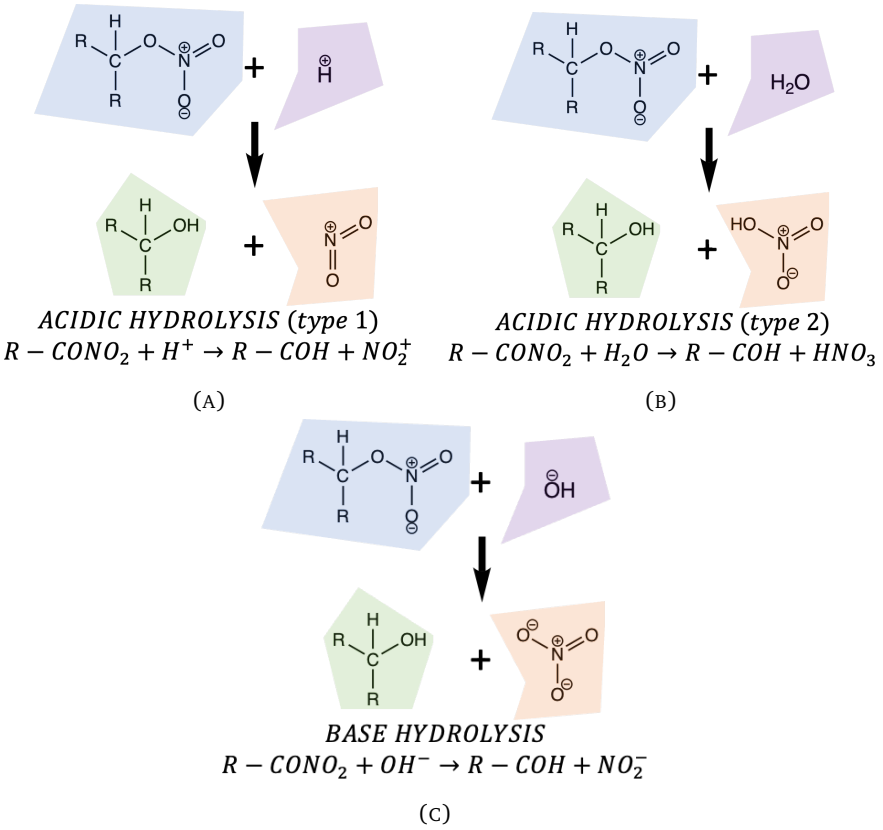


FIGURE 5.4: The components of the three different reactions that are being calculated. The reaction energies are found by geometry optimising the molecule in each of the blocks separately.

The energies initially calculated for these were very high, but also the relative energies calculated for each of the different types of reaction were the same, just shifted due to the energies of the other participant molecules, as we were adding constant energies from the H^+ and NO_2 / H_2O and HNO_3 / OH^- and NO_2^- for every different structure. This should not be the case, the acid hydrolysis and the base hydrolysis should have different reaction energies, due to the different mechanisms, which means there is a flaw in this method of calculation.

5.2.1 Electrostatic potential

Before continuing to calculate reaction energies, particularly in the context of such large systems, it was questioned whether there might be a way to predict which nitrate position is most likely to be most favourable for denitration. This was done by calculating the electrostatic potential for the chain unit, using ONETEP.

The electrostatic potential is calculated using the formula

$$V_{nuc}(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|} \tag{5.1}$$

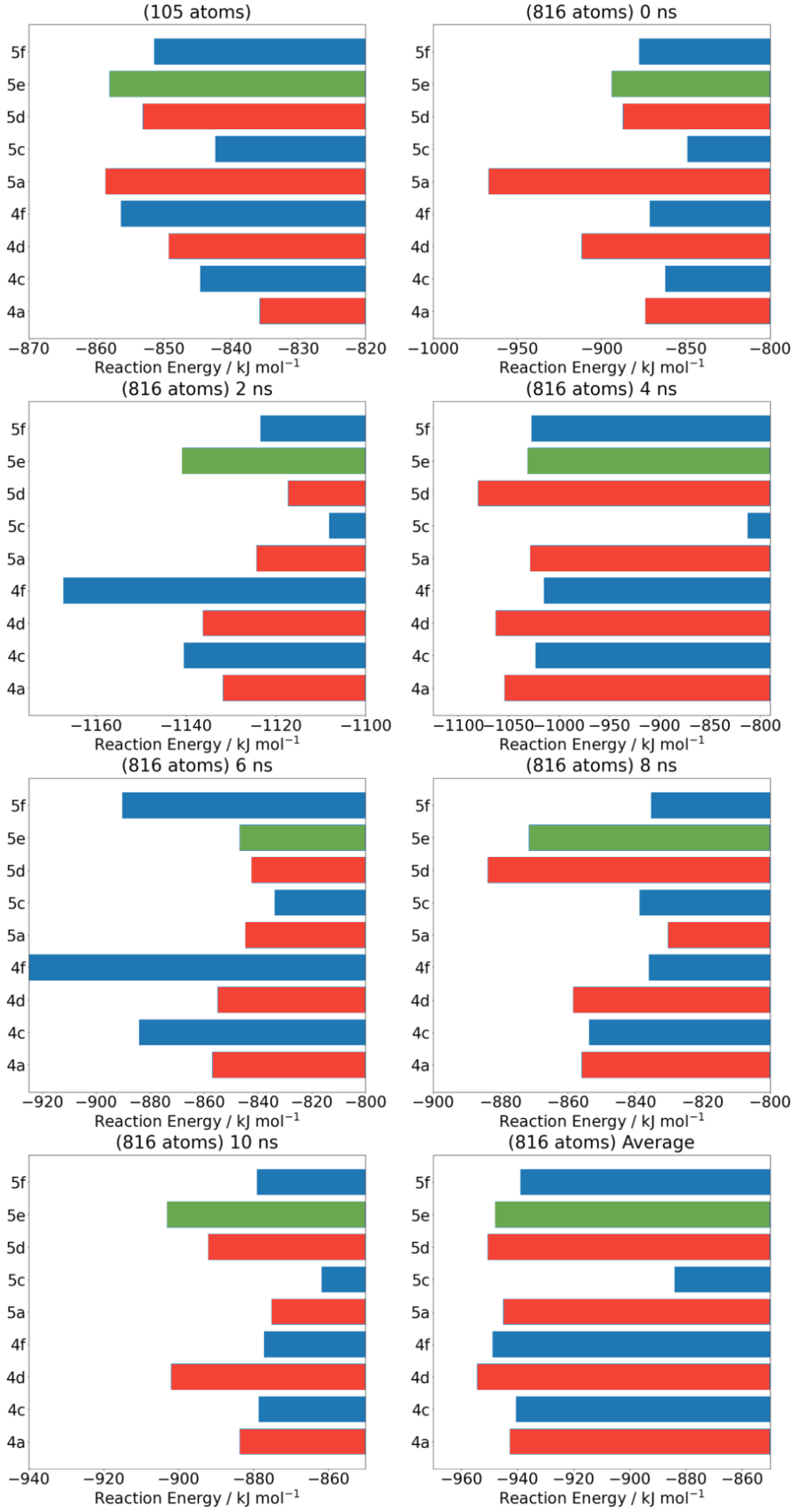


FIGURE 5.5: Reaction energies for selected frames of the chain and the block structure, based on the position that is being denitrated.

where $\mathbf{R}_A$ is the position of atom A, and Z_A is the nuclear charge. Figure 5.6 on the following page shows the electron potentials for the chain unit, the red showing where there is an excess of electrons.

The images in figure 5.6 show the electron potential in the plane of the O-N-O_2 group. For the **4d**, **5d** and **5e** the carbon is further out of the plane so is not labelled. Looking at the images, it can be seen that there is not a big difference depending on the position in the chain, no clear pairing of the position **c** and position **f** nitrates, as has often previous occurred particularly in chapter 3.. However, the positions with the lowest electron density are the positions on the end of the chains, **4a**, **4c**, **5d** and **5f**. This suggests that the relative distance from the end of the chain is what is causing the main effect here. This means that this investigation was unable to give us any meaningful information about which positions might be more likely to react. A longer chain might be able to provide this, although as there is not a significant difference between the inner nitrate positions, it may be that no conclusions can be found from this.

The reaction is happening through an S_N2 reaction, the mechanism of which is shown in figure 5.7 on page 135. This mechanism is based on the S_N2 reactions modelled by Shukla and Hill in their three papers written in 2012, previously discussed in section 2.5.[109,133,134] They discussed that there are two directions from which the OH^- can attack, angular and direct. The direct attack (from the opposite side) was found by them to have lower activation energies, so this was modelled here, as opposed to a direct attack (from the same side).

This means that the reaction is more likely to happen when the C-O bond is more polarisable. This would mean that the most energetically favourable reactions would be likely to happen at a position where there is more electron density at the -O- and less at the carbon.

5.3 Improving the Structures

Looking at the papers published by Shukla and Hill in 2012, it was noticed that the products they used did not directly swap the -NO_2 for an -OH group, but the group was also being changed from an equatorial to axial position. This makes sense for the base hydrolysis, as OH^- is a nucleophile, so this hydrolysis is likely to be an S_N2 reaction (figure 5.7). This means that the product structures being used for these calculations are not representative of the actual products of this denitration reaction, which might be a contributing factor to the reaction energies being not as expected. One of the defining factors of the S_N2 substitution is that the structure is inverted. Figure 5.8 shows how this affects the position of the molecules, meaning it is more complex to create a product. It is possible that the S_N2 reaction could also lead to changes to the

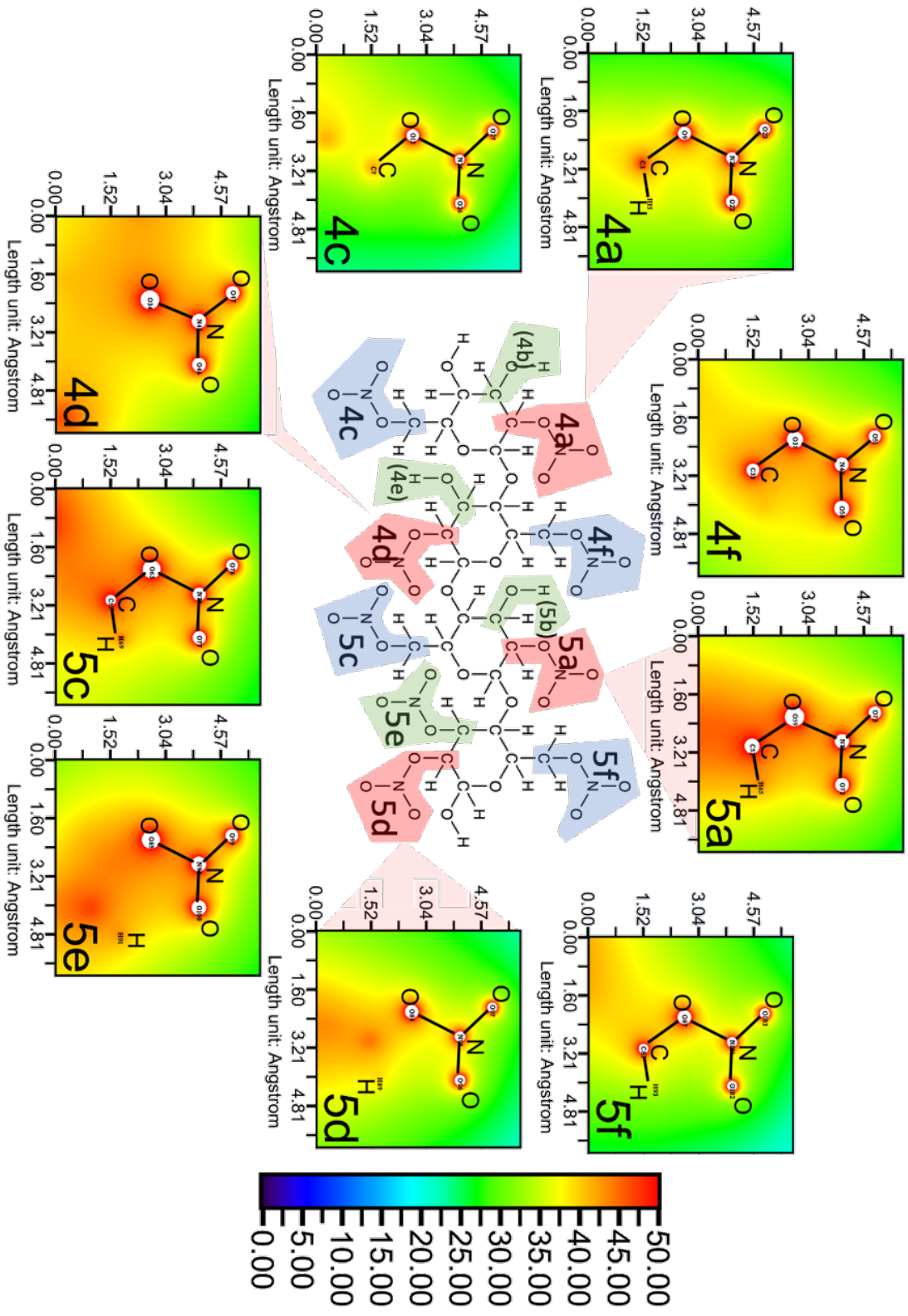


FIGURE 5.6: The electrostatic potential, plotted using MultiWFN. The images are each positioned so that the O-N-O_2 atoms are all in the plane of the image.

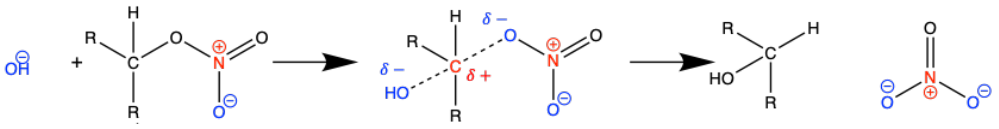


FIGURE 5.7: The mechanism of an S_N2 reaction. The partial bonds shown in the second step show how the bonds are polarised by the nucleophilic attack.

ring conformation, due to the change of functional groups, however, that would require further adaptations to the algorithm, so for now it has been simplified to exclude them.

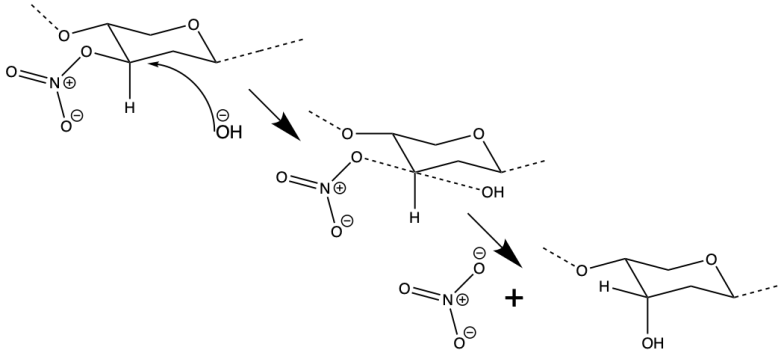


FIGURE 5.8: Mechanism for an S_N2 reaction on a ring in chair comformation. If the leaving group is in an equatorial position, as it is in this case, the group substituted in will be in an axial position. This is happening via a direct attack, as opposed to an angular attack, based on the mechanism investigated by Shukla and Hill [133].

The new products, with inverted chirality, were created for the chain structure using the processes shown in figures 5.9 on the following page and 5.10 on page 137.

For the initial calculations run with these structures the energies found were not very different from the previous results, with energies in the range of 800 - 880 kJ mol^{-1}, see figure 5.11. However this is the correct way for it to be calculated, so these inverted structures were used for further calculations.

The other improvement to the methodology was to try and create a single reactant system including both molecules. The concern with this method was whether it would be feasible to create an optimised system where the incoming group would be interacting with the correct position. The base hydrolysis was tested first as this is the reaction known to proceed by an S_N2 mechanism. Figure 5.12 shows the three systems now being calculated.

