Die Rolle der Molekulardynamik bei der Simulation der Form von organischen Molekülen / The Role of Molecular Dynamics in Simulating the Shapes of Organic Compounds

Leo Robinson

Klasse 8B

BRG XIX, Krottenbachstraße 11-13, 1190 Wien

1. Abstract

Molecular dynamics is a method of finding the three-dimensional arrangement of atoms in a molecule by simulating the nuclei as classical particles governed by Newtons' laws using a procedure called the Newton-Raphson method. In a mixed classical-quantomechanical approach, the positions of the nuclei, which are predicted using molecular dynamics, are used to calculate the orbital spin functions of the electrons in the molecule. This can be done using methods such as the Born-Oppenheimer approximation and the variational method. This project investigates classical and quantomechanical approaches to determining organic structures along with the methods and results of a molecular dynamics simulation program that was written as part of this project. The molecular dynamics program only performs the classical part of the mixed classical-quantomechanical approach. The full source code along with results and input data has been made publicly available at https://github.com/leorobinson11/Physics. Through the process of creating the molecular dynamics program, I aimed to not only work through the physics theory but also through the in-detail process of solving complex atomic systems through different approaches. This article illustrates the usefulness and limitations of molecular dynamics in predicting the shapes of complex molecules.

Table of Contents

2. Introduction

Hydrocarbons, which are molecules comprising only carbon and hydrogen, are the building blocks of all lifeforms; proteins, lipids etc. Carbon's ability to form four covalent bonds allows it to create complex structures through chains and rings (Greenshiel, 1958). The properties of hydrocarbons depend not only on the atoms they contain - the number of hydrogen and carbon atoms - but also on the structure that is created through which atoms are bonded to one another. Molecules that contain the same atoms but that are arranged in different bonding patterns are referred to in organic chemistry as isomers. The positions of atoms in relation to one another, through rotation or stretching of bonds, depend on physical forces rather than chemical reactions. Two molecules with the same atoms bonded in the same pattern can exhibit different properties due to a different spatial arrangement. These molecules that differ only in their spatial arrangement are named conformers. The focus of this work is to examine methods of finding the three-dimensional shapes of conformers where the bonding patterns are already established.

A historic example of the importance of understanding the shapes and behaviors of isomers is given by the drug thalidomide, which was prescribed to pregnant women in the 1950s to treat morning sickness. Thalidomide can take two forms: R(+) thalidomide and S(-) thalidomide; the two forms being mirrored versions of each other (Vargesson, 2015). R-thalidomide has sedative effects (the desired function), while S-thalidomide is teratogenic, meaning that it causes abnormalities after fetal exposure affecting over 10,000 children worldwide, named Contergan-children (Alliance & Health., 2010). This shows how the smallest changes in a molecule affect its behavior in vivo and the outcome when applied as a drug for treatment. Furthermore, it was later discovered that R thalidomide reacts in the stomach to form S thalidomide, further highlighting the importance of understanding molecules in relation to their dynamic surroundings.

The properties of a molecule are hard to predict due to several factors; foremostly that a molecule's properties only show in relation to its dynamic surroundings and the interactions with other molecules (Vyas, 2008). In addition, even the simplest of molecules aren't static. Rather, they are always changing in form, even in the absents of other molecules, giving them different properties as they evolve through time. Finally, as shown by thalidomide, they can only be understood in three dimensions.

Molecular dynamics has shown a wide range of applications in medicine and pharmaceutics. It is a particularly useful approach for large molecules such as proteins, where methods using quantum mechanical descriptions aren't able to solve for molecular shapes within a realistic time due to the complexity of the calculations involved. Molecular dynamics and quantum mechanical models are used to discover, design, and develop therapeutic chemical agents, by using simulations to determine the structures of drugs and test the effectiveness of binding to sites, receptors or proteins of known structure (Prieto-Martinez, 2019). With the reduced cost of computational drug discovery, it is hoped to find treatments for neglected diseases and perform multitarget drug discovery. The application of computer systems to condensed matter systems began in the 1950s, with two types of methods: Monte Carlo (MC) random sampling methods and molecular dynamics (MD), which are physics-based trajectory calculations (G. Ciccotti, 2022).

3. Methods

The motion of atoms in organic molecules was simulated using molecular dynamics, in which the movement of the nucleic positions in three dimensions is modelled using classical laws to find the configuration in which the potential energy of the molecule is at a minimum, which is the most stable state of the molecule. This approach was applied to alkanes - molecules containing single-bond hydrogen and carbons. The goal of the project is to create a molecular dynamics program for simple alkenes.

3.1. N-Body Problems and time complexity

When solving quantum systems, the main problem that needs to be dealt with is that the complex nature of the equations leads to a rapid increase in the time required to solve them with an increase in the number of atoms. For example, with a system containing only a hundred silicon atoms using a more exact approach such as the coupled cluster (CCSD) method, it would take 2000 years to solve with current computer processors, in comparison to a quantum computational method called Density functional theory (DFT), which would only take five hours (Tomasik, 2012). Throughout this article, almost every method used is applied to lower the computation time. Even though small systems of a couple of hundred atoms can be solved using quantum mechanical approaches, biological bodies such as proteins contain up to 2000 amino acids (Alberts, 2002), each with around 20 atoms, making the application of even the

fastest quantum computational methods impossible. One of the approximations that molecular dynamics makes is to treat the system as a classical system that evolves through time according to newtons laws of motion instead of the Schrödinger equation. For most N-Body problems, systems with three or more bodies, no closed analytical solution exists even using classical laws (Liao, 2013). Instead, the problem is described by Lorenz equations which consist of an infinite sum. As the bodies undergo chaotic motion, which is non-repeating, their paths cannot be described by equations with finite terms. Nevertheless, finding solutions to n-body problems isn't difficult when using approximate methods. Throughout this article, capital O notation will be used to describe the time complexity of different methods. Capital O notation and time complexity describe the way the computation time of an algorithm rises proportionally to an input size n (Simplilearn, 2022). With one interaction between each pair of bodies, as with the simplest case depicted in figure 1 below, many-body classical simulations have a polynomial time complexity of $O(n^2)$, as each body interacts with each other body, meaning every time a new body is added to the system the number of interactions increases by the number of bodies already present in the system. This is visualized in figure 1, where the circles represent the bodies and arrows represent the pairwise interactions.
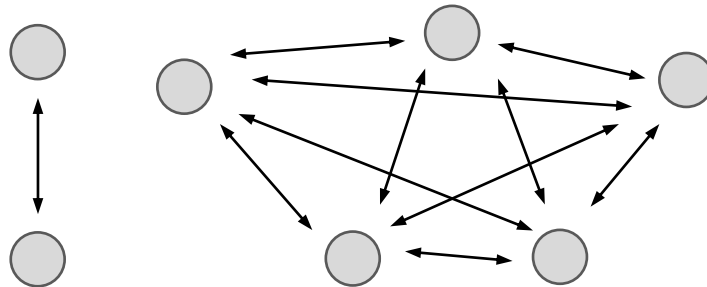
Figure 1: The multitude of interactions between classical bodies – left: 2-bodies, right: 5-bodies

Classical systems are easy to solve because they can be described by only the position and momentum of the set of bodies, which only requires 6 differential equations; 1 for each spatial dimension of the position and momentum, as a function of time. As we will see later, quantum systems are harder to solve due to an effect called quantum entanglement, which means that each particles' position and momentum depends on that of every other particle in the system (Horodecki, 2009). Methods such as the Born Oppenheimer approximation and the variational approach will be discussed, to show how the movements of the particles can be partially separated.

3.2. Visualization

To fully understand results from simulations of molecular dynamic simulations, the molecules need to be visualized. Unfortunately, as all particles on the subatomic level exist as probability density distributions, they are hard to visualize on a two-dimensional surface because of the amount of data that needs to be depicted to fully cover the characteristics of the molecule. The models used to display data on the molecules can be classified as either atomistic or abstract (Kozlíková, 2015). Atomistic models depict the expected positions (mean position) of all the atoms, displaying only the structure of the molecule, not the internal forces of the atoms. The first and most common form of the atomistic model is the planar representation, most commonly used in chemistry to display different isomers of simple molecules. The atoms are represented by their atomic symbol and bonds through a line. As the name suggests, planar representation is best suited for molecular structures that don't contain a large depth. This article will mainly use the ball and stick model, in which atoms are represented through colored spheres and the bonds through sticks connecting the bonded atoms. For extremely large molecules such as proteins, even the ball and stick model becomes messy. To show reoccurring structures, called alpha spirals or beta sheets, a cartoon view can be used, but as the molecules discussed in this article are relatively simple, they will be represented using the ball and stick model. As already mentioned, for a ball and stick model the atoms are modelled as classical bodies, with one distinct position. This means that the in the simplest case the information about the configuration of the entire molecule can be recorded in a matrix with a vector of x, y and z components for each atom. A data structure called a dictionary is used to keep records of which atoms are bonded to which. The dictionary consists of a set of all the atoms in the system, the keys, each with another associated set of all the atoms to which the atom is bonded. The dictionary functions so that from an input of an atom, it returns a set of the atoms to which it is bonded. This type of data structure is used to find the interacting sets of atoms used as inputs for the potential energy and force equations.