At first, the OH^- was added by hand to an already optimised reactant molecule using VMD, and a single-point energy calculation run. The OH^- was positioned approximately 2Å, directly behind the carbons that it was required to attack, as shown in figure 5.13.

The results, in figure 5.14, are more promising - the denitration of the off-chain positions, **c** and **f**, are now at $\sim$ 300 kJ mol^{-1} which is still higher than expected, but it

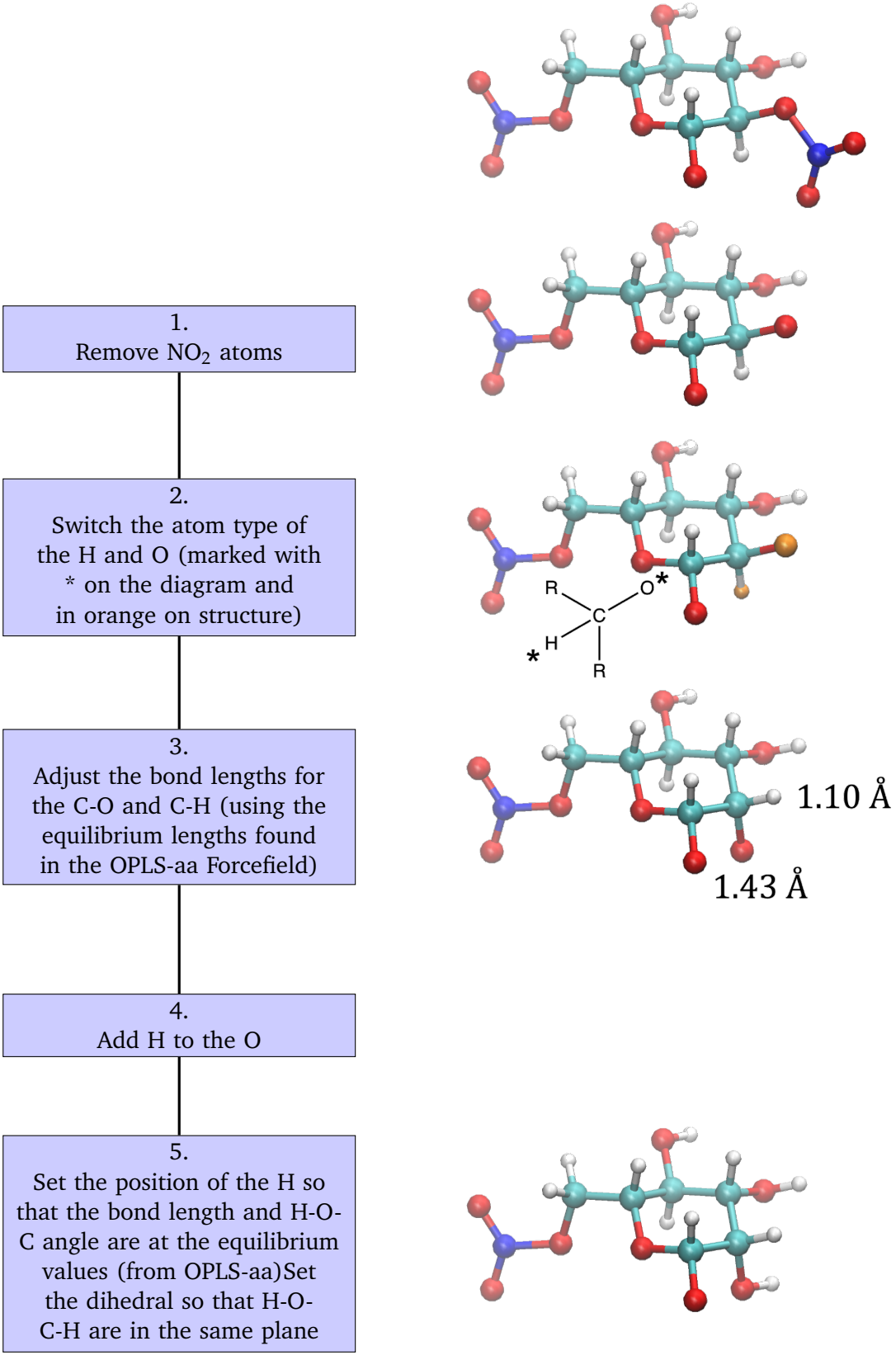


FIGURE 5.9: Procedure for creation of an accurate product in an on-ring position, with images for the denitration of a **4a** position.

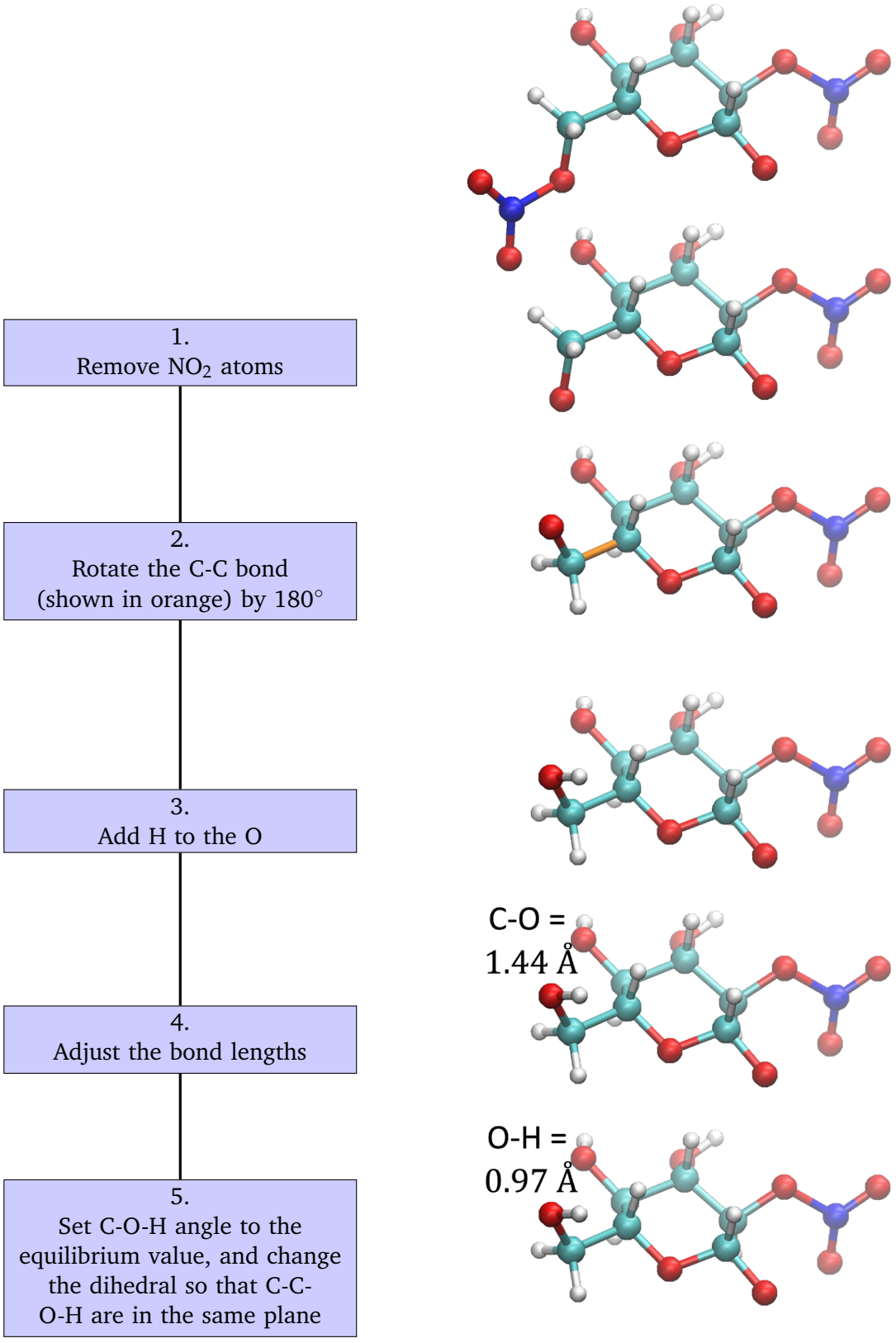


FIGURE 5.10: Procedure for creation of an accurate product in an off-ring position, with images for the denitration of a **4c** position.

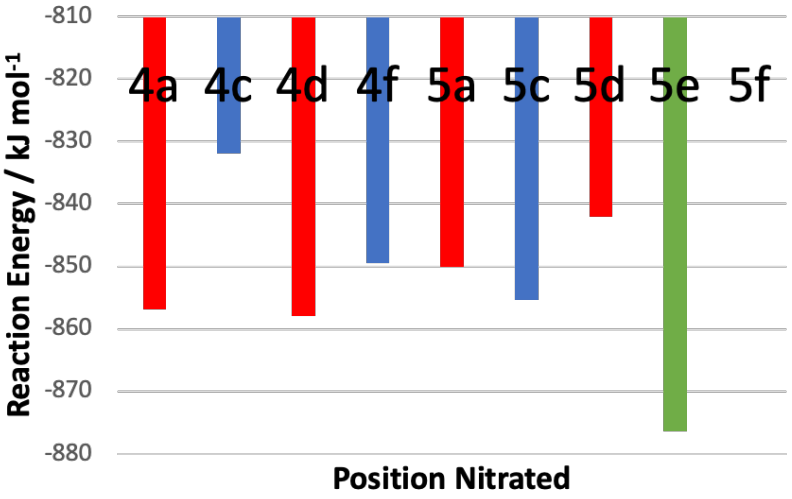


FIGURE 5.11: Results using products with inverted geometry. The **5f** value calculated has been left off as it was much higher.

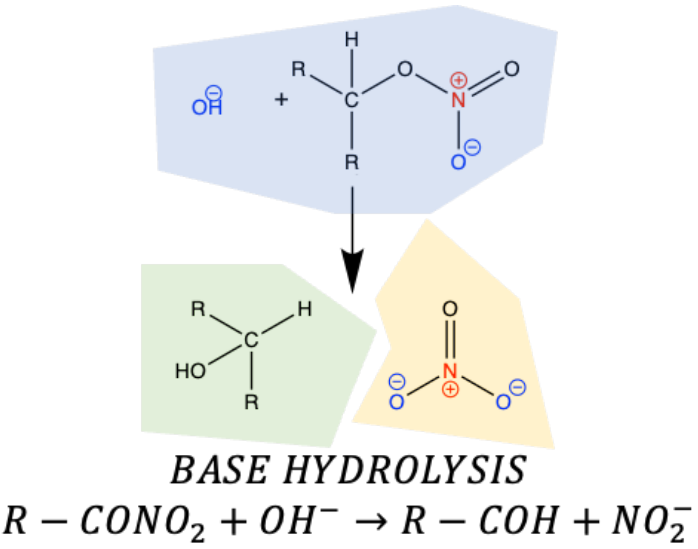


FIGURE 5.12: The components of the base hydrolysis that are now being calculated.

definitely has been reduced. This implied that the concept was correct, but the addition of the OH^- needed to be improved.

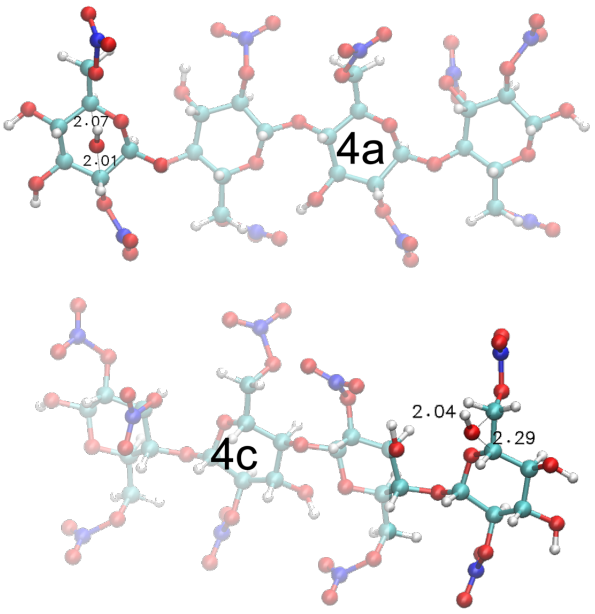


FIGURE 5.13: Illustration of how the OH^- nucleophile was added to the system approximately 2 Å behind the carbon it is going to attack. Other distances measured were based on the geometry details provided in the third paper by Shukla and Hill in 2012 [133]

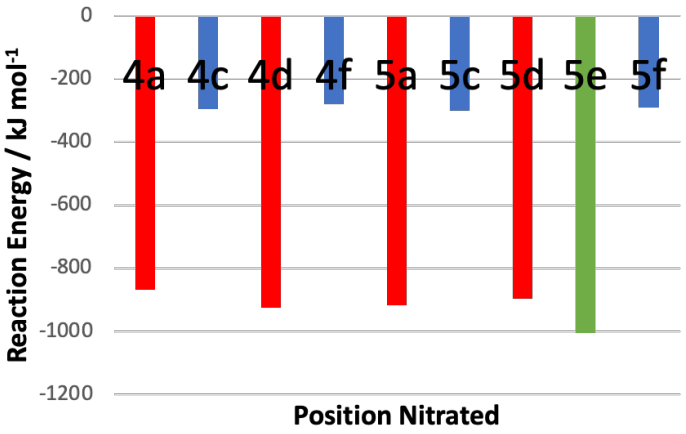


FIGURE 5.14: Results using optimised reactants with OH^- ions added in.

For this, a procedure was developed, using the Python script in Appendix B,to add that OH^- precisely 2 Å away from the carbon it is attacking, by projecting the C-O bond backwards, a process illustrated in figure 5.15.

The products from this were then geometry optimised, and the structures visualised, to check that the OH^- was still in the correct area for the desired denitration. In every case, the OH^- had moved away from the chain molecule, but it was still evident where the nucleophilic attack would happen. The reaction energies now calculated were $\sim$ 125 kJ mol^{-1}, which is much more likely than the previous energies calculated. The energies, and also the distance the OH^- moved are shown in figure 5.16. It can also be

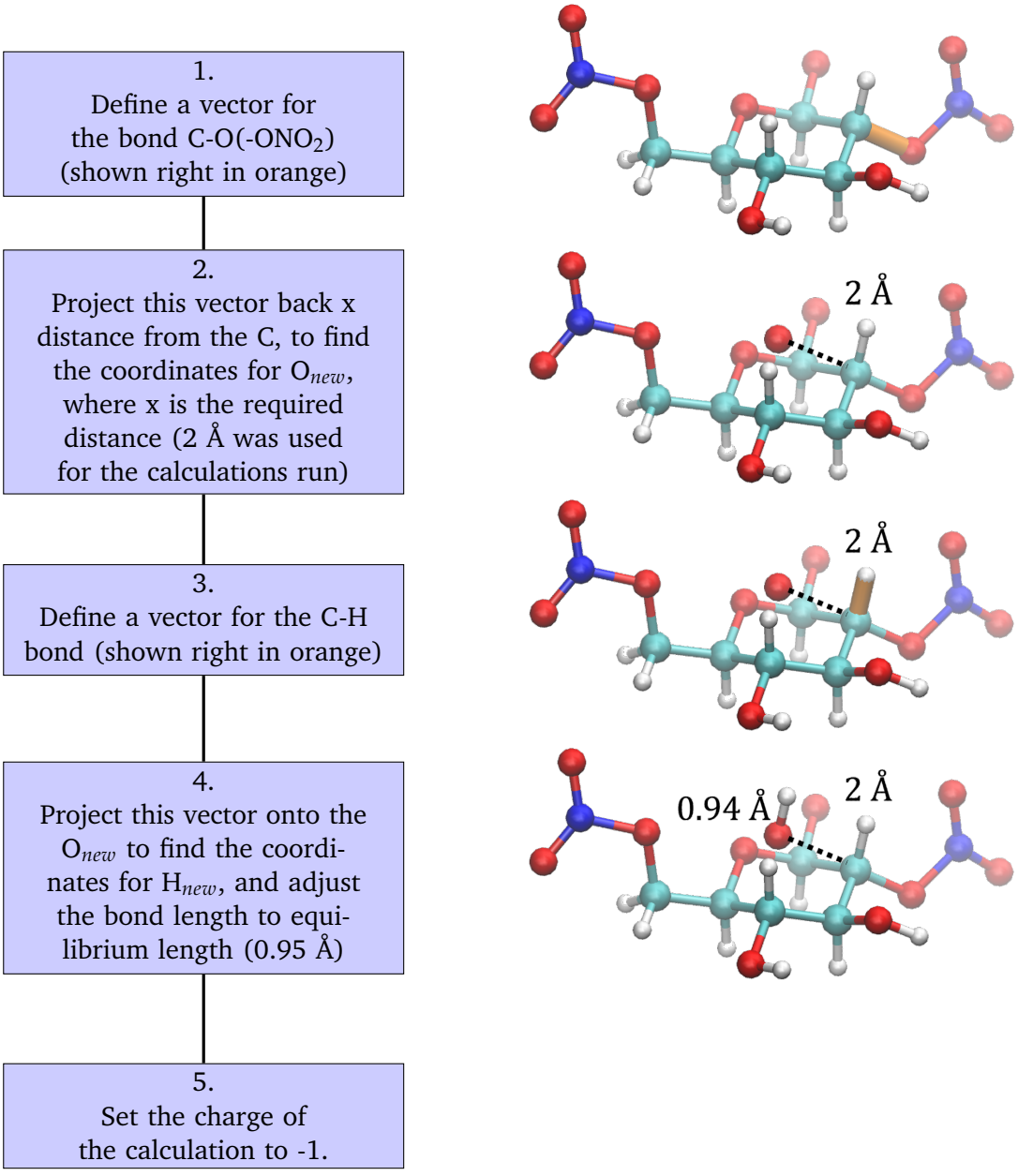


FIGURE 5.15: Procedure for adding OH^- ions to the reactant structures, with images for adding OH^- to attack the **4a** position.

seen that generally, the optimised structures for the off-ring nitrate positions moved the OH^- group further away. This is due to the proximity of the H on the ring, shown in figure 5.17.

The structures now being used are the most chemically accurate, and they are giving reasonable reaction energies, meaning that this method should be able to be used on larger systems to get the reaction energies for large-scale systems as well.

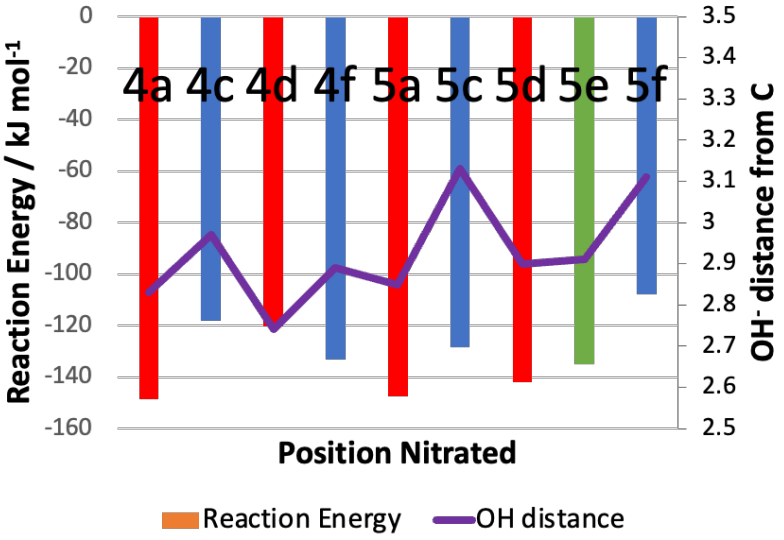
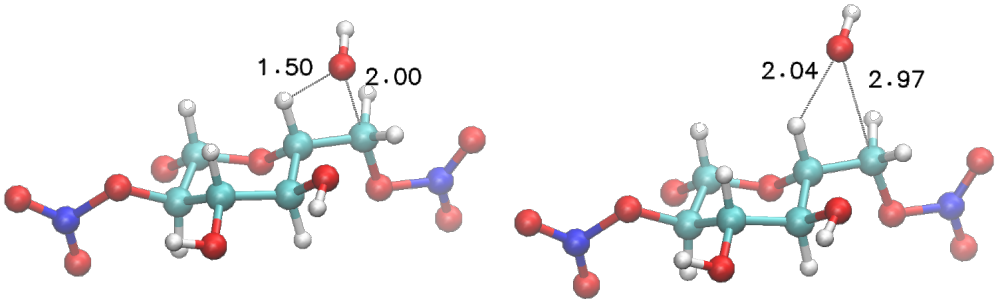


FIGURE 5.16: The reaction energies calculated when the reactants have the OH^- nucleophile added at 2 Å away, and the structures optimised. The purple line plotted also shows how far away the OH^- has moved from the molecule.



(A) Generated position (B) Optimised position

FIGURE 5.17: In the case of the off-ring carbons, the initial position of the OH^- ion is close to one of the on-ring H atoms. This means that in the optimised structure it has moved further away, to avoid this hydrogen.

5.4 Conclusions and further work

This section initially calculated reaction energies for the denitration of different positions in both a chain and also a small block system of 12 wt% nitrocellulose, containing 816 atoms. Comparing the reaction energies showed only a little correlation between the energy and the position that was being denitrated. The difference in energy between the reactants and products was also much larger than expected, which lead to a more careful study of the mechanism that the reaction was expected to proceed by.

By looking at the expected mechanism for base hydrolysis, two improvements were tested to see if they could improve the calculations. First, as the reaction is expected to follow an S_N2 substitution mechanism, the new -OH group should be in the opposite

type of position (axial or equatorial) from the leaving group. Changing this made the calculations more likely to correctly represent the required reaction, but did not make the calculated energies more reasonable.

Secondly, the "reactant" energy was calculated with both the molecule and the OH^- in the same system. Combining the two reactants was not considered previously as there was concern that these systems could not be created in such a way that could provide specificity to the position being denitrated. After promising initial tests where the two molecules were added together by hand with the nitrate position being considered approximately 2 Å away from the OH^- and not optimised, a process was created to add the nucleophile exactly 2 Å behind the point of attack. When these structures were optimised the OH^- remained close enough to the selected carbon to make it clear which position it is heading towards.

Both of these improvements lead to much more realistic reaction energies for calculations performed on the chain unit. However, there was not enough time in this project for the calculations to be performed on the small block system to study the reaction energy in a large-scale system.

The methods for creating the corrected reactants and products have been fully automated using the scripts attached in Appendix B , which means a direct continuation of this work, performing the calculations on block systems of nitrocellulose, would be very fruitful. Also, considering the improvement to the calculation found from optimising the two reactants together in the same system, the impact of optimising the products in a combined system should also be investigated.

Based on current literature this is one of the first attempts to use DFT to study the degradation reactions of nitrocellulose in the context of a bulk nitrocellulose system, as opposed to as up to trimers in a solvent solution.[133] Looking at the nitrocellulose in such a large context (816-atoms) was possible because ONETEP was being used, and this is a linear scaling program.

Chapter 6

Conclusions

The aim of this thesis was to test different methods of simulation to find which approaches could be used to investigate the degradation reactions of nitrocellulose. Over the last few chapters, a method has been developed to create large-scale models of nitrocellulose, and then techniques have been investigated for studying the degradation reactions, which using ONETEP are able to be used on increasingly large systems.

A protocol has been developed to create systems of nitrocellulose of different levels of nitration which reflect realistic systems. The crystallinity of the simulated structures was verified by comparing the RDF structures before and after. They were not fully amorphous, where everything was random. This is because the periodic boundary conditions mean that the chains must go in one side of the simulation cell, and out the other side. In this case, the choice had to be made between using the PBCs to represent an infinite system and being able to make a completely amorphous structure of a limited size. One of the aims of this project was to be able to create systems that DFT calculations could also be performed on, it was decided that creating systems with periodic boundary conditions was more important, as the affect of the bulk was of interest. The paracystallinity of the structures created were deemed sufficient, and the densities were similar enough to experimental data for the structures to be thought to be recreating realistic systems.

By comparing the differently nitrated structures, some conclusions were able to be drawn from this - the more nitrated the system, the “easier” it is to break the crystallinity, leading to the conclusion that the interchain interactions which hold the structure in its crystal form are obstructed by the nitrate groups.

Before running DFT calculations on these large-scale systems, as was the final aim of the thesis, it was decided that the methods should be tested on smaller molecules, in this case nitrate esters, to see how this can be used to give us information about the

energy of the reaction. Transition state searches were used to make estimates of the hydrolysis reaction on nitrate esters, and the study of solvent models on this was also measured. The activation energies for these reactions were calculated to be around 48 kJ mol^{-1}, which is a little lower than literature values for nitrocellulose, although it is reasonable, as the nitrate esters being studied were smaller.

The homolytic fission reaction was also studied using DFT, and information about the Coulson-Fischer point was found, although this did not provide further information about the reaction energy.

Finally, DFT was performed on the unit used to create the structures in the previous chapter and found that the unit is in a low-energy configuration, which supports its use as a model system.

Lastly, some of these methods were applied to larger structures, and by improving the method used, reaction energies of - 100-150 kJ mol^{-1} were calculated, a value that aligns with what is expected based on literature, which is promising evidence that as these calculations can provide valuable information when applied to increasingly large systems.

6.1 Further Work

There are many aspects of this work that could be taken further and either completed or investigated to deepen the understanding of the system required.

1. Compare results from other force fields. This work only uses the OPLS-aa force field, as it had been used for similar work, and it was implemented in LAMMPS, which was the program being used. However, there are other force fields that have been used on nitrocellulose that could be tested to see if they give any better results.

2. Capping the chains to see how finite chains behave. It has been theorised that the systems created for this thesis are not fully amorphous, as the periodic boundary conditions require every chain to be the same length and to go in one side of the simulation cell and out the other. The aim of this project was to simulate structures that represented bulk systems by using PBCs, so this was prioritised over creating a fully amorphous system. However, it would be interesting to test this by creating systems with capped chains that aren't required to pass through the boundaries, to see how they react and how amorphous they become.

3. Investigating how the system size affects the crystallinity. In the last point, it was mentioned that the boundary conditions do limit the amount of randomness in

the system. Increasing the size of the system can mitigate this issue, but again, the point of the PBCs is to be able to approximate a larger system without increasing the calculation size as much. A more thorough investigation into the system size could find the optimal system that would give a required level of paracrystallinity, whilst minimising the calculation time.

4. Run annealing cycles on all structures. When the cellulose structure remained crystalline, after all the nitrocellulose structures became at least somewhat paracrystalline after the protocol was performed, further annealing steps were run. This involved heating and rapidly cooling the cellulose structure to see if this could break the crystallinity. It was seen that the structure did become a little more paracrystalline at the higher temperature but always returned to crystalline when reduced to room temperature. There are a lot of ways these cycles could be adapted - the maximum temperature increase, changing the pressure as well, using different timescales, so it would have been interesting to investigate further these cycles to try and create paracrystalline cellulose. It also would have been interesting to test the annealing cycles on more of the structures and see how they affect the crystallinity of the already paracrystalline nitrocellulose structures.

5. Create structures with randomised nitration patterns. One limitation of the large-scale models, when compared to real systems, is that due to how the structures were initially generated the nitrated positions are in a repeated pattern, whereas in natural nitrocellulose it is likely to be random. When the calculations were being run this was an acknowledged limitation of the program being used (Moltemplate), but it may be possible with other packages. If randomisation of the initial units was possible this could also improve the models.

6. Further investigation into 3Ncfe. The preliminary investigation into why a phase shift was seen in the 3Ncfe structures shows that a small arbitrary shift in one atom in each unit can prevent this. Further investigation into why this is the case could be of interest. This also has implications for the other structures, are all of the structures as inherently unstable - a tiny change could change the model significantly - and it is just fortunate that for the other structures a working initial structure was created each time.

7. Using a reactive force field like ReaxxFF to study the reactions. A reactive force field can be used for studying the breaking and making of chemical bonds, meaning that one could be used to study the break-down reactions being focused on in this project, and the results compared to those found using DFT.

8. A custom force field could be created by using DFT calculations to create potential energy surfaces for the different parameters. This could be done using smaller molecules such as NO_2, nitrate esters and nitrocellulose dimers. Parameters could also be taken from a combination of other force fields, as long as they have been

obtained through the same methodology. The advantage of this is that it would mean the force field is directly tailored to nitrocellulose structures.

9. Testing of nitroceullose monomers and dimers. Although a lot of work was put into investigating nitrate esters, as smaller molecules that contain a nitrate group, the next step was to apply the methods straight to bulk nitrocellulose. This caused issues as the introduction of the ring affected how the reactant and product structures should be created. Investigations into monomers and dimers of nitrocellulose by Shukla and Hill, but calculations into these based on the methods used in this thesis could be run and might lead to refining the methods for the large-scale systems further.[109,133,134]

10. Investigation of the pyrolysis reaction. This project mostly focused on the thermolysis and hydrolysis reactions of nitrocellulose. The pyrolysis reactions are quite well understood for cellulose, with mechanisms well defined. A potentially interesting path of research would be to test these mechanisms, such as those found in reference [94],[97] and apply them to nitrocellulose.

11. The investigation of homolytic fission. Finding the Coulson-Fischer point of the reaction was interesting, however, it was not pursued further as it was decided that it strayed from the original remit of the project. However, it could be an interesting area for further investigation.

12. Application of the algorithms developed in Chapter 5. By using the scripts to create the correct reactants and products, accurate calculations should now be possible to run to calculate reaction energies for large-scale systems. This can be tested on the 816-atom structures used in Chapter 5, but if this produces results that show there is a correlation to be studied between the reaction energy and the nitrate position, then the methods could also be tested on the other structures created by the protocol, and also on the 52,000-atom systems created in Chapter 3.

13. Taking the calculations further and performing transition state searches on the large-scale structures. The calculations of reaction energy can help us compare different nitrate positions, but finding transition states would mean that we could find rate constants, and therefore understand how the conditions can affect the rate of degradation. This will require transition state searches such as NEB to be run on nitrocellulose structures as opposed to nitrate esters. For base hydrolysis, the understanding of the reaction to be S_N2 can also help, as an estimated transition state can be found based on what we know of this mechanism. The positioning of the OH^- for the coordinated transition state can be informed by its position in the optimised structures shown in figure 5.16 on page 141. Acidic hydrolysis will need more consideration, as it is not going to proceed by nucleophilic substitution.

This project was entirely computational, as it aimed to study chemicals that are highly reactive and not easy to study experimentally. However, it might have been beneficial to validate the methods used in this thesis by comparing simulations and experimental data for less reactive substances.

Appendix A

Appendix: Chapter 3

A.1 Additional Parameters for OPLS-aa

Additional parameters were needed to describe the nitrate groups of nitrocellulose. The values illustrated in figure A.1 were found using LigParGen to obtained parameters for HNO_3.