3.3. Force Fields

In both quantum and classical mechanics, the most stable configuration of a molecule is the configuration with the lowest potential energy (in classic mechanics), or the ground state (in quantum mechanics). This is because any system will be the most stable in the minimum energy

state, as that is the state it will evolve towards. Given some permutation of an initial position vector set, $\vec{x}(t = 0)$, obtained from experimental data, molecular dynamics explores the potential energy landscape to find a local minimum configuration. The potential energy equations ($U(\vec{x})$) are not derived from first principles and are only functions that aim to match the shape of the types of interactions. They are fitted to results from quantomechanical simulations or experimental data using a set of constants specific to each combination of pairs of atoms. As the constants are specific to pairs of atoms, the model implemented here is limited to hydrocarbons. This limits the number of combinations and the amount of data required. The molecular mechanics program developed was implemented from scratch, but using equations and constants from the Amber molecular dynamics program (Case, 2013). The bond and angle potentials (U) are modelled as springs described by Hooke's law as a quadratic potential.

$$U_i(\vec{x}_i) = \sum_{j \,\epsilon\, bonds,angles} \frac{1}{2} k_{ij} \left(r_{ij} - r_0 \right)^2 \tag{1}$$

The distance r between two atoms i and j is calculated by the magnitude of the vector connecting the two positional vectors, $\vec{x}_i$ and $\vec{x}_j$. The angle calculated is the angle formed by the two vectors connecting the atoms i and j to the central atom. The potential increases with bond length or angle r from the equilibrium displacement/angle (r_0), at which the bond/angle is at its lowest energy. The spring constant k_{ij} is, like the equilibrium displacement/angle, a constant depending on the atoms i, j and determines the stiffness of the bond/angle and the steepness of the potential energy function. For a chain of 4 or more atoms, the atoms can rotate around the central bond, as shown in figure 2 c). The associated potential energy is encapsulated in the torsion potential and acts between two atoms bonded to adjacent atoms.

$$U_i(\vec{x}_i) = \sum_{j \,\epsilon\, torsions} k_{ij}(1 - cos(n\omega)) \tag{2}$$

The spring constant k_{ij} determines the stiffness of the bond rotation and n is the periodicity of the rotations, the number of cycles in 2 pie radians, as the torsion can have multiple equilibrium angles in a full rotation. The images below, in Figure 2, were produced with the help of the Avogadro molecular modelling software. Mathematically, the torsion angle is defined as the angle formed by the two planes going through the atoms i and the two central atoms, and j and the two central atoms. Many simple molecular mechanics simulations only consider bonded

interactions, although this gives rise to the possibility that the atoms pass through each other (an unphysical behaviour) because there aren't any interactions between some non-bonded atoms. This illustrates the importance of including non-bonded interactions. The Van der Waal force is a general term for the interaction between non-bonded atoms in the molecule (Margenau, 1939). The two types of Van der Waal forces are the dipole-dipole interactions and the London dispersion force. Dipole-dipole interactions arise from the attraction or repulsion of two opposite or equal partial charges. Partial charges form in a molecule by certain atoms attracting the electrons more than others because of a higher electronegativity. London dispersion forces arise for the same reason, but the partial charge in the atoms exists only temporarily because of the random movement of electrons. In molecular mechanics, to reduce the computational time the two Van der Waal forces are summarized in one equation called the Lenard Jones potential (Maghfiroh, 2020).

$$U_i(\vec{x}_i) = \sum_{j \, \epsilon \, non-bonded} 4\varepsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{3}$$

The Lenard Jones Potential has terms for the attractive and repulsive forces between two particles. The attractive force acts due to the aforementioned partial charges as long as the distance between the two atoms r_{ij} is larger than the minimum distance sigma, σ, at which non-bonding particles are attracted to one another before being repulsed. As the particles move away from one another the force falls off quickly, like most nuclear forces as they only act over short distances, tending towards zero at infinite distance. The repulsive force, named Pauli repulsion, acts when the particles so-called Van der Waal orbitals cross as the electrons are not allowed to occupy the same space for long, due to Pauli's exclusion principle, which will be discussed later in detail. The repulsive force is far greater than the attractive force, with the potential rising to infinity as the distance between the particles approaches past the resting distance towards zero.
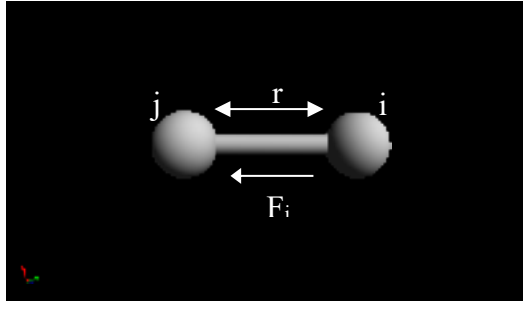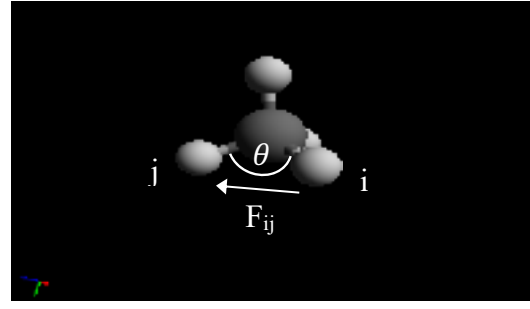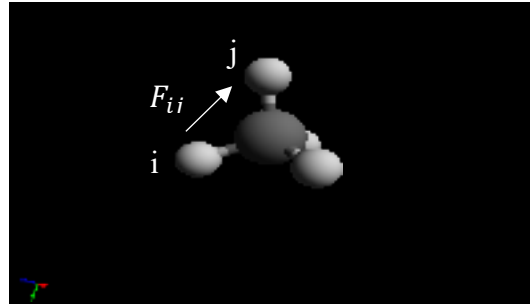
a) Bond Force   b) Angle Force
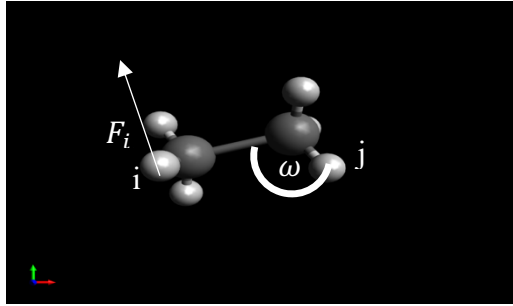
c) Torsion Force   d) Van der Waal Force

Figure 2: Figure 2: Types and directions of forces acting on atoms in molecules

3.4. Energy Minimization Method – Newton - Raphson

The approximation of atoms as classical particles allows Newton's second law - that force is the time derivative of momentum - to be applied to model the atom's trajectories (Shattuck, 2003), which forms the basis for the Newton-Raphson Method. The force $(\vec{F}_{ij})$ on a single-point particle i, exerted by another particle j, is by definition the negative of the gradient of potential energy over space, which is the reason why bodies will evolve into a lower potential energy configuration (Shattuck, 2003, p. 29)

$$\vec{F}_{ij} = -\vec{\nabla} U_{ij} \hat{r}_{ij} \qquad (4)$$

Where the gradient vector in three dimensions is defined as:

$$\vec{\nabla} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}, \qquad (5)$$

9

where $\hat{r}_{ij}$ is a unit vector comprising of a linear combination of the basis vector set, determining the direction of the exerted force. As a unit vector, $\hat{r}_{ij}$ does not affect the magnitude of the force:

$$|\widehat{r}_{ij}| = \sqrt{\vec{r_x}^2 + \vec{r_y}^2 + \vec{r_z}^2} = 1. \tag{6}$$

For a single-point particle, the momentum (p) is the product of mass (m) and velocity (v):

$$\overrightarrow{F}_i = \frac{d\vec{p}_i}{dt} = m_i \frac{d\vec{v}_i}{dt} = m_i \vec{a}_i. \tag{7}$$

This allows for rearranging to solve for the acceleration ($\vec{a}_i$) of a particle i from the force exerted on the particle:

$$\vec{a}_i = \frac{\overrightarrow{F}_i}{m_i} \tag{8}$$

Three body problems become analytically unsolvable, so the double integral of the acceleration is approximated as being constant and the change in the atom's velocity or acceleration is approximated by calculating it over a short period delta t, by performing numerical integration (Shattuck, 2003, p. 29). In the equations below it is assumed that the atoms are initially at rest. In Newton-Rapson the equation of velocity is calculated as

$$\vec{v}_i(t + \Delta t) = \int_t^{t + \Delta t} \frac{\vec{F}_{ij}}{m_i} dt \approx \vec{v}_i(t) + \frac{\sum_{j=1}^{n} \vec{F}_{ij}}{m_i} \cdot \Delta t \tag{9}$$

and position

$$\vec{x}_i(t + \Delta t) = \iint_t^{t + \Delta t} \frac{\vec{F}_{ij}}{m_i} dt^2 \approx \vec{x}_i(t) + \vec{v}_i(t + \Delta t) \cdot \Delta t \tag{10}$$