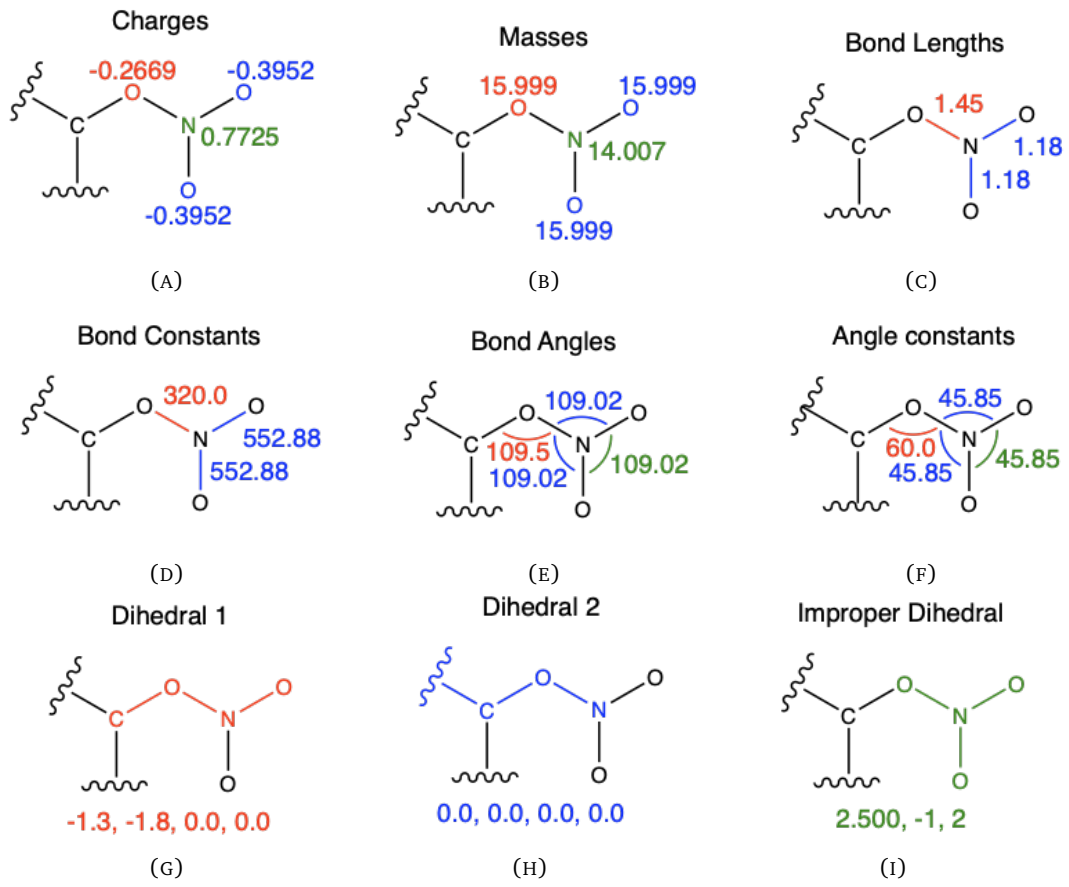


FIGURE A.1: Parameters for ONO_2 groups

To add these parameters to the forcefield, the following lines were added to the oplsaa.lt file provided on the MolTemplate GitHub.

```
OPLSAA {
    write_once("In Init") {
    ...
    improper_style cvff
    ...
    } #(end of init parameters
    write_once("In Charges") {
        ...
        set type @atom:907 charge  0.7725   # N
        set type @atom:908 charge -0.3952   # -O
        set type @atom:909 charge -0.2669   # -O-
        ...
    } #(end of atom partial charges)

    write_once("Data Masses") {
        ...
        @atom:907 14.007
        @atom:908 15.999
        @atom:909 15.999
```

```
    ...
} #(end of atom masses)

...
replace{ @atom:907 @atom:907_b111_a111_d111_i111 }
replace{ @atom:908 @atom:908_b112_a112_d112_i112 }
replace{ @atom:909 @atom:909_b020_a020_d020_i020 }
...

write_once("In Settings") {
    ...
    pair_coeff @atom:907_b111_a111_d111_i111 @atom:907
_b111_a111_d111_i111 0.17 3.3
    pair_coeff @atom:908_b112_a112_d112_i112 @atom:908
_b112_a112_d112_i112 0.17 2.96
    pair_coeff @atom:909_b020_a020_d020_i020 @atom:909
_b020_a020_d020_i020 0.14 2.9
    ...
} #(end of pair_coeffs)

write_once("In Settings") {
    ...
    bond_coeff @bond:020_111 320.00 1.45
    bond_coeff @bond:111_112 552.88 1.18
    ...
} #(end of bond_coeffs)

write_once("Data Bonds By Type") {
    ...
    @bond:020_111 @atom:*_b020_a*_d*_i* @atom:*_b111_a*_d*_i*
    @bond:111_112 @atom:*_b111_a*_d*_i* @atom:*_b112_a*_d*_i*
    ...
} #(end of bonds by type)

write_once("In Settings") {
    ...
    angle_coeff @angle_013_020_111  60.00 190.50
    angle_coeff @angle_020_111_112  45.85 109.02
    angle_coeff @angle_112_111_112. 45.85 109.02
    ...
} #(end of angle_coeffs)

write_once("Data Angles By Type") {
    ...
    @angle:013_020_111 @atom:*_b*_a013_d*_i* @atom:*_b*_a020_d*_i*
@atom:*_b*_a111_d*_i*
    @angle:020_111_112 @atom:*_b*_a020_d*_i* @atom:*_b*_a111_d*_i*
@atom:*_b*_a112_d*_i*
    @angle:112_111_112 @atom:*_b*_a112_d*_i* @atom:*_b*_a111_d*_i*
@atom:*_b*_a112_d*_i*
    ...
```

```
    } #(end of angles by type)

    write_once("In Settings") {
        ...
        dihedral_coeff @dihedral:112_111_020_013 -1.3 -1.8  0.0  0.0
        dihedral_coeff @dihedral:111_020_013_046  0.0  0.0  0.0  0.0
        ...
    } #(end of dihedral_coeffs)

    write_once("Data Dihedrals By Type") {
        ...
        @dihedral:112_111_020_013 @atom:*_b*_a*_d112_i* @atom:*_b*_a*
    _d111_i* @atom:*_b*_a*_d020_i* @atom:*_b*_a*_d013_i*
        @dihedral:111_020_013_046 @atom:*_b*_a*_d111_i* @atom:*_b*_a*
    _d020_i* @atom:*_b*_a*_d013_i* @atom:*_b*_a*_d046_i*
        ...
    } #(end of dihedrals by type)

    write_once("In Settings") {
        ...
        improper_coeff @improper:020_112_111_112 2.500 -1 2 # (
    moltemplate)
        ...
    } #(end of improper_coeffs)

    write_once("Data Impropers By Type (opls_imp.py)") {
        ...
        @improper:020_112_111_112 @atom:*_b*_a*_d*_i020* @atom:*_b*_a*_d*
    _i112* @atom:*_b*_a*_d*_i111* @atom:*_b*_a*_d*_i112*
        ...
    } #(end of impropers by type)
}
```

LISTING A.1: Lines added to OPLSAA file provided on the Moltemplete GitHub

A.2 Unit Cell Moltemplate Input File

```
import "my_oplsaa.lt"

cUnit_xyz inherits OPLSAA {

  write('Data Atoms') {
    $atom:01C   $mol @atom:135  0.00     0.148  -0.342   0.449
    $atom:02C   $mol @atom:100  0.00    -0.032  -1.499  -0.536  # aCell
    $atom:02C   $mol @atom:125  0.00    -0.032  -1.499  -0.536  # aNCell
    $atom:03C   $mol @atom:100  0.00     0.439  -1.116  -1.918  # bCell
    $atom:03C   $mol @atom:125  0.00     0.439  -1.116  -1.918  # bNCell
    $atom:04C   $mol @atom:911  0.00    -0.082   0.244  -2.336
    $atom:05C   $mol @atom:125  0.00     0.055   1.296  -1.296
```

```
    $atom:06C   $mol @atom:99   0.00    -0.662   2.599  -1.588  # cCell
    $atom:06C   $mol @atom:124  0.00    -0.662   2.599  -1.588  # cNCell
    $atom:01O   $mol @atom:122  0.00    -0.535  -0.701   1.636
    $atom:02O   $mol @atom:96   0.00     0.766  -2.559  -0.003  # aCell
    $atom:02O   $mol @atom:909  0.00     0.766  -2.559  -0.003  # aNCell
    $atom:03O   $mol @atom:96   0.00     0.032  -2.127  -2.864  # bCell
    $atom:03O   $mol @atom:909  0.00     0.032  -2.127  -2.864  # bNCell
    $atom:04O   $mol @atom:122  0.00     0.535   0.701  -3.554
    $atom:05O   $mol @atom:122  0.00    -0.528   0.807  -0.055
    $atom:06O   $mol @atom:96   0.00    -0.001   3.284  -2.637  # cCell
    $atom:06O   $mol @atom:909  0.00    -2.049   2.357  -1.743  # cNCell
    $atom:01H   $mol @atom:136  0.00     1.093  -0.153   0.621
    $atom:02H   $mol @atom:118  0.00    -0.972  -1.774  -0.567  # aCell
    $atom:02H   $mol @atom:127  0.00    -0.972  -1.774  -0.567  # aNCell
    $atom:03H   $mol @atom:118  0.00     1.775  -1.075  -1.898  # bCell
    $atom:03H   $mol @atom:127  0.00     1.775  -1.075  -1.898  # bNCell
    $atom:04H   $mol @atom:127  0.00    -1.041   0.143  -2.517
    $atom:05H   $mol @atom:127  0.00     1.004   1.484  -1.131
    $atom:06H   $mol @atom:118  0.00    -1.576   2.411  -1.846  # cCell
    $atom:06H   $mol @atom:127  0.00    -0.518   3.215  -0.855  # cNCell
    $atom:07H   $mol @atom:118  0.00    -0.676   3.155  -0.794  # cCell
    $atom:07H   $mol @atom:118  0.00    -0.308   2.998  -2.398  # cNCell
    $atom:08H   $mol @atom:97   0.00     0.581  -2.603   0.955  # aCell
    $atom:01N   $mol @atom:907  0.00     2.168  -2.028   0.042  # aNCell
    $atom:12O   $mol @atom:908  0.00     3.221  -2.839   0.608  # aNCell
    $atom:13O   $mol @atom:908  0.00     2.467  -0.713  -0.474  # aNCell
    $atom:09H   $mol @atom:97   0.00     0.593  -1.956  -3.651  # bCell
    $atom:02N   $mol @atom:907  0.00    -1.464  -2.232  -2.870  # bNCell
    $atom:14O   $mol @atom:908  0.00    -2.196  -1.555  -1.825  # bNCell
    $atom:15O   $mol @atom:908  0.00    -2.172  -3.016  -3.856  # bNCell
    $atom:11H   $mol @atom:97   0.00    -0.569   4.012  -2.958  # cCell
    $atom:03N   $mol @atom:907  0.00    -2.822   3.298  -0.867  # cNCell
    $atom:16O   $mol @atom:908  0.00    -4.266   3.269  -0.875  # cNCell
    $atom:17O   $mol @atom:908  0.00    -2.123   4.234  -0.017  # cNCell
    $atom:07C   $mol @atom:135  0.00    -0.148   0.342   5.639
    $atom:08C   $mol @atom:100  0.00     0.032   1.499   4.654  # dCell
    $atom:08C   $mol @atom:125  0.00     0.032   1.499   4.654  # dNCell
    $atom:09C   $mol @atom:100  0.00    -0.439   1.116   3.272  # eCell
    $atom:09C   $mol @atom:125  0.00    -0.439   1.116   3.272  # eNCell
    $atom:10C   $mol @atom:911  0.00     0.082  -0.244   2.855
    $atom:11C   $mol @atom:125  0.00    -0.055  -1.296   3.911
    $atom:12C   $mol @atom:99   0.00     0.662  -2.599   3.602  # fCell
    $atom:12C   $mol @atom:124  0.00     0.662  -2.599   3.602  # fNCell
    $atom:08O   $mol @atom:96   0.00    -0.766   2.559   5.187  # dCell
    $atom:08O   $mol @atom:909  0.00    -0.766   2.559   5.187  # dNCell
    $atom:09O   $mol @atom:96   0.00    -0.032   2.127   2.326  # eCell
    $atom:09O   $mol @atom:909  0.00    -0.032   2.127   2.326  # eNCell
    $atom:10O   $mol @atom:122  0.00     0.528  -0.807   5.135
    $atom:11O   $mol @atom:96   0.00     0.001  -3.284   2.553  # fCell
    $atom:11O   $mol @atom:909  0.00     2.064  -2.394   3.586  # fNCell
    $atom:12H   $mol @atom:136  0.00    -1.093   0.153   5.811
```

```
    $atom:13H   $mol @atom:118  0.00    0.972   1.774   4.623  # dCell
    $atom:13H   $mol @atom:127  0.00    0.972   1.774   4.623  # dNCell
    $atom:14H   $mol @atom:118  0.00   -1.420   1.078   3.280  # eCell
    $atom:14H   $mol @atom:127  0.00   -1.420   1.078   3.280  # eNCell
    $atom:15H   $mol @atom:127  0.00    1.041  -0.143   2.673
    $atom:16H   $mol @atom:127  0.00   -1.004  -1.484   4.059
    $atom:17H   $mol @atom:118  0.00    1.576  -2.411   3.334  # fCell
    $atom:17H   $mol @atom:127  0.00    0.444  -3.258   4.280  # fNCell
    $atom:18H   $mol @atom:118  0.00    0.676  -3.155   4.396  # fCell
    $atom:18H   $mol @atom:127  0.00    0.371  -2.929   2.738  # fNCell
    $atom:20H   $mol @atom:97   0.00   -0.581   2.603   6.145  # dCell
    $atom:04N   $mol @atom:907  0.00    0.122   3.697   5.596  # dNCell
    $atom:18O   $mol @atom:908  0.00    1.550   3.647   5.386  # dNCell
    $atom:19O   $mol @atom:908  0.00   -0.478   4.857   6.212  # dNCell
    $atom:21H   $mol @atom:97   0.00   -0.593   1.956   1.539  # eCell
    $atom:05N   $mol @atom:907  0.00    1.460   2.096   2.172  # eNCell
    $atom:20O   $mol @atom:908  0.00    2.288   2.961   1.184  # eNCell
    $atom:21O   $mol @atom:908  0.00    2.288   1.214   2.961  # eNCell
    $atom:22H   $mol @atom:97   0.00    0.569  -4.012   2.232  # fCell
    $atom:06N   $mol @atom:907  0.00    2.518  -1.944   4.943  # fNCell
    $atom:22O   $mol @atom:908  0.00    1.528  -1.818   5.986  # fNCell
    $atom:23O   $mol @atom:908  0.00    3.906  -1.660   5.226  # fNCell
  }

  write('Data Bond List') {
    $bond:01C02C        $atom:01C       $atom:02C
    $bond:01C01O        $atom:01C       $atom:01O
    $bond:01C05O        $atom:01C       $atom:05O
    $bond:01C01H        $atom:01C       $atom:01H
    $bond:02C03C        $atom:02C       $atom:03C
    $bond:02C02O        $atom:02C       $atom:02O
    $bond:02C02H        $atom:02C       $atom:02H
    $bond:03C04C        $atom:03C       $atom:04C
    $bond:03C03O        $atom:03C       $atom:03O
    $bond:03C03H        $atom:03C       $atom:03H
    $bond:04C05C        $atom:04C       $atom:05C
    $bond:04C04O        $atom:04C       $atom:04O
    $bond:04C04H        $atom:04C       $atom:04H
    $bond:05C06C        $atom:05C       $atom:06C
    $bond:05C05O        $atom:05C       $atom:05O
    $bond:05C05H        $atom:05C       $atom:05H
    $bond:06C06O        $atom:06C       $atom:06O
    $bond:06C06H        $atom:06C       $atom:06H
    $bond:06C07H        $atom:06C       $atom:07H
    $bond:02O08H        $atom:02O       $atom:08H       # aCell
    $bond:02O01N        $atom:02O       $atom:01N       # aNCell
    $bond:01N12O        $atom:01N       $atom:12O       # aNCell
    $bond:01N13O        $atom:01N       $atom:13O       # aNCell
    $bond:03O09H        $atom:03O       $atom:09H       # bCell
    $bond:03O02N        $atom:03O       $atom:02N       # bNCell
    $bond:02N14O        $atom:02N       $atom:14O       # bNCell
```

```
115     $bond:02N15O        $atom:02N       $atom:15O       # bNCell
116     $bond:06O11H        $atom:06O       $atom:11H       # cCell
117     $bond:06O03N        $atom:06O       $atom:03N       # cNCell
118     $bond:03N16O        $atom:03N       $atom:16O       # cNCell
119     $bond:03N17O        $atom:03N       $atom:17O       # cNCell
120     $bond:07C08C        $atom:07C       $atom:08C
121     $bond:07C10O        $atom:07C       $atom:10O
122     $bond:07C12H        $atom:07C       $atom:12H
123     $bond:08C09C        $atom:08C       $atom:09C
124     $bond:08C08O        $atom:08C       $atom:08O
125     $bond:08C13H        $atom:08C       $atom:13H
126     $bond:09C10C        $atom:09C       $atom:10C
127     $bond:09C09O        $atom:09C       $atom:09O
128     $bond:09C14H        $atom:09C       $atom:14H
129     $bond:10C11C        $atom:10C       $atom:11C
130     $bond:10C01O        $atom:10C       $atom:01O
131     $bond:10C15H        $atom:10C       $atom:15H
132     $bond:11C12C        $atom:11C       $atom:12C
133     $bond:11C10O        $atom:11C       $atom:10O
134     $bond:11C16H        $atom:11C       $atom:16H
135     $bond:12C11O        $atom:12C       $atom:11O
136     $bond:12C17H        $atom:12C       $atom:17H
137     $bond:12C18H        $atom:12C       $atom:18H
138     $bond:08O20H        $atom:08O       $atom:20H       # dCell
139     $bond:08O04N        $atom:08O       $atom:04N       # dNCell
140     $bond:04N18O        $atom:04N       $atom:18O       # dNCell
141     $bond:04N19O        $atom:04N       $atom:19O       # dNCell
142     $bond:09O21H        $atom:09O       $atom:21H       # eCell
143     $bond:09O05N        $atom:09O       $atom:05N       # eNCell
144     $bond:05N20O        $atom:05N       $atom:20O       # eNCell
145     $bond:05N21O        $atom:05N       $atom:21O       # eNCell
146     $bond:11O22H        $atom:11O       $atom:22H       # fCell
147     $bond:11O06N        $atom:11O       $atom:06N       # fNCell
148     $bond:06N22O        $atom:06N       $atom:22O       # fNCell
149     $bond:06N23O        $atom:06N       $atom:23O       # fNCell
150   }
151
152 } # cUnit_xyz
```

LISTING A.2: moltemplate input file to create single unit of nitrocellulose.