The above calculations are performed repeatedly, with the result from the previous time step being used as the input for the next. The Newton-Raphson method is the simplest of molecular mechanics, which was one of the reasons it was chosen as the simulation method. In addition to its simplicity, it gives an insight into the molecule's movement and vibrations, as opposed

to Monty Carlo-based Methods, which sample random positions to find the minimum energy. The largest downside of Newton-Raphson simulations is their tendency to get stuck in local minima and not explore the larger potential energy landscape. The basic procedure is described again by the flow chart in figure 3.
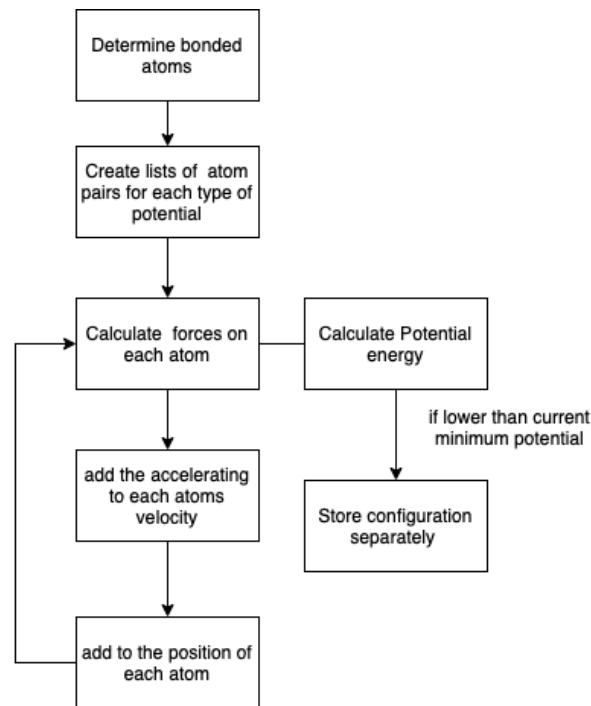
Figure 3: Steps in the molecular dynamics program

In summary, the force on each atom is calculated and used to determine the acceleration of each atom. Instead of integrating to calculate the change in velocity and then integrating again for the change in position, the Newton-Rapson method approximates the forces as being linear and instead multiplies by some small-time increment delta *t*. The whole process is repeated with the position vector calculated in the last time step as the input. Between each recalculation, the potential energy of the current configuration is compared to that of the previous minimum, and recorded along with the positional vector, if smaller.

3.5. Quantum mechanics postulates

This section briefly outlines the fundamentals of quantum mechanics relevant to any methods discussed in this article. The entire information of a system in quantum mechanics is described by a wave function, usually denoted by an upper-case psi (Ψ). Psi as a function of time and

position ($\Psi(\vec{x}, t)$) has a wavenumber (k), affecting the height of the wave depending on position, and an angular frequency (ω) shifting the wave over time. Psi is a complex function $\Psi \in \mathbb{C}$. Psi squared of $\vec{x}$ is the probability density function of finding the particle at position $\vec{x}$ (Phillips, 2003).

$$\mathbb{P}(X = \vec{x}) = |\Psi(\vec{x})|^2 dx \qquad (11)$$

This allows the expected values of any property associated with the wave equation to be calculated. Nevertheless, the certainty in the value will depend on the equation and it is not possible to have a wave equation with absolute certainty in all quantum numbers. Every characteristic of a wave function has an associated operator that acts on the wave equation, transforming it in some way. Operators and objects form a basis for much of the mathematics performed in quantum physics. Objects like the wave function can be compared to vectors in linear algebra, as in that they hold some information. Operators act on objects to transform them in some way to form a new object, a lot like matrices. In quantum physics, each characteristic of a system has an associated operator, denoted by the symbol of the operator with a circumflex on top ex. $\hat{x}$. For every linear transformation of a matrix, there is a set of eigenvectors. An eigenvector associated with a transformation is a vector that only changes in magnitude and not direction when acted on by the matrix. The eigenvalue is the factor by which the magnitude of the vector is changed. A similar concept exists for operators. An operator's eigenfunctions are functions that only change by some constant factor, the eigenvalue, when acted on by the operator:

$$\hat{A}\varphi = A\varphi \qquad (12)$$

Every eigenfunction of an operator is a function with no uncertainty in the value associated with the operator. The most important equation in quantum mechanics is the Schrödinger equation, which provides a means to find the wave equation for any system:

$$\hat{H}\Psi = \hat{E}\Psi \qquad (13)$$

which associates the Hamiltonian $\hat{H}$, the sum of the potential and the kinetic energy of the system, to the energy operator, all acting on the wave function. If the wave function is an eigenfunction of the Hamiltonian, the eigen Schrödinger equation allows the calculation of

energy. As the energy is constant with respect to time, this form of the Schrödinger equation is also called the time-independent form of the Schrödinger equation:

$$\left[\frac{\hat{p}^2}{2m} + V(\hat{x})\right]\Psi(\mathrm{x}, \mathrm{t}) = E\,\Psi(\mathrm{x}). \tag{14}$$

As the Schrödinger equation is a linear function, any linear combination of solutions to the Schrödinger equation are also solutions to the Schrödinger equation. This will be important, as most solutions will be found through a linear combination of a basis set of eigenfunctions, as will be seen later on.

3.6. Born Oppenheimer Approximation

One of the most significant shortcomings of molecular mechanics is that it only models the atomic structure using the positions of the nuclei and does not describe electron behavior in any way. Quantum effects being relative to size, however, affect electrons much more than the nuclei, making a classical approach to modelling the behavior even more inaccurate. This is why in some cases quantum mechanical methods should be used. We will see, however, that for a solvable system, even for an implementation to be solved using computers, a few approximations need to be made.

As in every quantum system, in an n-body system, we first solve for the Hamiltonian. In a system of N nuclei, with positions and momenta $\vec{X}_\alpha, \vec{P}_\alpha$, and n electrons, with positions and momenta $\vec{x}_i, \vec{p}_i$, the total Hamiltonian for the entire system is simply the sum of all of the kinetic energies of the nuclei as well as all the pairwise potential energy interactions, nucleus-nucleus, electron-nucleus and electron-electron;

$$H = \sum_{\alpha=0}^{N} \frac{\hat{P}_\alpha^{\,2}}{2M_\alpha} + \sum_{i=0}^{n} \frac{\hat{p}_i^{\,2}}{2m} + V_{NN}(\vec{X}) + V_{eN}(\vec{x}, \vec{X}) + V_{ee}(\vec{x}). \tag{15}$$

This is true for any system, independent of the potential energies used. In this version, the electrons' wave functions are dependent not only on their positions and motion but also on those of the nuclei. In quantum mechanics, a system undergoes an adiabatic process when the gradually changing conditions allow the system to adapt to the conditions, meaning that if the

system starts in an eigenstate of the Hamiltonian the system will again fall back into the eigenstate of the Hamiltonian given that it can adapt to the change in conditions fast enough (Born, 1928). This allows us to perform a timescale separation on the n-body Hamiltonian. Timescale separation allows us to remove the dependence of the fast-evolving bodies on the movement of the slow-moving ones, as they can be approximated as being stationary (Gunawardena, 2014). The goal of the Born-Oppenheimer approximation is to separate the electronic and nucleic components of the Hamiltonian, removing dependence on each other, lowering the computation time, and allowing us to finally write the total wave function as a product of the nuclear and electric terms (Sherrill, 2005).

$$\widehat{H}_e = \sum_{i=0}^{n} \frac{\hat{p}_i^2}{2m} + V_{eN}(\vec{x}, \vec{X}_a) + V_{ee}(\vec{x}) + V_{NN}(\vec{X}_a) \tag{16}$$

$$\Psi(\vec{x}, \vec{X}) = \Psi_e(\vec{x}, \vec{X}_a)\Psi_N(\vec{X}) \tag{17}$$

What prevents the separation of the Hamiltonians is the cross term $V_{eN}(\vec{x}, \vec{X})$, which depends on both electron and nuclear positions. As the electron velocities are far greater than those of the nuclei, we can consider a fixed nuclear skeleton, due to the larger M_α. We fix the nuclear positions at some position $\vec{X}_a$ and solve the wave function for $\Psi(\vec{x}, \vec{X}_a)$ with only a parametric dependence on the positions of the nuclei. The electronic Hamiltonian is independent of the nucleic kinetic energy, as the nuclei are stationary, and the nucleus-nucleus potential energy is simply a constant. To solve for the electronic wave function, we simply solve the Schrödinger equation, named the electronic Schrödinger equation, using the electronic Hamiltonian.

$$\left(\sum_{i=0}^{n} \frac{\hat{p}_i^2}{2m} + V_{eN}(\vec{x}, \vec{X}_a) + V_{ee}(\vec{x}) + V_{NN}(\vec{X}_a) \right) \Psi(\vec{x}, \vec{X}) = \widehat{E}_e(\vec{X})\Psi_e \tag{18}$$

$$\widehat{H}_e(\vec{x}, \vec{X}_a)\Psi(\vec{x}, \vec{X}) = \widehat{E}_e(\vec{X})\Psi_e \tag{19}$$

Although it has not been mentioned until now, the same adiabatic approximation is made in molecular dynamics, separating the motion of the nuclei from that of the electrons (dynamics, 1998). Due to their high velocities, the average position of the electrons can be fixed relative

to the positions of the protons. The electronic Schrödinger equation gives us a method of solving the electrons' wave functions given the positions of the nuclei. But as we calculate the nucleic positions through molecular dynamics, we can set $\vec{X}_a$ to be equal to the classically-calculated nuclei positions for any given time.