The boundary conditions were dealt with by creating boundary units with only the atoms that belong on that side of the border. For example, atoms on the face **x(0)** in figure A.2 (value 0 in the x column in table A.1) were written into a unit labelled **0yz**, and on the **x(1)** face contains atoms written in a unit names **1yz**. All bonds through the boundary had to be specified in the chain / plane files.

A.3 Block Structure Moltemplate Input Files

```
import "XN_unit_xyz.lt" # load in the unit cell
import "XN_unit_xy0a.lt"
import "XN_unit_xy1a.lt"

import "new_oplsaa.lt"

cChain_xyza inherits OPLSAA {

  create_var {$mol}

  # Create the chain arranged along the x-axis

  unit_xyz01 = new cUnit_xy0
  unit_xyz02 = new cUnit_xyz.move(0.00, 0.00, 10.38)
  unit_xyz03 = new cUnit_xy1.move(0.00, 0.00, 20.76)

  # List the inter-unit bonds
  # (Angles, dihedrals, impropers will be automatically added later.)

  write('Data Bond List') {
# boundary bonds
    $bond:01C02C        $atom:unit_xyz01/01C    $atom:unit_xyz03/02C
    $bond:01C05O        $atom:unit_xyz01/01C    $atom:unit_xyz03/05O
    $bond:02O08H        $atom:unit_xyz03/02O    $atom:unit_xyz01/08H    #
     aCell
    $bond:02O01N        $atom:unit_xyz03/02O    $atom:unit_xyz01/01N    #
     aNCell
    $bond:01N13O        $atom:unit_xyz01/01N    $atom:unit_xyz03/13O    #
     aNCell
```

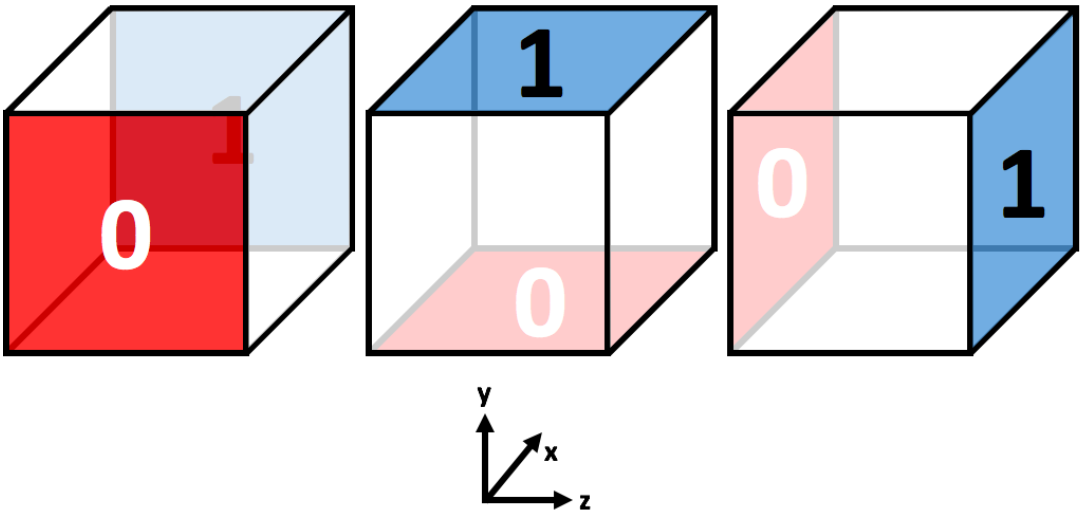


FIGURE A.2: Labels of the boundary

TABLE A.1: List of which boundaries each of the atoms sits on, based on figure A.2

Atom	x (0/1)	y (0/1)	z (0/1)	Atom	x (0/1)	y (0/1)	z (0/1)
01C	0	1	0	**07C**	1	0	0
02C (# aCell)	1	1	1	**08C (# dCell)**	0	0	0
02C (# aNCell)	1	1	1	**08C (# dNCell)**	0	0	0
03C (# bCell)	0	1	1	**09C (# eCell)**	1	0	0
03C (# bNCell)	0	1	1	**09C (# eNCell)**	1	0	0
04C	1	0	1	**10C**	0	1	0
05C	0	0	1	**11C**	1	1	0
06C (# cCell)	1	0	1	**12C (# fCell)**	0	1	0
06C (# cNCell)	1	0	1	**12C (# fNCell)**	0	1	0
01O	1	1	0	**08O (# dCell)**	1	0	0
02O (# aCell)	0	1	1	**08O (# dNCell)**	1	0	0
02O (# aNCell)	0	1	1	**09O (# eCell)**	0	0	0
03O (# bCell)	1	1	1	**09O (# eNCell)**	0	0	0
03O (# bNCell)	1	1	1	**10O**	0	1	0
04O	0	0	1	**11O (# fCell)**	0	1	0
05O	1	0	1	**11O (# fNCell)**	0	1	0
06O (# cCell)	1	0	1	**12H**	1	0	0
06O (# cNCell)	1	0	1	**13H (# dCell)**	0	0	0
01H	0	1	0	**13H (# dNCell)**	0	0	0
02H (# aCell)	1	1	1	**14H (# eCell)**	1	0	0
02H (# aNCell)	1	1	1	**14H (# eNCell)**	1	0	0
03H (# bCell)	0	1	1	**15H**	0	1	0
03H (# bNCell)	0	1	1	**16H**	1	1	0
04H	1	0	1	**17H (# fCell)**	1	1	0
05H	0	0	1	**17H (# fNCell)**	1	1	0
06H (# cCell)	1	0	1	**18H (# fCell)**	0	1	0
06H (# cCell)	1	0	1	**18H (# fNCell)**	0	1	0
06H (# cCell)	1	0	1	**20H (# dCell)**	1	0	0
06H (# cNCell)	1	0	1	**04N (# dNCell)**	0	0	0
06H (# cNCell)	1	0	1	**18O (# dNCell)**	0	0	0
07H (# cCell)	0	0	1	**19O (# dNCell)**	0	0	0
07H (# cNCell)	0	0	1	**21H (# eCell)**	1	0	0
08H (# aCell)	0	1	0	**05N (# eNCell)**	0	0	0
01N (# aNCell)	0	1	0	**20O (# eNCell)**	0	0	0
12O (# aNCell)	0	1	0	**21O (# eNCell)**	0	0	0
13O (# aNCell)	0	1	1	**22H (# fCell)**	0	1	0
09H (# bCell)	0	1	1	**06N (# fNCell)**	0	1	0
02N (# bNCell)	1	1	1	**22O (# fNCell)**	0	1	0
14O (# bNCell)	1	1	1	**23O (# fNCell)**	0	1	0
15O (# bNCell)	1	1	1				
11H (# cCell)	1	0	1				
03N (# cNCell)	1	0	1				
16O (# cNCell)	1	0	1				
17O (# cNCell)	1	0	1				

```
# unit to unit
    $bond:unit01unit02  $atom:unit_xyz01/07C    $atom:unit_xyz02/04O
    $bond:unit02unit03  $atom:unit_xyz02/07C    $atom:unit_xyz03/04O
  }

} # cChain_xyza
```

LISTING A.3: Chain input file

```
import "chain_xyza.lt"
import "chain_0yza.lt"
import "chain_1yza.lt"

import "new_oplsaa.lt"

cPlane_xyza inherits OPLSAA {

  create_var {$mol}
   chain_xyza01 = new cChain_0yza
   chain_xyza02 = new cChain_xyza.move( 7.78, 0.00, 0.00)
   chain_xyza03 = new cChain_1yza.move(15.57, 0.00, 0.00)

  # list the bonds

  write('Data Bond List') {
# unit to unit
    $bond:unit01unit02  $atom:chain_xyza03/unit_1yz01/07C       $atom:
    chain_xyza01/unit_0yz02/04O
    $bond:unit02unit03  $atom:chain_xyza03/unit_1yz02/07C       $atom:
    chain_xyza01/unit_0yz03/04O
# edges
    $bond:01C02C.end.2  $atom:chain_xyza01/unit_0yz02/01C       $atom:
    chain_xyza03/unit_1yz02/02C
    $bond:01C01O.end.2  $atom:chain_xyza01/unit_0yz02/01C       $atom:
    chain_xyza03/unit_1yz02/01O
    $bond:01C05O.end.2  $atom:chain_xyza01/unit_0yz02/01C       $atom:
    chain_xyza03/unit_1yz02/05O
    $bond:02C03C.end.2  $atom:chain_xyza03/unit_1yz02/02C       $atom:
    chain_xyza01/unit_0yz02/03C
    $bond:02C02O.end.2  $atom:chain_xyza03/unit_1yz02/02C       $atom:
    chain_xyza01/unit_0yz02/02O
    $bond:03C04C.end.2  $atom:chain_xyza01/unit_0yz02/03C       $atom:
    chain_xyza03/unit_1yz02/04C
    $bond:03C03O.end.2  $atom:chain_xyza01/unit_0yz02/03C       $atom:
    chain_xyza03/unit_1yz02/03O
    $bond:04C05C.end.2  $atom:chain_xyza03/unit_1yz02/04C       $atom:
    chain_xyza01/unit_0yz02/05C
    $bond:04C04O.end.2  $atom:chain_xyza03/unit_1yz02/04C       $atom:
    chain_xyza01/unit_0yz02/04O
    $bond:05C06C.end.2  $atom:chain_xyza01/unit_0yz02/05C       $atom:
    chain_xyza03/unit_1yz02/06C
```

```
31     $bond:05C05O.end.2  $atom:chain_xyza01/unit_0yz02/05C       $atom:
       chain_xyza03/unit_1yz02/05O
32     $bond:06C07H.end.2  $atom:chain_xyza03/unit_1yz02/06C       $atom:
       chain_xyza01/unit_0yz02/07H
33     $bond:03O09H.end.2  $atom:chain_xyza03/unit_1yz02/03O       $atom:
       chain_xyza01/unit_0yz02/09H       # bCell
34     $bond:07C08C.end.2  $atom:chain_xyza03/unit_1yz02/07C       $atom:
       chain_xyza01/unit_0yz02/08C
35     $bond:07C10O.end.2  $atom:chain_xyza03/unit_1yz02/07C       $atom:
       chain_xyza01/unit_0yz02/10O
36     $bond:08C09C.end.2  $atom:chain_xyza01/unit_0yz02/08C       $atom:
       chain_xyza03/unit_1yz02/09C
37     $bond:08C08O.end.2  $atom:chain_xyza01/unit_0yz02/08C       $atom:
       chain_xyza03/unit_1yz02/08O
38     $bond:09C10C.end.2  $atom:chain_xyza03/unit_1yz02/09C       $atom:
       chain_xyza01/unit_0yz02/10C
39     $bond:09C09O.end.2  $atom:chain_xyza03/unit_1yz02/09C       $atom:
       chain_xyza01/unit_0yz02/09O
40     $bond:10C11C.end.2  $atom:chain_xyza01/unit_0yz02/10C       $atom:
       chain_xyza03/unit_1yz02/11C
41     $bond:10C01O.end.2  $atom:chain_xyza01/unit_0yz02/10C       $atom:
       chain_xyza03/unit_1yz02/01O
42     $bond:11C12C.end.2  $atom:chain_xyza03/unit_1yz02/11C       $atom:
       chain_xyza01/unit_0yz02/12C
43     $bond:11C10O.end.2  $atom:chain_xyza03/unit_1yz02/11C       $atom:
       chain_xyza01/unit_0yz02/10O
44     $bond:12C17H.end.2  $atom:chain_xyza01/unit_0yz02/12C       $atom:
       chain_xyza03/unit_1yz02/17H
45     $bond:08O04N.end.2  $atom:chain_xyza03/unit_1yz02/08O       $atom:
       chain_xyza01/unit_0yz02/04N       # dNCell
46     $bond:09O21H.end.2  $atom:chain_xyza01/unit_0yz02/09O       $atom:
       chain_xyza03/unit_1yz02/21H       # eCell
47 # ends
48     $bond:01C02C.end.end        $atom:chain_xyza01/unit_0yz01/01C
       $atom:chain_xyza03/unit_1yz03/02C
49     $bond:01C01O.end.end        $atom:chain_xyza01/unit_0yz01/01C
       $atom:chain_xyza03/unit_1yz01/01O
50     $bond:01C05O.end.end        $atom:chain_xyza01/unit_0yz01/01C
       $atom:chain_xyza03/unit_1yz03/05O
51     $bond:02C03C.end.end        $atom:chain_xyza03/unit_1yz03/02C
       $atom:chain_xyza01/unit_0yz03/03C
52     $bond:02C02O.end.end        $atom:chain_xyza03/unit_1yz03/02C
       $atom:chain_xyza01/unit_0yz03/02O
53     $bond:03C04C.end.end        $atom:chain_xyza01/unit_0yz03/03C
       $atom:chain_xyza03/unit_1yz03/04C
54     $bond:03C03O.end.end        $atom:chain_xyza01/unit_0yz03/03C
       $atom:chain_xyza03/unit_1yz03/03O
55     $bond:04C05C.end.end        $atom:chain_xyza03/unit_1yz03/04C
       $atom:chain_xyza01/unit_0yz03/05C
56     $bond:04C04O.end.end        $atom:chain_xyza03/unit_1yz03/04C
       $atom:chain_xyza01/unit_0yz03/04O
```

```
57      $bond:05C06C.end.end        $atom:chain_xyza01/unit_0yz03/05C
        $atom:chain_xyza03/unit_1yz03/06C
58      $bond:05C05O.end.end        $atom:chain_xyza01/unit_0yz03/05C
        $atom:chain_xyza03/unit_1yz03/05O
59      $bond:06C07H.end.end        $atom:chain_xyza03/unit_1yz03/06C
        $atom:chain_xyza01/unit_0yz03/07H
60      $bond:03O09H.end.end        $atom:chain_xyza03/unit_1yz03/03O
        $atom:chain_xyza01/unit_0yz03/09H       # bCell
61      $bond:07C08C.end.end        $atom:chain_xyza03/unit_1yz01/07C
        $atom:chain_xyza01/unit_0yz01/08C
62      $bond:07C10O.end.end        $atom:chain_xyza03/unit_1yz01/07C
        $atom:chain_xyza01/unit_0yz01/10O
63      $bond:08C09C.end.end        $atom:chain_xyza01/unit_0yz01/08C
        $atom:chain_xyza03/unit_1yz01/09C
64          $bond:08C08O.end.end        $atom:chain_xyza01/unit_0yz01/08C
           $atom:chain_xyza03/unit_1yz01/08O
65      $bond:09C10C.end.end        $atom:chain_xyza03/unit_1yz01/09C
        $atom:chain_xyza01/unit_0yz01/10C
66      $bond:09C09O.end.end        $atom:chain_xyza03/unit_1yz01/09C
        $atom:chain_xyza01/unit_0yz01/09O
67      $bond:10C11C.end.end        $atom:chain_xyza01/unit_0yz01/10C
        $atom:chain_xyza03/unit_1yz01/11C
68      $bond:10C01O.end.end        $atom:chain_xyza01/unit_0yz01/10C
        $atom:chain_xyza03/unit_1yz01/01O
69      $bond:11C12C.end.end        $atom:chain_xyza03/unit_1yz01/11C
        $atom:chain_xyza01/unit_0yz01/12C
70      $bond:11C10O.end.end        $atom:chain_xyza03/unit_1yz01/11C
        $atom:chain_xyza01/unit_0yz01/10O
71      $bond:12C17H.end.end        $atom:chain_xyza01/unit_0yz01/12C
        $atom:chain_xyza03/unit_1yz01/17H
72      $bond:08O04N.end.end        $atom:chain_xyza03/unit_1yz01/08O
        $atom:chain_xyza01/unit_0yz01/04N       # dNCell
73      $bond:09O21H.end.end        $atom:chain_xyza01/unit_0yz01/09O
        $atom:chain_xyza03/unit_1yz01/21H       # eCell
74    }
75  } # cPlane_xyza
```

LISTING A.4: Plane input file

```
1  import "plane_xyza.lt"
2  import "plane_x0za.lt"
3  import "plane_x1za.lt"
4
5  import "new_oplsaa.lt"
6
7  cBlock_xyza inherits OPLSAA {
8
9    create_var {$mol}
10
11
12   plane_xyza01 = new cPlane_x0za
```

```
  plane_xyza02 = new cPlane_xyza.move(-0.93,  8.15, 0.00)
  plane_xyza03 = new cPlane_x1za.move(-1.86, 16.30, 0.00)


  # list bonds


  write('Data Bond List') {
    $bond:07C10O.2.end          $atom:plane_xyza01/chain_x0za02/
    unit_x0z01/07C  $atom:plane_xyza03/chain_x1za02/unit_x1z01/10O
    $bond:09C10C.2.end          $atom:plane_xyza01/chain_x0za02/
    unit_x0z01/09C  $atom:plane_xyza03/chain_x1za02/unit_x1z01/10C
    $bond:01C05O.2.end          $atom:plane_xyza03/chain_x1za02/
    unit_x1z01/01C  $atom:plane_xyza01/chain_x0za02/unit_x0z03/05O
    $bond:03C04C.2.end          $atom:plane_xyza03/chain_x1za02/
    unit_x1z03/03C  $atom:plane_xyza01/chain_x0za02/unit_x0z03/04C
    $bond:07C10O.2.2            $atom:plane_xyza01/chain_x0za02/
    unit_x0z02/07C  $atom:plane_xyza03/chain_x1za02/unit_x1z02/10O
    $bond:09C10C.2.2            $atom:plane_xyza01/chain_x0za02/
    unit_x0z02/09C  $atom:plane_xyza03/chain_x1za02/unit_x1z02/10C
    $bond:01C05O.2.2            $atom:plane_xyza03/chain_x1za02/
    unit_x1z02/01C  $atom:plane_xyza01/chain_x0za02/unit_x0z02/05O
    $bond:03C04C.2.2            $atom:plane_xyza03/chain_x1za02/
    unit_x1z02/03C  $atom:plane_xyza01/chain_x0za02/unit_x0z02/04C
    $bond:01C05O.end.end        $atom:plane_xyza03/chain_x1za01/
    unit_01z01/01C  $atom:plane_xyza01/chain_x0za03/unit_10z03/05O
    $bond:03C04C.end.end        $atom:plane_xyza03/chain_x1za01/
    unit_01z03/03C  $atom:plane_xyza01/chain_x0za03/unit_10z03/04C
    $bond:07C10O.end.end        $atom:plane_xyza01/chain_x0za03/
    unit_10z01/07C  $atom:plane_xyza03/chain_x1za01/unit_01z01/10O
    $bond:09C10C.end.end        $atom:plane_xyza01/chain_x0za03/
    unit_10z01/09C  $atom:plane_xyza03/chain_x1za01/unit_01z01/10C
    $bond:01C05O.end.2          $atom:plane_xyza03/chain_x1za01/
    unit_01z02/01C  $atom:plane_xyza01/chain_x0za03/unit_10z02/05O
    $bond:03C04C.end.2          $atom:plane_xyza03/chain_x1za01/
    unit_01z02/03C  $atom:plane_xyza01/chain_x0za03/unit_10z02/04C
    $bond:07C10O.end.2          $atom:plane_xyza01/chain_x0za03/
    unit_10z02/07C  $atom:plane_xyza03/chain_x1za01/unit_01z02/10O
    $bond:09C10C.end.2          $atom:plane_xyza01/chain_x0za03/
    unit_10z02/09C  $atom:plane_xyza03/chain_x1za01/unit_01z02/10C
  }
} # cBlock_xyza
```

LISTING A.5: Block input file

```
import "block_full.lt"
import "new_oplsaa.lt"
# Periodic boundary conditions: DEFINE HOW BIG YOUR SYSTEM VOLUME IS
write_once("Data Boundary") {
   -2    16 xlo xhi
   0    17   ylo yhi
```

```
   0   21   zlo zhi
}


c_block_1yz = new cBlock_full
```

LISTING A.6: System input file

A.4 Sample LAMMPS Input file

```
## run.in.nvt
## This file is a modified version of the "system.in" input script
## created by moltemplate.
##
## --  Usage  --
##
## lmp_g++ -i run.in.nvt
## (assuming lmp_g++ is the name of your lammps binary)
##
## --  Prerequisite Input Files:  --
## systen.data, system.in.init, system.in.settings
##
##     You can generate these files by running moltemplate this way:
## moltemplate.sh -atomstyle "full" system.lt
## ---------------------------------

## -- Init section --

include "system.in.init"       # <-- contains text written to "In Init"

## -- Atom definition section --

# For step 1
boundary p p p
read_data "system.data"
change_box all triclinic

change_box all xy final  -1.86
change_box all xz final   0.00
change_box all yz final   0.00
change_box all x final 0 15.57
change_box all y final 0 16.30
change_box all z final 0 20.76

## For other steps
#read_data       "system_after_minimisation.data"

## -- Settings Section --

```

```
include "system.in.settings"

# Only in step 1
include "system.in.charges"
## -- Run section --


timestep        0.5

variable d internal 0.0
variable dsq equal v_d*v_d

# Minimisation
min_style cg

neigh_modify delay 0 every 1 check yes
kspace_style pppm 1.0e-4
#thermo_style custom step temp press etotal pe
thermo_style custom step etotal pe cella cellb cellc cellalpha cellbeta
    cellgamma ## check the unit cell dimensions

dump 1 all custom 10 traj_min.lammpstrj id mol type x y z ix iy iz
dump 2 all xyz 10 traj_min.xyz
dump 3 all dcd 10 traj_min.dcd

## How often to output thermo data
thermo 1

## Minimisation
minimize 1.0e-4 1.0e-6 100000 400000

## NVT calculation
#fix fxnvt all nvt temp 800.0 800.0 $(200.0*dt)

## NPT calculation
#fix fxnpt all npt temp 300.0 300.0 30000.0 couple xy aniso 9869 9869 $
    (200.0*dt)

## Run NVT and NPT
#run            200000 # 100 ps

write_data  system_after_min.data
```

LISTING A.7: Example LAMMPS input file for a minimisation step

Appendix B

Appendix: Chapter 5

B.1 Structure Input for Testing Scripts

```
105

 H 13.454871975147 10.7295941650162 26.5236465751755
 C 14.4611086281028 10.4814598498988 31.1400655596015
 C 15.1352484820914 9.63278671615099 30.0309677054849
 C 14.8098222264233 10.2031237510543 28.6421391059915
 C 13.3094609046922 10.5173908765069 28.4544600060237
 C 12.8433641067544 11.3691348255944 29.6343738358511
 C 11.3837047145883 11.7944633351453 29.5990043720964
 O 14.5800895724651 9.81222345775777 32.3655894905186
 O 16.5720656225549 9.72184380065602 30.1316417643261
 O 15.1663682824393 9.30729209078766 27.5924141937981
 O 13.097656615798 11.2555894836804 27.260550338638
 O 13.0796906262344 10.6291525737226 30.8356888620075
 O 10.5710008140431 10.6577117913951 29.9549757416642
 H 14.9418193425782 11.4812850976342 31.1900460823368
 H 14.8084649398018 8.58551755662818 30.108942337552
 H 15.3514320298906 11.1621672693018 28.5262785079557
 H 12.7427873636019 9.56793768972317 28.4421208106337
 H 13.4296347732351 12.3120855357603 29.6311888769875
 H 11.2155099194301 12.5995902647511 30.3293833276629
 H 11.1282809305088 12.1601905628075 28.5943963678346
 N 17.2423394666128 8.65689165554066 30.9157341115729
 O 16.5822218032075 7.69430878845321 31.2317620741861
 O 18.4008277579597 8.92138109188183 31.1004373835098
 H 16.1361515579647 9.26446075213646 27.5360781469557
 N 9.15431061119086 10.9924164285626 30.0549672148244
 O 8.81725002417432 12.0890950756952 29.6599942678973
 O 8.51029586315187 10.0933194803321 30.5340288014477
 C 13.4752133353632 10.5925307062691 36.333272619524
 C 12.9908847421502 11.6600770897304 35.3378287474024
 C 13.2509250152771 11.2915345334845 33.857522818524
```

```
C 14.5902308832381 10.5778655653777 33.5904257556535
C 14.8742708244576 9.57096717632297 34.7122658265219
C 16.2071365528759 8.83952208743497 34.6607798475467
O 11.5722325618805 11.8303875353674 35.6231859694465
O 13.2694708772255 12.4733610214136 33.0617798941884
O 14.8119441211932 10.2575997929273 35.9587534300521
O 17.3244855496951 9.75022701398901 34.6403003191614
H 12.8393233625096 9.68317078942537 36.3143923704977
H 13.5297540956088 12.5888462740556 35.5815009873817
H 12.4437738240963 10.6075150997702 33.5312664425694
H 15.3918483625163 11.3364991442155 33.5510594693623
H 14.0974539229768 8.77817589661977 34.6670677957243
H 16.2942035606537 8.15515117421888 35.5136500974043
H 16.2860799496582 8.26181832369716 33.7288737007307
N 11.0895470638992 13.2217688857137 35.5398306635255
O 10.6303909542255 13.6269021408756 36.573404240638
O 11.1584527628058 13.7601026060021 34.4532972959949
H 12.4134757301174 12.9272728261999 33.1829608927942
N 18.0854003056425 9.85390674645556 35.8858123845082
O 17.749703355374 9.15372364783142 36.8225437278626
O 18.9842868000972 10.6435834478639 35.7883412765557
C 14.3306462316875 10.3388197629774 41.5443756588655
C 15.0183751447143 9.3671725665713 40.5841891178481
C 14.8509196101161 9.77480856533984 39.1121334320712
C 13.4467715421133 10.338299899288 38.7662165882142
C 12.9477006010542 11.2523237723748 39.9028933592249
C 11.5411076285136 11.813135564611 39.7484191765589
O 14.3281285123659 9.79944943745573 42.8269855865647
O 16.4435131854375 9.40237020648645 40.8159316926789
O 15.149054503785 8.63516755527921 38.3189161647153
O 13.4976276414574 11.1747276078471 37.5937974422703
O 12.9896517063371 10.5191471627435 41.1208466164114
O 10.6247589211976 10.7738904778508 39.3683362708367
H 14.8588615016583 11.3179292047041 41.5381459201808
H 14.628567020258 8.35304715146506 40.7458387317877
H 15.5665818300134 10.5925065757884 38.9101550643338
H 12.7367215640163 9.5084074763337 38.6171311758136
H 13.6222103499113 12.1329921772439 39.9601736698778
H 11.2343189941219 12.2581540701523 40.7036441520713
H 11.5247808711882 12.5824956712965 38.9614963662603
N 17.0096139980353 8.31753258731987 41.6378396253258
O 16.276169360148 7.40892621882078 41.9566431753142
O 18.1743121839216 8.51333524536546 41.8676123909164
H 15.9040453325563 8.85163310161273 37.7447701122666
N 9.33210060624313 10.80294733377 40.0766518706334
O 8.57034170681311 9.99196130119607 39.6187879587847
O 9.20280658018603 11.5933313863526 40.9822859565889
C 13.1303743196441 10.8481408136436 46.6128065431468
C 12.5772556527445 11.807253336599 45.5449795370119
C 12.912986893696 11.384834507928 44.1031052616395
C 14.3099826554692 10.7207675420994 43.9276894201323
```

```
C 14.6350103877817 9.84247493122465 45.1497332700567
C 16.0358589300553 9.2327574809986 45.1335487026472
O 11.1455775638216 11.8177264414816 45.6261633642587
O 12.5645039108029 12.4681957756011 43.2022716067291
O 14.5010936813037 10.5859068893173 46.3507884172861
O 17.0536832938476 10.2333644822239 45.0812096900122
H 12.5677546993112 9.88863812833247 46.5810447450097
H 12.9785308885515 12.8094992525561 45.769325484496
H 12.172845094896 10.6299589868704 43.7893415424652
H 15.093987635559 11.481479623176 43.8077148926437
H 13.925531286057 8.98560738257092 45.1452898748741
H 16.1625687192157 8.61884995271951 46.0351052665419
H 16.1591016030646 8.59551180905502 44.2476372099982
N 10.5573118463479 12.9001963630651 46.4827856921941
O 11.2585447267178 13.8450357066874 46.7391311442172
O 9.41809513057285 12.6435246203753 46.7594942006989
N 13.3009917938974 13.7542482127112 43.379428176663
O 12.5704950434844 14.7013278280448 43.2693872049755
O 14.4997136929799 13.6906442344191 43.5391375623684
N 17.843491337039 10.374649293397 46.3317103618193
O 17.5688078309571 9.64808030789491 47.2569988301392
O 18.6853581935336 11.2221853311596 46.2087752466035
H 13.2406521510501 10.8546370988881 48.5455670815329
O 13.0045681534847 11.4827793353173 47.841143050926
```

LISTING B.1: chain unit xyz file used in Chapter 5 which can be used to test the scripts.