3.7. Variation principle

Just as in molecular dynamics, the aim of most numerical methods of solving the Schrödinger equation is to minimize the energy of the system, as the lowest energy configuration, or ground state, is the most stable (Epstein, 1974). In the variation method, the wave function describing the system is changed until the lowest possible energy is found. To do this, a basis set of functions, a set of trial functions, is first chosen (Nagy, 2017). Each basis function has an associated constant c which will be varied to find the linear combination of functions that give the lowest energy, forming the total wave function.

$$\Psi = \sum_{i=0}^{k} c_i \Psi_i \tag{20}$$

As we have a system with multiple electrons, we will have a many-electron wave function, comprising of multiple single-electron wave functions. The Hartree product is the simplest method of writing the total wave function in terms of the single electron wave functions, by writing the total wave function as a product of the individual spin functions ($\chi_i(\vec{x}_i)$) (Gharabaghi, 2016). The one over n factorial square rooted is just the normalizing factor, so that the square of the function follows the rule for probability density functions of integrating to 1.

$$\Psi(\vec{x}_0, \vec{x}_1 \ldots \vec{x}_n) = \frac{1}{\sqrt{n!}} \chi_0(\vec{x}_0) \chi_1(\vec{x}_1) \ldots \chi_n(\vec{x}_1) = \frac{1}{\sqrt{n!}} \prod_{i=0}^{n} \chi_i(\vec{x}_i) \tag{21}$$

A spin function is a wave function that describes a particle not only as a function of the spatial, x, y and z coordinates, but also the particle's property named spin. A particle's spin is a quantum property equivalent to angular momentum in classical mechanics, which determines a lot of particle properties such as its magnetic moment (Varshslovich, 1988). Unfortunately, the Hartree product does not follow the Pauli exclusion principle. The Pauli exclusion principle states that two fermions, which are particles such as electrons that have half-odd-integer spin

(1/3, 3/2, 5/2, etc.) and which make up all the mass in the universe, cannot occupy the same quantum states at the same time (Kaplan, 2016). Because of this principle, a particle cannot have the same spin as a particle it shares the same position with, meaning the two particles cannot exist in the same space for long. The Pauli exclusion principle is what prevents all the matter in the universe from collapsing in on itself.

To show why the Hartree product is an invalid way of writing the total wave function of a fermion system, an exchange operator needs to be introduced. The exchange operator $\hat{P}_{ij}$ acts on a multibody wave function by switching the dependencies of electrons i and j;

$$\hat{P}_{01}\chi(\vec{x}_0, \vec{x}_1) = \chi(\vec{x}_1, \vec{x}_0). \tag{22}$$

All functions are eigenfunctions of the exchange operator. If the exchange operator is applied twice, we should return to the original function;

$$\hat{P}_{01}^{\ 2}\chi(\vec{x}_0, \vec{x}_1) = \hat{P}_{01} \pm \chi(\vec{x}_1, \vec{x}_0) = \chi(\vec{x}_0, \vec{x}_1), \tag{23}$$

meaning that the operator acting on the original functions can have both eigenvalues of 1, -1. Functions formed from Hartree products have eigenvalues of 1 due to the commutability of products:

$$\hat{P}_{01}\, \chi_0(\vec{x}_0)\chi_1(\vec{x}_1) = \chi_0(\vec{x}_1)\chi_1(\vec{x}_0). \tag{24}$$

The equation before and after the substitution is the same due to the indistinguishability of electrons. The exchange in electrons resulting in the original function cannot be accurate for a fermion system as it would require them to have the same quantum numbers, defying Pauli's exclusion principle, telling us that the above equation can only represent a system of bosons, which are integer-spin subatomic particles (such as gluons). A fermion system will always have the form of

$$\hat{P}_{01}\,(\chi_0(\vec{x}_0)\chi_1(\vec{x}_1) - \chi_0(\vec{x}_1)\chi_1(\vec{x}_0)\,) = \chi_0(\vec{x}_1)\chi_1(\vec{x}_0) - \chi_0(\vec{x}_0)\chi_1(\vec{x}_1)$$
$$= -(\chi_0(\vec{x}_0)\chi_1(\vec{x}_1) - \chi_0(\vec{x}_1)\chi_1(\vec{x}_0)), \tag{25}$$

16

as the switching of electrons causes the position sign to switch because of the rearrangement of the terms. This is why bosons are described as symmetric and fermions are described as antisymmetric. As the solution to a determinant involves the permutation of all the items of the columns and row, the simplest form of writing the full wave equation is in a determinant named the Slater Determinant (Matsen, 1969), with a column for each wave function and a row for each electron. For example, a two-particle system would be written as

$$\Psi(\vec{x}_0, \vec{x}_1) = \frac{1}{\sqrt{2}} \begin{vmatrix} \chi_0(\vec{x}_0) & \chi_1(\vec{x}_0) \\ \chi_0(\vec{x}_1) & \chi_1(\vec{x}_1) \end{vmatrix} = \frac{1}{\sqrt{2}} \left(\chi_0(\vec{x}_0)\chi_1(\vec{x}_1) - \chi_0(\vec{x}_1)\chi_1(\vec{x}_0) \right). \qquad (26)$$

The Slater determinant can be generalized for any n-body System;

$$\Psi(\vec{x}_0, \vec{x}_1 \dots \vec{x}_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_0(\vec{x}_0) & \cdots & \chi_n(\vec{x}_0) \\ \vdots & \ddots & \vdots \\ \chi_0(\vec{x}_n) & \cdots & \chi_n(\vec{x}_n) \end{vmatrix}. \qquad (27)$$

Because of this, the basis set for a multielectron system comprising of multiple Slater determinants. In the variation theorem small changes to wave function $\delta\Psi$ are added until the energy stops changing, indicating that the energy has reached at least a local minimum:

$$\Psi \to \Psi + \delta\Psi. \qquad (28)$$

To find the total energy of the system, the Hamiltonian operator has to act on the wave function. But as the square of the wave function gives the probability distribution of the wave function being in a certain state, to find the expectation value of any operator the transformed wave function has to be averaged over all states. Dirac notation is used as a shorthand for writing this expression, with the square brackets representing the expectation values of the operator acting on the wave function (Marshman, 2017):

$$E(\Psi) = \langle \Psi | \hat{H}_e | \Psi \rangle = \int_{-\infty}^{\infty} \Psi^* \hat{H}_e \Psi d\vec{x}. \qquad (29)$$

After the variation is added, the adjusted energy becomes the expected value of the Hamiltonian acting on the new wave function (Vallance, 2021):

$$E(\Psi + \delta\Psi) = \langle\Psi + \delta\Psi|\hat{H}_e|\Psi + \delta\Psi\rangle. \tag{30}$$

This allows us to know when at least a local minimum has been reached - when the variation in the energy is zero. The total wave function comprises of multiple wave functions, so to find the expected energy, the energy has to be averaged over the probabilities of each function given every other function:

$$E(\Psi) = \langle\Psi|\hat{H}_e|\Psi\rangle = \sum_{i=0}^{k}\sum_{j=0}^{k} c_i^* c_j \langle\Psi_i|\hat{H}_e|\Psi_j\rangle. \tag{31}$$

As the square of the wave function is a probability density distribution, it has to be normalized. This means that inaccuracy in the energy will arise due to the denormalization of wave function from the addition of the variation in the wave function. To overcome this, the equation is held subject to the condition that the expectation value of the Hamiltonian equals that of the energy of the normalized wave function, forming the below equality:

$$\sum_i\sum_j c_i^* c_j \langle\Psi_i|\hat{H}_e|\Psi_j\rangle - E\left(\sum_i\sum_j c_i^* c_j \langle\Psi_i|\Psi_j\rangle - 1\right) = 0. \tag{32}$$

To vary this equality, we can vary the constants of either variable in the function:

$$\sum_i\sum_j \delta c_i^* c_j \langle\Psi_i|\hat{H}_e|\Psi_j\rangle \;-\; E\,\delta c_i^* c_j \langle\Psi_i|\Psi_j\rangle$$
$$+ c_i^* \delta c_j \langle\Psi_i|\hat{H}_e|\Psi_j\rangle - E\,c_i^* \delta c_j \langle\Psi_i|\Psi_j\rangle = 0. \tag{33}$$

To simplify the term, a Hamiltonian matrix H, with the rows and columns consisting of the expected Hamiltonian integrals of the combinations of the wave functions, and an overlap matrix S, with the overlap integrals, are defined. The matrix S of the integrals of the combinations of wave functions is called the overlap matrix because one of the functions has its complex conjugate taken. The sum over j and i can be separated into two terms forming the below equality:

$$\sum_i \delta c_i{}^* \left(\sum_j H_{ij}c_j - ES_{ij}c_j \right) + \sum_j \delta c_j \left(\sum_i H_{ij}c_i{}^* - ES_{ij}c_i{}^* \right) = 0. \qquad (33)$$

Because of the symmetry of the terms, for the entire term to be zero, each of the terms individually needs to be zero:

$$\sum_j H_{ij}\, c_j = E \sum_j S_{ij}c_j \qquad (34)$$

The above term can also be written in matrix form, where H is a matrix of the different expected energies for each combination of wave functions, c is a vector of the constants and S is the overlap matrix that forces normalization condition on the wave function.

$$Hc = ESc \qquad (35)$$

In the special case where S is the identity matrix, and we have an orthonormal basis set, the overlap matrix vanishes, simplifying the equation to the eigenvalue matrix form of the Schrödinger equation, which means that by diagonalising the equation (a method of finding eigenvalues) we can find the eigenvector which is the coefficient vector c. The issue is that c and H both depend on each other simultaneously. To overcome this problem an algorithm called a self-consistent field procedure is used (J.P.Desclaux, 1969). The self-consistent field procedure is performed as follows: First of all, the molecule of interest is specified, which requires the set of coordinates for the nuclei which, using the Born-Oppenheimer approximation, stay fixed in the variational method. In a mixed quantum-classical approach, the position of the nuclei is found using a molecular dynamics simulation. The other starting condition that has to be determined before the self-consistent field procedure can begin is the set of basis functions. As the systems will almost always contain multiple electrons, the basis set will consist of a linear combination of Slater determinants, each containing all the different spin-orbital functions for each of the electrons. Using the initial predetermined set of basis functions, the computer calculates the overlap matrix. The initial set of coefficients is generated randomly. Using the set of random coefficients, the Hamiltonian matrix is formed. The resulting equation is solved using the new Hamiltonian matrix to find a new set of coefficients and energies. This can be done by diagonalising the Hamiltonian matrix and solving the equation as an eigenvalue equation with c as the eigenvector and E as the eigenvalue. The

procedure is repeated using the new coefficients to form the Hamiltonian matrix, which in turn is used to find the new set of coefficients. This procedure is repeated until the set of coefficients stops changing from one step to the next.

3.8. Finding Shapes through Spectroscopy

Both molecular dynamics and quantomechanical methods need a starting set of atomic positions. A common method for determining these, for proteins, is X-ray crystallography, in which the diffraction of x-rays is used to calculate the approximate positions of the atoms in a crystalized molecule. Unfortunately, this type of procedure requires expensive equipment. Prior to this project, simpler experiment was conducted (sperate from the molecular dynamics program) to determine atomic structure using a simpler method based on visible light absorption spectroscopy. Absorbance spectra of the organic molecule beta carotene were used to show how the initial positions of atoms in molecules can be inferred using experimental data. In brief, 4 ml of acetone, 3 ml of water and 3 ml of petroleum were added (in that order) to 1-2 grams of ground leaf material, generating a layer of solution containing only the plant pigment. This formed on top of the leaf material and was separated and applied to a chromatography strip. The strip was lowered into a chromatography jar, which contained 5ml of acetone-petroleum ether 10:90, which caused the individual pigments to diffuse to different positions along the chromatography strip, where they could be identified by colour. The strips were cut at the intersections between colours and dissolved in separate cuvettes which contained a small amount of acetone. A spectrometer was used to measure the light absorption of the pigment over wavelengths ranging from 380 nm to 720 nm at 10 nm intervals (Kennedy, 2021).

In beta carotene, a plant pigment, the chain in the molecule consists of a chain of single and double bonds. The electronic orbitals of the individual atoms in the chain overlap, forming one larger pi-orbital that covers the entire chain. This means that any of the electrons can move freely along the entire chain of carbon atoms, along the pi-orbital. In quantomechanical terms it is more accurate to say that the electrons are highly delocalized, meaning that they can exist anywhere along the pi-orbital. This means that the energy required to promote electrons to higher orbitals is less. When light collides with the pi orbital, the wavelengths with the energy required to promote the electrons are absorbed. There exists a relationship between the wavelength with the highest absorption and the length of the chain in beta-carotene. More

exactly, if we assume that the electrons can move along only the pi orbital and nowhere else, then we can set up a Hamiltonian with a one-dimensional potential energy of zero as long as the electron is between 0 and the length of the chain, and infinity everywhere else. This particular solution to the Schrödinger equation is called a particle in a box or infinite well model and gives an energy shown below obtained through solving for the eigenvalues of the Energy-eigenfunction in the time-independent Schrödinger equation:

$$\Delta E = \frac{(n_i{}^2 - n_j{}^2)h^2}{8mL^2}. \tag{36}$$

Pauli's exclusion rule, which states that two fermions cannot have the same quantum numbers at the same time, means that no more than two electrons can have the same energy levels, one for each direction of spin. This means that beta carotene pi orbital, which has 22 electrons, will have eleven energy levels filled in its ground state. This means that the lowest energy photon that is absorbed, promotes electrons from the eleventh to the twelfth energy shell:

$$\Delta E_{min} = E_{12} - E_{11}. \tag{37}$$

After some simplification, the formula for the difference in energy between the highest occupied and lowest unoccupied energy level for N electrons becomes:

$$\Delta E = \frac{(2N + 1)h^2}{8mL^2} \tag{38}$$

The energy of a photon like every other wave is its frequency times the planks constant.

$$E = hk \tag{39}$$

The relationship between the wavelength and the frequency of light

$$k = \frac{c}{\lambda} \tag{40}$$

forming a relationship between the absorbed wavelength and the length of the conjugated chain given the number of occupied energy levels.

$$\frac{c}{\lambda} = \frac{(2N+1)h}{8mL^2}$$

(41)

Rearranging allows us to calculate the length of the chain using the wavelength with the highest absorption, which is determined experimentally as described above:

$$L = \sqrt{\frac{(2N+1)h\lambda}{8mc}}.$$

(42)

4. Results

4.1. Molecular Dynamics Simulations

To test whether the molecular dynamics program not only result in lower potential energy configurations but also simulates a realistic evolution of the atoms, the simulation was left to run beyond when a local potential energy minimum was found. To test the program to begin the simplest molecules were used, as the chaotic behavior increases with the number of atoms, which makes the results harder to analyze. The simplest possible molecule to simulate is a Hydrogen bonded to a carbon. Figure 4 below shows the potential energy of the bond connecting the two atoms.
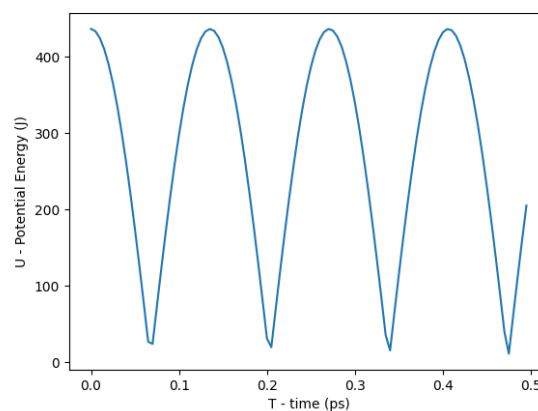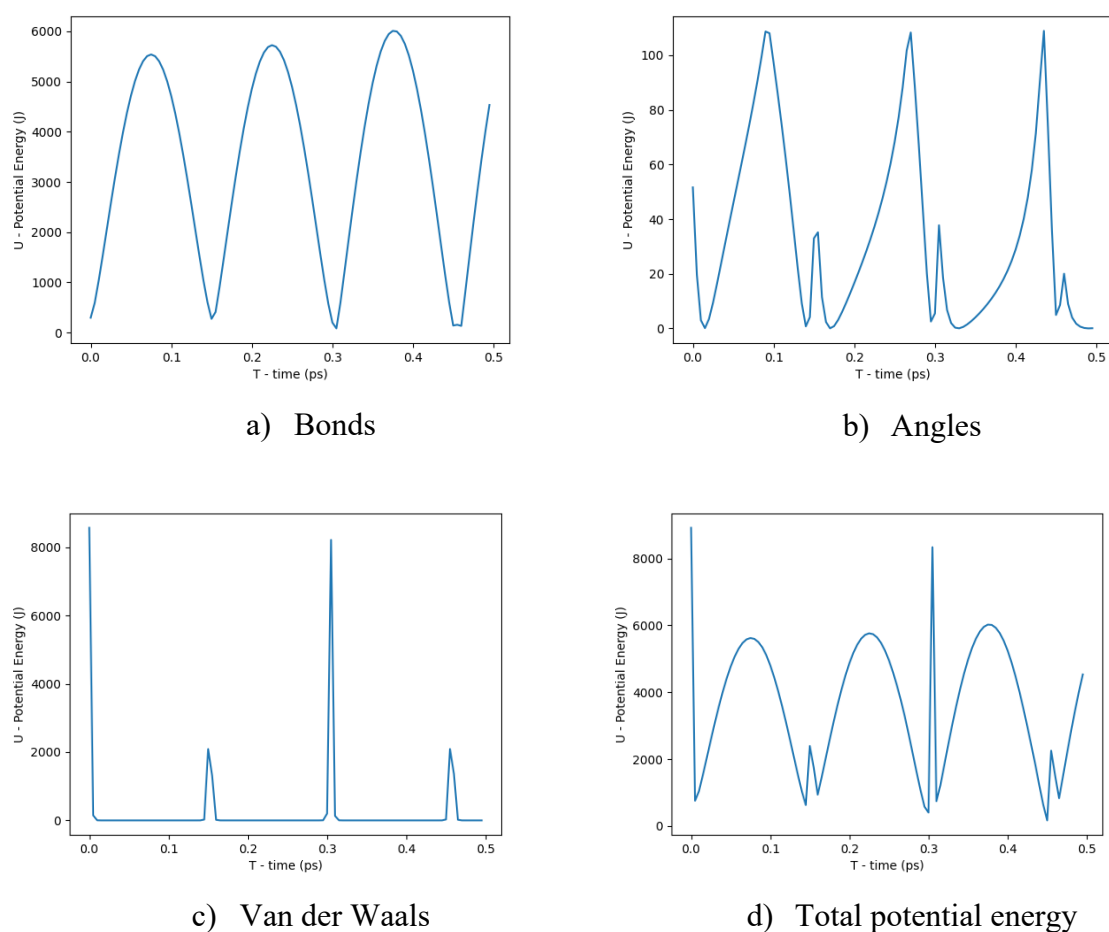