B.2 Script to Create Corrected Reactants

```
#!/usr/bin/env python3
import argparse
import re
import csv
import numpy as np

######################
## Define functions ##
######################

def read_original(file_name):
    """
    Read in the xyz file and make a list of each of
    the columns: atom name, x, y and z coordinates
    The coordinates must be in xyz form, where the
    first line contains only the number of atoms,
    and the atoms being to be defined on line 2
    """
    with open(file_name,'r') as f:
        atoms = []
```

```
        x = []
        y = []
        z = []
        row_no=0
        reader = csv.reader(f,delimiter=' ')
        for row in reader:
            if row_no > 1:
                # space at the beginning of line is row[0], so
                #  save data from [1] onwards
                atoms.append(row[1])
                x.append(row[2])
                y.append(row[3])
                z.append(row[4])
    #             print(row)
            row_no += 1
    return atoms, x, y, z

def write_atom(index, x, y, z):
    """
    For a specific atom with a given index, save the
    x,y,z coordinates in the form of a list of integers
    """
    atom = []
    atom.append(float(x[index]))
    atom.append(float(y[index]))
    atom.append(float(z[index]))
    return atom

def find_vector(atom1, atom2):
    """
    Find the vector between two atoms
    """
    vec = []
    i = 0
    while i < 3:
        vec.append(atom2[i] - atom1[i])
        i += 1
    return vec

def find_new_O(C_original, distance, vec):
    """
    Find position of the O of the OH, positioned "distance"
    away from the carbon it is attacking.
    """
    new_vec = scale_vector(vec, distance)
    new_O = []
    i = 0
    while i < 3:
        new_O.append(C_original[i] + new_vec[i])
        i += 1
    return new_O
```

```
72
73 def find_new_H_ring(CH_vec, O_new):
74     """
75     Pick the location for the H atom.
76     Step 1: translate CO_vector to O_new
77     """
78     OH_vec = scale_vector(CH_vec, 0.945)
79     H_new = []
80     i = 0
81     while i < 3:
82         H_new.append(O_new[i] + OH_vec[i])
83         i += 1
84     return H_new
85
86 def find_new_H_chain(CH_vec, O_new):
87     """
88     Pick the location for the H atom.
89     Step 1: translate CO_vector to O_new
90     """
91     OH_vec = scale_vector(CH_vec, 0.945)
92     H_new = []
93     i = 0
94     while i < 3:
95         H_new.append(O_new[i] + OH_vec[i])
96         i += 1
97     return H_new
98
99 def scale_vector(vec, scale):
100     """
101     Scales any vector of any number of dimensions by a certain amount
102     """
103     i = 0
104     sum_vec = 0
105     while i < 3:
106         sum_vec += vec[i]**2
107         i += 1
108     len_vec = np.sqrt(sum_vec)
109 #     len_vec = np.sqrt(vec[0]**2+vec[1]**2+vec[2]**2)
110 #     print(len_vec)
111     new_vec = []
112     i=0
113     while i < 3:
114         new_vec.append(vec[i]*scale/len_vec)
115         i += 1
116     return new_vec
117
118 def find_oh_ring(C_index, O_index, H_index, distance, x, y, z, yn_print):
119     """
120     finds the coordinates for OH- atoms, on "in-ring" positions
121     This could be subject to change if it is decided the
122     OH bond needs rotating
```

```
    """
    C_original = write_atom(C_index, x, y, z)
    O_original = write_atom(O_index, x, y, z)
    H_original = write_atom(H_index, x, y, z)
    CO_vec = find_vector(C_original, O_original)
    CH_vec = find_vector(C_original, H_original)
    O_new = find_new_O(C_original, -2, CO_vec)
    H_new = find_new_H_ring(CH_vec, O_new)
    if yn_print != 0 :
        print("Calcuating for a position which is on the ring.")
        print("The new coordinates found are:")
        print("O coordinate:", O_new)
        print("H coordinate:", H_new)
    return O_new, H_new

def find_oh_chain(C_index, O_index, H_index, distance, x, y, z, yn_print)
    :
    """
    finds the coordinates for OH- atoms, on "off-ring" positions
    """
    C_original = write_atom(C_index, x, y, z)
    O_original = write_atom(O_index, x, y, z)
    H_original = write_atom(H_index, x, y, z)
    CO_vec = find_vector(C_original, O_original)
    CH_vec = find_vector(C_original, H_original)
    O_new = find_new_O(C_original, -2, CO_vec)
    H_new = find_new_H_chain(CH_vec, O_new)
    if yn_print != 0 :
        print("Calcuating for a position which is off ring.")
        print("The new coordinates found are:")
        print("O coordinate:", O_new)
        print("H coordinate:", H_new)
    return O_new, H_new

def write_atom_new(atom, atom_type, position):
    """
    Writes on of the newly found atoms into the output file
    """
    with open("reactant_{}.xyz".format(position), "a") as f:
        f.write(atom_type)
        f.write(" ")
        f.write(str(atom[0]))
        f.write(" ")
        f.write(str(atom[1]))
        f.write(" ")
        f.write(str(atom[2]))
        f.write("\n")
    return None

def write_atom_old(atom, x, y, z, position):
    """
```

```
    Writes the atoms from the original structure into the output file
    """
    with open("reactant_{}.xyz".format(position), "a") as f:
        f.write(atom)
        f.write(" ")
        f.write(x)
        f.write(" ")
        f.write(y)
        f.write(" ")
        f.write(z)
        f.write("\n")
    return None

def write_output(Oh, oH, position, atoms, x, y, z):
    """
    After the positions of the OH atoms are found,
    this writes the new output file.
    """
    no_of_atoms = len(atoms) + 2
    with open("reactant_{}.xyz".format(position), "w") as f:
        f.write(str(no_of_atoms))
        f.write("\n\n")

    i=0
    while i < len(atoms):
        write_atom_old(atoms[i], x[i], y[i], z[i], position)
        i += 1
    write_atom_new(Oh, "O", position)
    write_atom_new(oH, "H", position)
    return None

def add_oh(C_index, O_index, H_index, position, distance, ring_or_chain,
    input_file, yn_print=0):
    """
    This finds the positions for OH based on the
    indexes of the atoms around a certain position,
    and writes them into a new output .xyz file
    """
    atoms, x, y, z = read_original(input_file)
    #print(atoms)
    if ring_or_chain == 'ring':
        Oh, oH = find_oh_ring(C_index, O_index, H_index, distance, x, y,
    z, yn_print)
    if ring_or_chain == 'chain':
        Oh, oH = find_oh_chain(C_index, O_index, H_index, distance, x, y,
     z, yn_print)
        #print(oH,)
    write_output(Oh, oH, position, atoms, x, y, z)
    return None

##########################
```

```
## Begin running script ##
##########################

my_parser = argparse.ArgumentParser(
    prog='create_reactants.py',
    usage='%(prog)s [-h]',
    description='create reactant geometries')

args = my_parser.parse_args()

print("Name of original geometry xyz file")
filename = input()

another = 0
while another==0:
    print("What position is being attacked?")
    print("label for the naming of output file")
    position_name = input()

    print("Is this a ring or chain position?")
    print("ring/chain")
    roc = input()

    print("Index of carbon atom")
    c_index = int(input())

    print("Index of oxygen atom")
    o_index = int(input())

    print("Index of hydrogen atom")
    h_index = int(input())

    print("What distance should the OH- be placed at?")
    print("units = angstrom")
    distance = input()

    add_oh(c_index, o_index, h_index, position_name, distance * -1, roc,
    filename)

    print("Would you like to generate another structure?")
    print("y/n")
    yon = input()
    if yon == 'n':
        another = 1
```

LISTING B.2: Python script used for creating the reactant structures with the OH- nucleophile added.

B.3 Script to Create Improved Products

```
#!/usr/bin/env python3
import argparse
import re
import csv
import numpy as np

######################
## Define functions ##
######################

def read_original(file_name):
    """
    Read in the xyz file and make a list of each of
    the columns: atom name, x, y and z coordinates
    The coordinates must be in xyz form, where the
    first line contains only the number of atoms,
    and the atoms being to be defined on line 2
    """
    with open(file_name,'r') as f:
        atoms = []
        x = []
        y = []
        z = []
        row_no=0
        reader = csv.reader(f,delimiter=' ')
        for row in reader:
            if row_no > 1:
                # space at the beginning of line is row[0], so
                #  save data from [1] onwards
                atoms.append(row[1])
                x.append(row[2])
                y.append(row[3])
                z.append(row[4])
    #             print(row)
            row_no += 1
    return atoms, x, y, z

def change_atom_type(index, type_new, atoms):
    atoms[index] = type_new
    return None

def invert_ring(c_index, o_index, h_index, atoms, x, y, z):
    change_atom_type(o_index, 'H', atoms)
    change_atom_type(h_index, 'O', atoms)

### Write function to add H in plane of h_index, c_index, and o_index, at
    a distance of 0.97, and so that the bond angle of h_new-h_index-
   c_index is 120 degrees

```

```
48      return h_new, atoms, x, y, z
49
50  def invert_chain(c_index, o_index, h_index, atoms, x, y, z):
51      c2_index = c_index - 1
52      h2_index = h_index + 1
53  ### Define an axis using c_index - c2_index vector
54  ### Rotate o_index, h_index and h2_index 180 degrees around the axis
55  ### Define a new atoms h_new, which is in the plane of c2_index, c_index,
        o_index, at a distance of 0.97, so that the bond angle of h_new-
        o_index-c_index is 120 degrees
56      return h_new, atoms, x, y, z
57
58  def write_atom_old(atom, x, y, z, position):
59      """
60      Writes the atoms from the original structure into the output file
61      """
62      with open("product_{}.xyz".format(position), "a") as f:
63          f.write(atom)
64          f.write(" ")
65          f.write(x)
66          f.write(" ")
67          f.write(y)
68          f.write(" ")
69          f.write(z)
70          f.write("\n")
71      return None
72
73  def write_atom_new(atom, atom_type, position):
74      """
75      Writes on of the newly found atoms into the output file
76      """
77      with open("product_{}.xyz".format(position), "a") as f:
78          f.write(atom_type)
79          f.write(" ")
80          f.write(str(atom[0]))
81          f.write(" ")
82          f.write(str(atom[1]))
83          f.write(" ")
84          f.write(str(atom[2]))
85          f.write("\n")
86      return None
87
88  def write_output(n_index, o2, o3, h_new, position, atoms, x, y, z):
89      """
90      This writes the new output file with the corrected positions
91      and removing the NO2 atoms
92      """
93      no_of_atoms = len(atoms) - 3
94  ## When H_new is sucessfully added use len(atoms) - 2
95  #     no_of atoms = len(atoms) - 2
96
```

```
97      with open("product_{}.xyz".format(position), "w") as f:
98          f.write(str(no_of_atoms))
99          f.write("\n\n")
100
101     i=0
102
103     while i < len(atoms):
104 #         write_atom(atoms[i], x[i], y[i], z[i], position)
105         if i != n_index:
106             if i != o2:
107                 if i != o3:
108                     write_atom(atoms[i], x[i], y[i], z[i], position)
109 # Currently this is not removing the correct oxygens?
110         i += 1
111     write_atom_new(H_new, "H", position)
112     return None
113
114
115 def create_product(c_index, o_index, h_index, n_index, o2, o3,
        position_name, roc, filename):
116     atoms, x, y, z = read_original(filename)
117     if roc == 'ring':
118         h_new, atoms, x, y, z = invert_ring(c_index, o_index, h_index,
        atoms, x, y, z)
119     if roc == 'chain':
120         h_new, atoms, x, y, z = invert_chain(c_index, o_index, h_index,
        atoms, x, y, z)
121
122
123     write_output(n_index, o2, o3, h_new, position_name, atoms, x, y, z)
124     return None
125
126 ##########################
127 ## Begin running script ##
128 ##########################
129
130 my_parser = argparse.ArgumentParser(
131     prog='create_reactants.py',
132     usage='%(prog)s [-h]',
133     description='create product geometries')
134
135 args = my_parser.parse_args()
136
137 print("Name of original geometry xyz file")
138 filename = input()
139
140 another = 0
141 while another==0:
142     print("What position has been attacked?")
143     print("label for the naming of output file")
144     position_name = input()
```

```
145
146     print("Is this a ring or chain position?")
147     print("ring/chain")
148     roc = input()
149
150     print("Index of carbon atom")
151     c_index = int(input())
152
153     print("Index of oxygen attached to the carbon atom")
154     o_index = int(input())
155
156     print("Index of hydrogen atom")
157     print("In case of chain position, give the lower index")
158     h_index = int(input())
159
160     print("Index of nitrogen atom")
161     n_index = int(input())
162
163     print("Index of the two oxygens of NO2 group")
164     print("two numbers seperated by a space")
165     os = input().split()
166     o2 = os[0]
167     o3 = os[1]
168
169     create_product(c_index, o_index, h_index, n_index, o2, o3,
    position_name, roc, filename)
170
171     print("Would you like to generate another structure?")
172     print("y/n")
173     yon = input()
174     if yon == 'n':
175         another = 1
```

LISTING B.3: Incompleted Python script used for creating improved product structures.

B.4 Indexes for Different Nitrate Structures

TABLE B.1: The index numbers required to run the script for the chain unit shown above.

Nitrate Position	C index	O index	H index	N index	O2	O3
4a	2	8	14	20	21	22
4c	6	12	18	24	25	26
4d	28	33	38	44	45	46
4f	32	36	42	48	49	50
5a	52	58	64	70	71	72
5c	56	62	68	74	75	76
5d	78	83	88	94	95	96
5e	79	84	89	97	98	99
5f	82	86	92	100	101	102

References

[1] F. Jensen, *Introduction to Computational Chemistry*, Wiley & Sons, Ltd. Publications, Chichester, 2nd edn., 2007.

[2] C. o. M. C. f. C. Chemistry, in *Mathematical Challenges from Theoretical/Comupational Chemistry*, 1995, ch. 2, pp. 13–18.

[3] D. Fabregat-Traver, A. E. Ismail and P. Bientinesi, *Journal of Computational Science*, 2018, **27**, 77–90.

[4] D. Frenkel and B. Smit, in *Understanding Molecular Simulation*, Academic Press, 2nd edn., 2002, ch. 4, pp. 63–105.

[5] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama and A. C. Van Duin, *npj Computational Materials*, 2016, **2**, 15011.

[6] E. Paquet and H. L. Viktor, *Advances in Chemistry*, 2018, **2018**, 1–14.

[7] A. R. Leach, in *Molecular Modelling Principals and Applications*, Pearson Education Limited, 2nd edn., 2001, ch. 4, p. 165 to 252.

[8] S. Corporation, *improper_style cvff command*, 2023, https://docs.lammps.org/improper_cvff.html.

[9] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society*, 1996, **118**, 11225–11236.

[10] H. Sun, *Journal of Physical Chemistry B*, 1998, **102**, 7338–7364.

[11] V. Spiwok, B. Králová and I. Tvaroška, *Carbohydrate Research*, 2010, **345**, 530–537.

[12] L. A. Richards, A. Nash, A. Willetts, C. Entwistle and N. H. De Leeuw, *RSC Advances*, 2018, **8**, 5728–5739.

[13] W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society*, 1988, **110**, 1657–1666.

[14] A. C. Van Duin, S. Dasgupta, F. Lorant and W. A. Goddard, *Journal of Physical Chemistry A*, 2001, **105**, 9396–9409.

[15] A. Strachan, E. M. Kober, A. C. Van Duin, J. Oxgaard and W. A. Goddard, *Journal of Chemical Physics*, 2005, **122**, 054502.

[16] S. Kshirsagar, K. K. Mandadapu and P. Papadopoulos, *Computational Materials Science*, 2016, **120**, 127–134.

[17] L. Verlet, *Physical Review*, 1967, **159**, 98–103.

[18] M. P. Allen and D. J. Tildesley, *Computer Simularions of Liquids*, Oxford University Press, Oxford, 2007.

[19] J. L. Bregado, A. R. Secchi, F. W. Tavares, D. de Sousa Rodrigues and R. Gambetta, *Fluid Phase Equilibria*, 2019, **491**, 56–76.

[20] R. Pathria and P. D. Beale, in *Statistical Mechanics*, Elsevier Ltd, 3rd edn., 2011, ch. 2, pp. 25–38.

[21] P. Atkins and J. de Paula, *Atkin's Physical Chemistry*, Oxford University Press, 9th edn., 2010, pp. 335–365.

[22] Y. Zhang, S. E. Feller, B. R. Brooks and R. W. Pastor, *The Journal of Chemical Physics*, 1995, **103**, 10252–10266.

[23] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford Graduate Texts, Oxford, 2019, p. 180.

[24] A. S. Lemak and N. K. Balabaev, *Molecular Simulation*, 1994, **13**, 177–187.

[25] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.