Figure 4: Potential energy of the bond (J) over time (ps) in CH, calculated using the molecular dynamics program during the simulation

Because of the conservation of energy, if the system stays isolated and no energy can enter or leave the system, the total amount of energy stays constant. As the total amount of energy (E) is the sum of kinetic (K) and potential energy (U), the kinetic energy has to increase as the potential energy decreases.

$$E = K + U \qquad\qquad (43)$$

This means that as the molecule reaches the minimum potential energy, it also reaches its maximum kinetic energy (and velocity). As it stretches beyond the resting bond length the force acts in the opposite direction as the atom's velocity decelerating the atom until it changes directions. This leads the atoms to undergo oscillating movement. As the system is closed, meaning the total amount of energy stays constant, the atoms undergo periodic movement forever. In the case of the CH molecule, as the only degree of freedom is along the bond, the oscillation behavior is relatively simple. Even with the addition of only one more atom, for example with CH_2 as can be seen in figure 5 below, the molecule exhibits more complex behavior. As every force affects the atoms' positions, which affects the magnitude of the potential energy, the different forces interfere with each other, resulting in more complex behavior. For example, as the angle force attracts the atoms to one another, at some point the atoms come close enough for the Van der Waal force to repel the atoms from one another, changing the angle between them, resulting in a different angle potential than what would occur if the angle force acted by itself. Note that the angle force is graphed on a smaller scale, from 0 to only 100 joules, than the other forces. This can be seen in figures 5 b) and c) as the lower peaks of the angle potential match the peaks of the Van der Waal potential. This also means that in some cases, potential energies will increase even at the beginning of the simulation as other acting forces are stronger, for example, the bond potential as can be seen in figure 5 a).

a) Bonds

b) Angles

c) Van der Waals

d) Total potential energy

Figure 5: Calculated potential energies of the Bonds, Angles and Van der Waals (J) and Total over time (ps) in CH2

Even though the movement is more complex, the molecules still return to a similar shape and will eventually return to the same shape. However, the more atoms present in the system, the more complicated the path that the atoms take is, until returning to their original positions. The smallest real molecule which included all the different types of potentials was C_2H_6. Figure 6 shows the different energies over time. The torsion potential was omitted as the energy was so low that it was affected too much by the other forces, meaning it does not give any information about the system.

a) Bonds

b) Angles

c) Van der Waals

d) Total potential energy
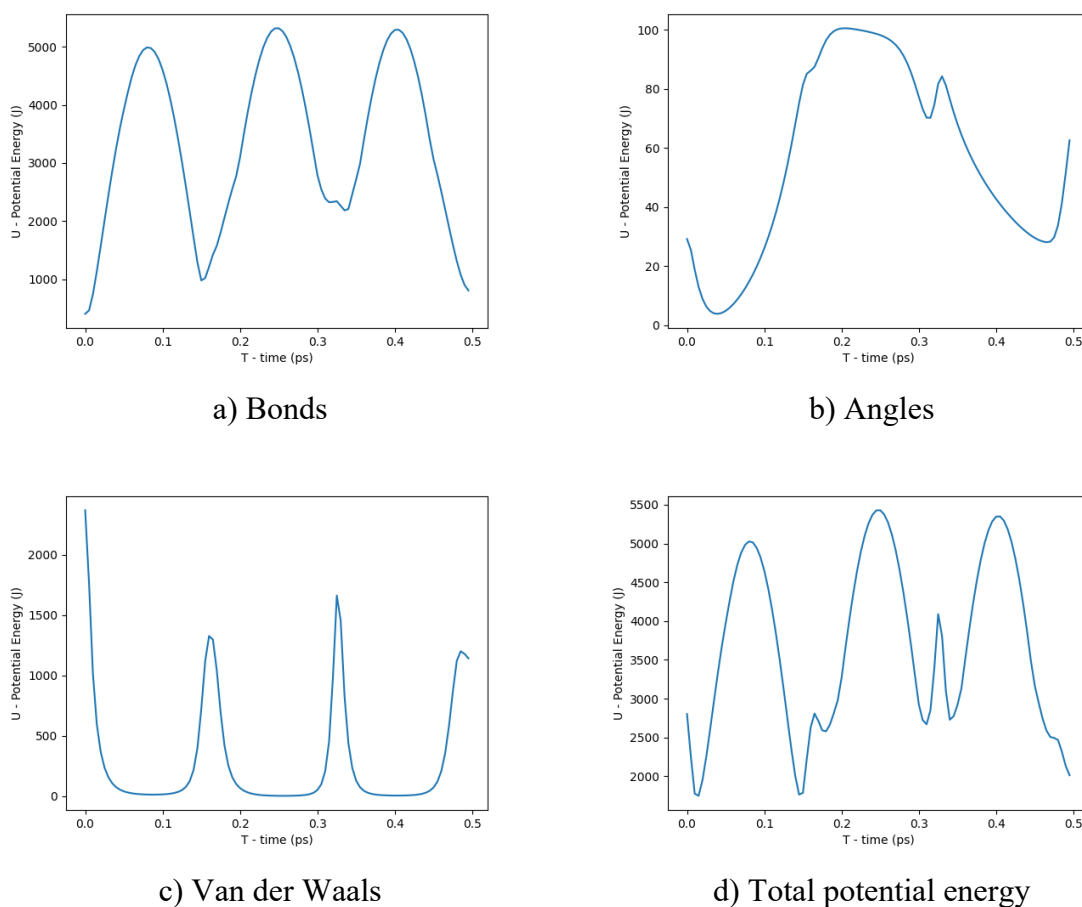
Figure 6: Potential energies of the bonds, angles and Van der Waals (J) over time (ps) in
C2H6

Interestingly, that the atom will fall into a minimum energy configuration very early on during the simulation, indicating that maybe only a local minimum is found. Because of the nature of molecular dynamics, preventing the atoms from moving past their starting distances when moving away from the resting values of the potential energy functions, the atoms may not be able to explore the larger energy landscape. Solutions to this could be to run the simulation multiple times from different starting positions or to combine molecular dynamics with a sampling approach such as Monte Carlo, which puts the atoms in random positions to test the larger energy surface, to find an optimal starting configuration. The results of a simulation of C_2H_6 can be seen in figure 7 below, where both images depict the same molecule from different angles.
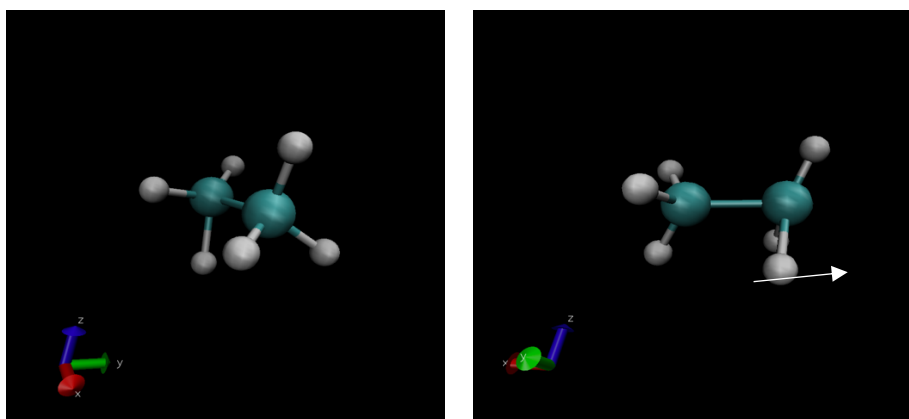
*Figure 7: Calculated minimum energy configuration of C2H6.*

The largest error in the outcome is that two of the external hydrogen atoms, on the right, are angled towards the central atoms and should be angled outwards as shown by the arrow on the right of figure 7 on the right. This type of angle would be taken care of by an out-of-plane potential, which was omitted for sake of simplicity. The other potential sources of error are discussed in the next chapter

4.2. Sources of Simulation errors

One of the largest sources of error in molecular dynamics is truncation error, which comes from the numerical double integration of the force to find the position. Truncation error is the divergence between the sum and the actual integral, as non-constant functions cannot be integrated by multiplying by the change in the integration variable.

$$\varepsilon = \int_a^b f(t)dt - \sum_{i=1}^n f(t_i)\,\mathrm{dt} \tag{44}$$

The smaller delta t is chosen to be, the smaller the truncation error in the simulation is. However, smaller a delta t leads to a longer computational time, as more timesteps (and thereby, repetitions) are needed for the same time evolution. To test whether there were any coding errors, the distance of the centre of mass from its original position was measured over time. This test was performed as the centre of mass should stay constant, as because of the symmetry of Newton's equation the forces acting on the two forces are equal in opposite directions so that the changes in the centre of mass cancel out. Nevertheless, there was a very small change in the centre of mass, as can be seen in figure 8 below.
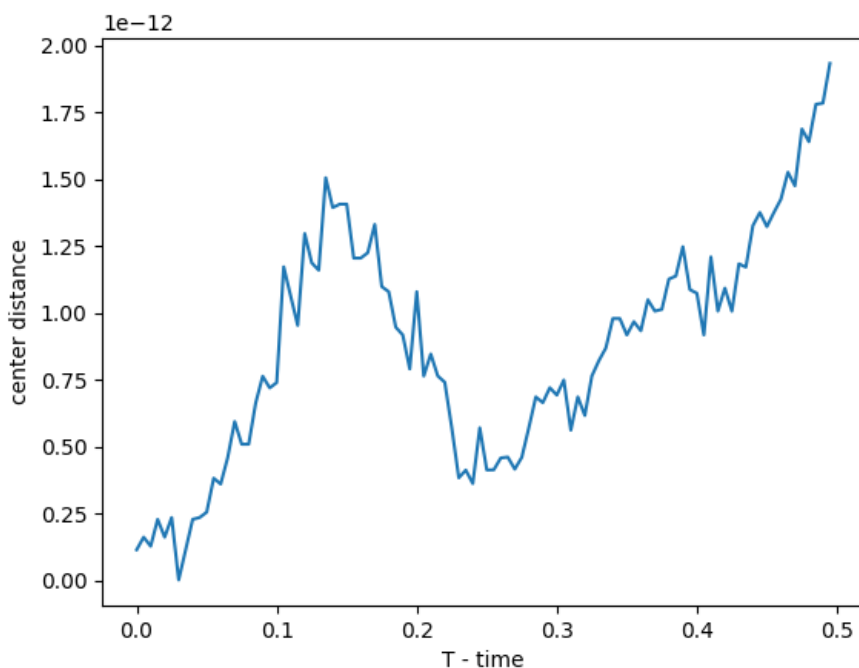
Figure 8: Centre of mass distance from the original position of the centre of mass in C3H8 for 10000 repetitions with a step size of 5*10-5

As the change was small, reaching only a distance of $2*10^{-12}$Angstrom in over 10 thousand steps, it can be mostly attributed to rounding and floating-point errors. In computers, floating point numbers which allow the use of decimals, are represented using a fixed number of binary numbers (Lande, 2014). The integer and decimal parts are split into two binary sequences. The integer part is calculated through the sum of the powers with the base 2 and the powers of the index of the binary numbers with a value of 1, starting at 0. This representation allows the formation of every possible number. The same method holds for the calculation of the decimal component of the number, with negative powers. Unfortunately, not all decimals can be represented as a sum of fractions, using a finite number of binary numbers. Computers use a finite number of binary numbers leading to a small error named floating point error.

4.3. Performance

As the main goal of molecular dynamics is to provide a method of solving molecular systems within a realistic time, the runtime was measured for a range of molecules, for simulations with ten thousand repetitions. Three trials were performed to account for deviation in the computer's computing power. The mean time along with the standard deviation is displayed in figure 9

below. The times were measured for systems with 2 up to 12 atoms. Systems of 7, 9 and 10 were excluded as the main aim was to find a trend of the increase in computational time in relationship to the number of atoms and no sensible molecules exist with these numbers of atoms.
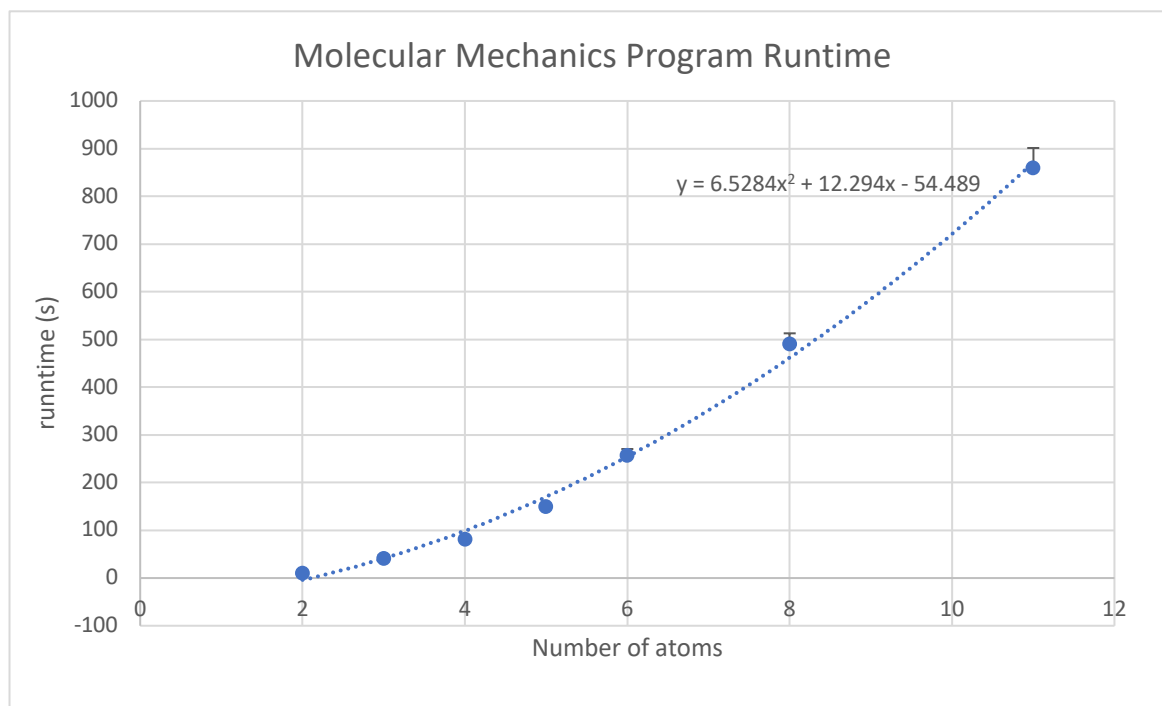

Figure 9: Runtime of 10000 repetitions against number of atoms

As predicted, the runtime increases in a quadratic relationship with the number of atoms, i.e., as $O(n^2)$. The reason for this time complexity is that, due the nonbonded interactions, each pair of atoms has at least one interaction. The Van der Wall force however, as can be seen in the results displayed in figures 5 and 6 c), falls off except for when the atoms come into close contact. This means, that it would be possible for larger molecules to introduce a cut-off range, so that interactions between non-bonded atoms beyond a certain distance that would never come close enough to exhibit a strong interaction, could be ignored. This further approximation would allow to lower the time complexity to a linear one of $O(n)$, as atoms would interact with neighbouring atoms. An alternative solution to a cut-off range often seen in molecular dynamics is a neighbour list, where the sets of atoms that should have their interactions calculated are specified before running the simulation.

4.4. Pi Orbital length – Particle in an infinite Potential Well

Using the methodology described previously for the second experiment, beta carotene was extracted and the absorption spectra for different wavelengths was measured for the wavelengths of 380 to 720 nm in 10 nm intervals, as can be seen in figure 10 below.
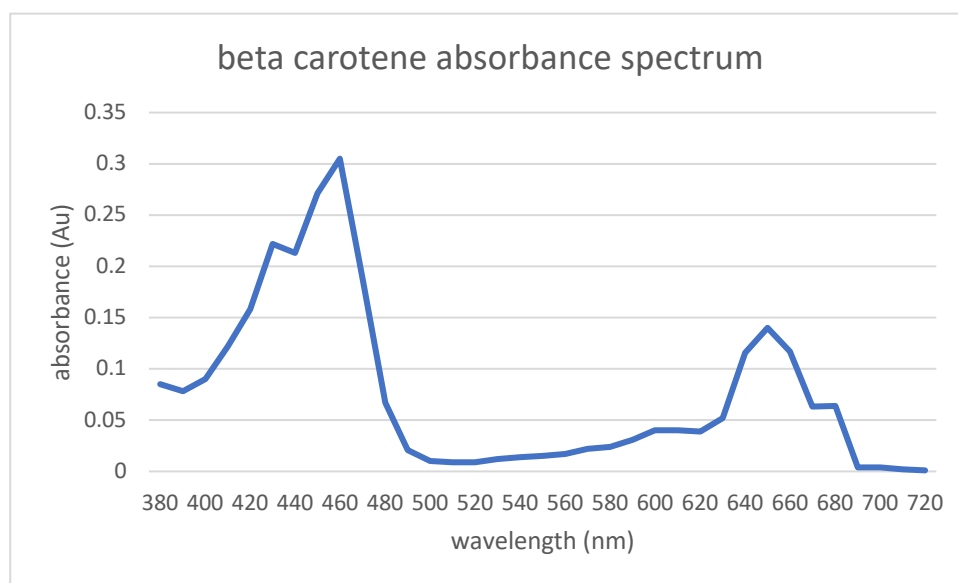
Figure 10: Absorbance spectra as a function of light wavelength for 4 extracted plant pigments.