[26] G. S. Larsen, P. Lin, K. E. Hart and C. M. Colina, *Macromolecules*, 2011, **44**, 6944–6951.

[27] C. J. Cramer, *Essentials of Computational Chemistry, Theories and Models*, John Wiley and Sons, LTD, 2nd edn., 2003, pp. 1720–1720.

[28] Y. Nishiyama, H. Chanzy and P. Langan, *Journal of the American Chemical Society*, 2002, **124**, 9074–9082.

[29] F. Giustino, *Materials Modelling using Density Functional Theory*, Oxford University Press, 5th edn., 2020.

[30] R. M. Martin, L. Reining and D. M. Ceperley, *Interacting Electrons*, Cambridge University Press, Cambridge, 1st edn., 2016.

[31] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, 2nd edn., 2001.

[32] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory*, Dover Publications, New York, 1st edn., 1996.

[33] R. M. Martin, *Electronic Structure; Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, 1st edn., 2020, p. 397.

[34] P. Hohenberg and W. Kohn, *Physical Review*, 1964, **136**, 864–871.

[35] L. J. Sham and W. Kohn, *Physical Review*, 1966, **145**, 561–567.

[36] A. D. Becke, *Journal of Chemical Physics*, 2014, **140**, 1–18.

[37] J. P. Perdew and K. Schmidt, *AIP Conf. Proc.*, 2001, **1**, 577.

[38] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, Cambridge University Press, Cambridge, 3rd edn., 2021.

[39] J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865–3868.

[40] N. Mardirossian and M. Head-Gordon, *Molecular Physics*, 2017, **115**, 2315–2372.

[41] C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785–789.

[42] D. A. Scherlis, J. L. Fattebert, F. Gygi, M. Cococcioni and N. Marzari, *Journal of Chemical Physics*, 2006, **124**, 1–12.

[43] J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, 2005, **105**, 2999–3093.

[44] D. M. Dolney, G. D. Hawkins, P. Winget, D. A. Liotard, C. J. Cramer and D. G. Truhlar, *Journal of Computational Chemistry*, 2000, **21**, 340–366.

[45] W. Hierse and E. B. Stechel, *Physical Review B*, 1994, **50**, 17811–17819.

[46] Weitao Yang, *Physical Review Letters*, 1991, **66**, 1438–1441.

[47] W. Yang and T. S. Lee, *The Journal of Chemical Physics*, 1995, **103**, 5674–5678.

[48] S. Goedecker, *Reviews of Modern Physics*, 1999, **71**, 1085–1123.

[49] C. Brouder, G. Panati, M. Calandra, C. Mourougane and N. Marzari, *Physical Review Letters*, 2007, **98**, 1–4.

[50] F. Mauri, G. Galli and R. Car, *Physical Review B*, 1993, **47**, 9973–9976.

[51] J. Kim, F. Mauri and G. Galli, *Physical Review B*, 1995, **52**, 1640–1648.

[52] J. Noedal and S. J. Wright, in *Numerical Optimization*, Springer, 2nd edn., 2006, ch. 6, pp. 135–163.

[53] N. D. Hine, M. Robinson, P. D. Haynes, C. K. Skylaris, M. C. Payne and A. A. Mostofi, *Physical Review B - Condensed Matter and Materials Physics*, 2011, **83**, 1–10.

[54] Y. P. Wang, S. L. Tao and Q. Chen, *Inverse Problems in Science and Engineering*, 2015, **23**, 1342–1365.

[55] N. Govind, M. Petersen, G. Fitzgerald, D. King-Smith and J. Andzelm, *Computational Materials Science*, 2003, **28**, 250–258.

[56] G. Mills and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, Proceedings of the International School of Physics, LERICI, Villa Marigola, 7 – 18 July 1997, 1999, ch. 16, p. 20.

[57] D. Sheppard, R. Terrell and G. Henkelman, *Journal of Chemical Physics*, 2008, **128**, 134106.

[58] G. Henkelman, G. Johannesson and H. Jónsson, in *in Theoretical Methods in Condensed Phase Chemistry*, Springer, 2002, ch. 10, pp. 269–300.

[59] G.Henkelman, B.P.Uberuaga and H.Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.

[60] H. Eyring, *Journal of Chemical Physics*, 1935, **3**, 107–155.

[61] G. A. Petersson, *Theoretical Chemistry Accounts*, 2000, **103**, 190–195.

[62] E. Aprà, T. Straatsma, M. Valiev, T. Windus, N. Govind, H. Van Dam, D. Wang, K. Kowalski, W. de Jong, J. Nieplocha and E. J. Bylaska, *Computer Physics Communications*, 2010, **181**, 1477–1489.

[63] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin Del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la Roza, B. Palmer,

A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao and R. J. Harrison, *The Journal of chemical physics*, 2020, **152**, 184102.

[64] A. Hjorth Larsen, J. JØrgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. SchiØtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics Condensed Matter*, 2017, **29**, 1–30.

[65] J. C. Prentice, J. Aarons, J. C. Womack, A. E. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M. Dubois, K. K. Duff, J. M. Escartín, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. Phipps, L. E. Ratcliff, Á. R. Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. Hine, A. A. Mostofi, M. C. Payne and C. K. Skylaris, *The Journal of chemical physics*, 2020, **152**, 174111.

[66] C. K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *Journal of Chemical Physics*, 2005, **122**, 1–10.

[67] A. A. Mostofi, C. K. Skylaris, P. D. Haynes and M. C. Payne, *Computer Physics Communications*, 2002, **147**, 788–802.

[68] C. K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard and M. C. Payne, *Computer Physics Communications*, 2001, **140**, 315–322.

[69] C. K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez and M. C. Payne, *Physical Review B - Condensed Matter and Materials Physics*, 2002, **66**, 1–12.

[70] J. L. Fattebert and F. Gygi, *Journal of Computational Chemistry*, 2002, **23**, 662–666.

[71] D. S. BIOVIA, *Chemical & Engineering News*, 2000, **78**, ibc.

[72] M. M. Nardai and M. A. Bohn, *New Trends in Research of Energetic Materials*, 2017, 863–876.

[73] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Computer Physics Communications*, 2022, **271**, 108171.

[74] *lammps*, http://lammps.sandia.gov.

[75] L. A. Richards, A. Nash, M. J. S. Phipps and N. H. De Leeuw, *New Journal of Chemistry*, 2018, **42**, 17420–17428.

[76] H. William, A. Dalke and S. Klaus, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

[77] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 0–17.

[78] V. . G. r. C1fcc5b), *Avogadro: an open-source molecular builder and visualisation tool.*, 2022.

[79] D. S. BIOVIA, *Discover*, 2001, 0–4.

[80] A. I. Jewett, D. Stelter, J. Lambert, S. M. Saladi, O. M. Roscioni, M. Ricci, L. Autin, M. Maritan, S. M. Bashusqeh, T. Keyes, R. T. Dame, J. E. Shea, G. J. Jensen and D. S. Goodsell, *Journal of Molecular Biology*, 2021, **433**, 166841.

[81] T. Lu and F. Chen, *Journal of Computational Chemistry*, 2012, **33**, 580–592.

[82] R. Wei, S. Huang, Z. Wang, X. Wang, C. Ding, R. Yuen and J. Wang, *Journal of Thermal Analysis and Calorimetry*, 2018, **136**, 651–660.

[83] M. A. Fernandez de la Ossa, M. Lopez-Lopez, M. Torre and C. Garcia-Ruiz, *TrAC - Trends in Analytical Chemistry*, 2011, **30**, 1740–1755.

[84] R. Ergun, J. Guo and B. Huebner-Keese, *Encyclopedia of Food and Health*, 2015, 694–702.

[85] Y. He, Y. He, J. Liu, P. Li, M. Chen, R. Wei and J. Wang, *Journal of Hazardous Materials*, 2017, **340**, 202–212.

[86] Y. T. Tsai, Y. Yang, H. C. Huang and C. M. Shu, *AIChE Journal*, 2020, **66**, 1–12.

[87] R. M. Brown, in *Cellulose: Molecular and Structural Biology*, Springer, 2007, ch. 15 - Cellu, pp. 257–284.

[88] D. N-s, *Cellulose*, 1994, **1**, 1–25.

[89] A. D. French and G. P. Johnson, *Journal of Chemical Information and Modeling*, 2008, **53**, 1689–1699.

[90] Q. Wang, H. Song, S. Pan, N. Dong, X. Wang and S. Sun, *Scientific Reports*, 2020, **10**, 1–18.

[91] W. Chen, G. C. Lickfield and C. Q. Yang, *Polymer*, 2004, **45**, 1063–1071.

[92] D. Meader, E. Atkins and F. Happey, *Polymer*, 1978, **19**, 1371–1374.

[93] K. Kulasinski, S. Keten, S. V. Churakov, D. Derome and J. Carmeliet, *Cellulose*, 2014, **21**, 1103–1116.

[94] G. Srinivas, X. Cheng and J. C. Smith, *Journal of Chemical Theory and Computation*, 2011, **7**, 2539–2548.

[95] S. Matsuoka, H. Kawamoto and S. Saka, *Journal of Analytical and Applied Pyrolysis*, 2014, **106**, 138–146.

[96] M. Zheng, Z. Wang, X. Li, X. Qiao, W. Song and L. Guo, *Fuel*, 2016, **177**, 130–141.

[97] Y. Zhang, C. Liu and X. Chen, *Journal of Analytical and Applied Pyrolysis*, 2015, **113**, 621–629.

[98] J. Wisniak, *Educactión química*, 2008, **19**, 71–81.

[99] C. F. Schönbein, *Improvement in Preparation of Cotton-Wool and Other Substances as Substitutes for Gun-Powder, US Patent US4874*, 1846.

[100] M. Freemantle, *Guncotton or nitrocellulose*, 2015, https://www.chemistryworld.com/podcasts/guncotton-or-nitrocellulose/9107.article.

[101] A. S. Greenberg and S. Broder, *Journal of the Patent Office Society*, 1926, **8**, 517–529.

[102] E. C. Worden, *Nature*, 1911, 69.

[103] F. A. Abel, *Improvement in the Manufacture of Gun-Cotton, US Patent US59888*, 1866.

[104] M. M. Fiume, W. F. Bergfeld, D. V. Belsito, R. A. Hill, C. D. Klaassen, D. C. Liebler, J. G. Marks, R. C. Shank, T. J. Slaga, P. W. Snyder, F. A. Andersen and L. Gill, *International Journal of Toxicology*, 2016, **35**, 50S–59S.

[105] M. G. Luft, *Industrial & Engineering Chemistry*, 1925, **8700**, 1037–1042.

[106] T. M. Letcher and N. S. Lutseke, *Journal of Chemical Education*, 1990, **67**, 361–363.

[107] J. E. Blomen, *Journal of the American Chemical Society*, 1895, 411–419.

[108] C. Christodoulatos, T.-L. Su and A. Koutsospyros, *Water Environment Research*, 2001, **73**, 185–191.

[109] M. K. Shukla and F. Hill, *Structural Chemistry*, 2012, **23**, 1905–1920.

[110] J. Brandrup, E. H. Immergut and E. A. Grulke, in *Polymer Handbook*, ed. A. Abe. and D. R. Bloch, Wiley & Sons, Ltd. Publications, Chichester, 4th edn., 1999, pp. 135–158.

[111] Q. Luo, T. Ren, H. Shen, J. Zhang and D. Liang, *Combustion Science and Technology*, 2018, **190**, 579–590.

[112] C. García-ruiz and M. Fernández de la Ossa, *Advances in Materials Science Research*, 2012, **7**, 201 – 220.

[113] *Nitrocellulose*, 2018, https://www.britannica.com/science/nitrocellulose.

[114] S. M. Pourmortazavi, S. G. Hosseini, M. Rahimi-Nasrabadi, S. S. Hajimirsadeghi and H. Momenian, *Journal of Hazardous Materials*, 2009, **162**, 1141–1144.

[115] K. Cieślak, K. Gańczyk-Specjalska, K. Drożdżewska-Szymańska and M. Uszyński, *Journal of Thermal Analysis and Calorimetry*, 2021, **143**, 3459–3470.

[116] M. A. Bohn, *Journal of Thermal Analysis and Calorimetry*, 2001, **65**, 103–120.

[117] M. A. Hiskey, K. R. Brower and J. C. Oxley, *Journal of Physical Chemistry*, 1991, **95**, 3955–3960.

[118] T. B. Brill and P. E. Gongwer, *Propellants, Explosives, Pyrotechnics*, 1997, **22**, 38–44.

[119] J. Kimura, *Propellants, Explosives, Pyrotechnics*, 1989, **14**, 89–92.

[120] P. Pfeil, H. Krause and N. Eisenreic, *Thermochimica Acta,*, 1985, **85**, 395–398.

[121] X. Ma, W. Zhu, J. Xiao and H. Xiao, *Journal of Hazardous Materials*, 2008, **156**, 201–207.

[122] A. P. Heiner, J. Sugiyama and O. Teleman, *Carbohydrate Research*, 1995, **273**, 207–223.

[123] S. Neyertz, A. Pizzi, A. Merlin, B. Maigret, D. Brown and X. Deglise, *Journal of Applied Polymer Science*, 2000, **78**, 1939–1946.

[124] K. Mazeau and L. Heux, *Journal of Physical Chemistry B*, 2003, **107**, 2394–2403.

[125] D. M. French, *Journal of Applied Polymer Science*, 1978, **22**, 309–313.

[126] R. Meyer, J. Köhler and A. Homburg, *Explosives*, Wiley-VCH, 6th edn., 2007, vol. 377.

[127] G. Wypych, *Handbook of Polymers*, 2016, 63–66.

[128] K. Kirschner, A. Yongye and S. Tschampel, *J Comput Chem2*, 8, **29**, 622–655.

[129] A. B. Poma, M. Chwastyk and M. Cieplak, *Cellulose*, 2016, **23**, 1573–1591.

[130] O. Beckstein, A. Fourrier and B. I. Iorga, *Journal of Computer-Aided Molecular Design*, 2014, **28**, 265–276.

[131] B. D. Rabideau, A. Agarwal and A. E. Ismail, *Journal of Physical Chemistry B*, 2013, **117**, 3469–3479.

[132] J. M. Andanson, E. Bordes, J. Devémy, F. Leroux, A. A. Pádua and M. F. Gomes, *Green Chemistry*, 2014, **16**, 2528–2538.

[133] M. K. Shukla and F. Hill, *Journal of Physical Chemistry A*, 2012, **116**, 7746–7755.

[134] M. K. Shukla and F. Hill, *Journal of Physical Chemistry A*, 2012, **116**, 4909–4921.

[135] C. Gibbon, P. Di Pietro, M. Storr, D. Broughton and C. K. Skylaris, *Physical Chemistry Chemical Physics*, 2023, **25**, 3190–3198.

[136] Y. Nishiyama, J. Sugiyama, H. Chanzy and P. Langan, *Journal of the American Chemical Society*, 2003, **125**, 14300–14306.

[137] L. S. Dodda, I. C. De Vaca, J. Tirado-Rives and W. L. Jorgensen, *Nucleic Acids Research*, 2017, **45**, W331–W336.

[138] A. Sihi, S. Lal and S. K. Pandey, *Condensed Matter*, 2017, 1–6.

[139] S. Sturniolo, P. J. Hasnip, P. Byrne, P. Hodgkinson, M. I. Probert and J. Kermode, *Support for advanced transition state search techniques in CASTEP*, 2022.

[140] D. Hait, A. Rettig and M. Head-Gordon, *Physiscs*, 2019, **21**, 21761–21775.

[141] D. Hait and M. Head-Gordon, *Angewandte Chemie - International Edition*, 2023, **62**, 1–7.

[142] N. Ferré, M. Filatov and M. Huix-Rotllant, *Topics in Current Chemistry 368 Density- Functional Methods for Excited States*, Springer, London, 2016.

[143] A. J. Lai, *Ph.D. thesis*, University College London, 2020.

[144] M. K. Shukla and F. Hill, *Journal of Physical Chemistry A*, 2012, **116**, 4909–4921.