The wavelength with the highest absorption was measured to be 460 nm. Using the derived formula for the length of the conjugated pi-orbital, the final result is that the conjugated chain in beta carotene has an approximate length of 17.41 Å. The article "A Particle in a Box Laboratory Experiment Using Everyday Compounds" (Wimpfheimer, 2015), which inspired this experiment, yielded a box length of 16.6 Å and another source yielded a box length of 17.79 Å. According to the article, the molecular modeling software spartan yields a length of 23.1 Å, slightly longer than any of the results derived through the particle in a box model. The reason why molecular modeling software is more accurate is that it uses a more exact shape to model the pi-orbital than the particle in a box model.

5. Discussion

The molecular mechanics software functions well for smaller molecules, producing quite accurate shapes, with the exception of the out-of-plane angle, as well as the physical movement of the atoms. Unfortunately, the runtime was too long for larger molecules which was a

fundamental problem stemming from the pairwise interactions in any N-body system. One of the major limitations of the molecular dynamics' software was that it only works for molecules containing only single-bonded hydrogen and carbon, and most biological molecules also contain other elements such as oxygen and nitrogen. Also, the set of parameters for the potential energy equations are only for a single arbitrary temperature. In addition, one of the most important factors determining the shape of a molecule is the environment it is in. This means that to find structures the molecules would realistically exhibit, their interactions with the solution they are in, as well as other molecules they would interact with, needs to be simulated. The issue with this is that in a finite system of molecules, the molecules would just disperse. To prevent this a repeating box model is used, in which copies of the system are placed next to each other. The solution surrounding the molecule can completely alter its shape. For example, proteins fold so that their non-polar structures, hydrophobic parts that repel water, are incased in their polar structures, which are hydrophilic and therefore are attracted to water molecules.

It isn't possible to say that molecular dynamic approaches are better than quantomechanical approaches, or vice versa, as each is suited to different applications. The main application of molecular dynamics is to find the positions of atoms in a molecule with the lowest potential energy. However, by simulating the movement using physical laws other properties such as the vibrational frequency and entropy can be found. Also, molecular dynamics is often used to simulate interactions between different molecules. Molecular mechanics is a good solution for modeling a large number of nuclei, but to find some molecular characteristics, the electronic or orbital shape of the molecule is required. For example, as can was demonstrated with beta carotene, the absorbance spectra of light depend on the orbital functions of the electrons of the molecule. This is why to find characteristics such as photon absorbance, polarity and bonding energy of a molecule quantomechanical methods such as the variational method were discussed in this article to highlight how mixed quantum-classical methods can be applied to overcome the limitations of molecular dynamics. In short, quantum mechanics is best suited for determining electronic wave functions, while molecular dynamics is often better for finding the positions of the nuclei in a molecule.

6. Conclusion

Molecular dynamics is a powerful computational tool that is at the forefront of modern chemistry and physics research. Its ability to accurately predict the three-dimensional

arrangement of atoms in a molecule is the result of its unique approach, which combines classical mechanics and quantum mechanics. By simulating the nuclei as classical particles subject to Newton's laws, the molecular dynamics method provides a detailed picture of the physical interactions between atoms.

The mixed classical-quantum mechanical approach used in molecular dynamics uses the positions of the nuclei predicted by the classical simulation to calculate the orbital spin functions of the electrons in the molecule. This can be achieved by techniques such as the Born-Oppenheimer approximation and the variational method. These techniques provide a deeper understanding of the underlying quantum mechanical processes that govern the behavior of electrons in a molecule.

This project, which explored classical and quantum mechanical approaches to determining organic structures, demonstrates the possibilities and limitations of molecular dynamics. The molecular dynamics program created as part of the project is a testament to the challenging, yet rewarding nature of this field of study.

In summary, molecular dynamics has proven to be a valuable tool for predicting the shape of complex molecules. The combination of classical mechanics and quantum mechanics provides a comprehensive picture of the physical interactions between atoms and allows us to gain a deeper understanding of the world at the molecular level. The development of the molecular dynamics program and its subsequent analysis demonstrates the importance of interdisciplinary approaches to scientific research and the potential for future advances in the field.

Bibliography

Library Alberts, B. (2002). The Shape and Structure of Proteins. *National Libary of medicine*.

Alliance, G., & Health., D. o. (2010). Understanding Genetics: A District of Columbia Guide for Patients and Health Professionals. *National Library of Medicine*.

Born, M. (1928). Beweis des Adiabatensatzes. *Zeitschrift für Physik*.

Case, D. A. (2013). Amber 12 Reference Manual. *ambermd*.

Duin, A. C. (1994). Delft Molecular Mechanics: A New Approach to Hydrocarbon Force Fields. *Journal of the Chemical Society, Faraday Transactions*.

dynamics, M. q.–c. (1998). John C. Tully. *Faraday Discuss*.

Epstein, S. (1974). The variation Method in quantum Chemistry. *University of Wisconsin*.

G. Ciccotti, C. D. (2022). Molecular simulations: past, present, and future (a Topical Issue in EPJB). *The European Physical Journal B volume*.

Gharabaghi, M. (2016). Some implications of the Hartree product treatment of the quantum nuclei in the ab initio nuclear–electronic orbital methodology. *Elsevier*.

Greenshiel, J. (1958). MOLECULAR STRUCTURE AND PROPERTIES OF HYDROCARBONS AND RELATED COMPOUNDS* . *The Journal of Physical Chemistry*.

Gunawardena, J. (2014). Time-scale separation--Michaelis and Menten's old idea, still bearing fruit. *Pubmed*.

Horodecki, R. (2009). Quantum entanglement. *APS*.

J.P.Desclaux. (1969). Hartree Fock Slater self consistent field calculations. *SienceDirect*.

Kaplan, I. G. (2016). Discovery of the Pauli Exlusion Principle and Early Developments. In I. G. Kaplan, *The Pauli Exclusion Principle: Origin, Verifications, and Applications*. Wiley.

Kennedy, U. (2021). PHOTOSYNTHESIS. In *PRACTICAL MANUAL 2021* (pp. 61-65). USQ.

Kozlíková, B. (2015). Visualization of Molecular Structure: The State of the Art. *Eurographics Conference on Visualization (EuroVis)* .

Lande, D. R. (2014). Development of Binary Number Systems and the Foundations of Computer Science. *The Mathematics Enthusiast*.

Liao, S. (2013). Physical limit of prediction for chaotic motion of three-body problem. *Elsevier*.

Maghfiroh, C. Y. (2020). Parameters (sigma, epsilon) of Lennard-Jones for Fe, Ni, Pb for Potential and Cr based on Melting Point Values Using the Molecular Dynamics Method of the Lammps Program. *IOPScience*.

Margenau, H. (1939). Van der Waals Forces. *Reviews of Modern Physics*.

Marshman, E. (2017). Investigating and improving student understanding of the expectation values of observables in quantum mechanics. *European Journal of Physics*.

Matsen, F. A. (1969). Spin-free quantum chemistry. VII. The Slater determinant. *ACS Publications*.

Nagy, B. (2017). Basis Sets in Quantum Chemistry. *Wiley Online Libary*.

Phillips, A. C. (2003). Introduction to Quantum mechanics. *Wiley*.

Prieto-Martinez, F. D. (2019). Computational Drug Design Methods—Current and Future Perspectives. *SienceDirect*.

Shattuck, T. W. (2003). Introduction Section 4 Energy Minimization. In T. W. Shattuck, *Colby College Molecular Mechanics Tutorial QUANTA Version* (pp. 21-26). Waterville, Maine 04901: Colby College.

Sherrill, C. D. (2005). The Born-Oppenheimer Approximation. *School of Chemistry and Biochemistry*.

Simplilearn. (2022). *Introduction to Big O Notation in Data Structure* . Retrieved from Simplilearn: https://www.simplilearn.com/big-o-notation-in-data-structure-article. accsessed on the 1.1.2023

Tomasik, M. (2012). 3. From many-body to single-particle: Quantum modeling of molecules. *MIT - opencourseware*.

Vallance, C. (2021). *Vallance Group*. Retrieved from http://vallance.chem.ox.ac.uk/pdfs/VariationPrincipleNotes.pdf accsessed on the 1.5.2023

Vargesson, N. (2015). Review Thalidomide-Induced Teratogenesis: History and Mechanisms. *Wiley Online Library*.

Varshslovich, D. A. (1988). Chapter 6: SPIN FUNCTIONS. In D. A. Varshslovich, *Quantum Theory of angular momentum* (pp. 170-195). World Sientific.

Vyas, D. (2008). Structures of Life: The Role of Molecular Structures in Scientists' Work. *University of twente*.

Wimpfheimer, T. (2015). A Particle in a Box Laboratory Experiment Using Everyday Compounds. *Scientific & Academic Publishing*.

Zhang, Y. (2009). Protein structure prediction: when is it useful? *Elsevier*.
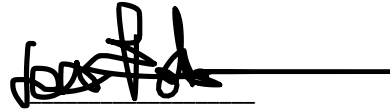
Table of Figures

Selbstständigkeitserklärung

Name: Leo Robinson

Ich erkläre, dass ich diese vorwissenschaftliche Arbeit eigenständig angefertigt und nur die im Literaturverzeichnis angeführten Quellen und Hilfsmittel benutzt habe.

____15.02.2023, Wien____

           Ort, Datum   Unterschrift