University of Nottingham
UK | CHINA | MALAYSIA

Quantum Biology to Enhance CO_2 Fixation

Student name: Jennifer Coughlan

Student ID: 20198790

Supervisors: Jonathan Hirst, Christof Jaeger, Daniela Quaglia

Date: August 2021

Abstract

β-methylmalyl-CoA lyase (MCL) is a bacterial metabolic enzyme with activity in acetyl-CoA assimilation pathways. Bifunctional MCL is found in the ethylmalonyl pathway where it reversibly converts (S)-malyl-CoA into acetyl-CoA and glyoxylate and combines propionyl-CoA with glyoxylate into β-methylmalyl-CoA. Interestingly, MCL has demonstrated activity with (S)-citramalyl-CoA in vitro, a reaction attributed to distantly related citrate lyase. Despite having different tertiary structures and diverged amino acid sequences, the participating residues of the active site of this family of enzymes are well conserved. The non-reversible mechanism of citrate lyase has been characterised both experimentally and computationally, but there is still much to learn about the mechanism of MCL, to address key questions such as how the reversible reaction is able to proceed. Two crystal structures of MCL from *R.sphaeroids* are available. In this work, molecular dynamics (MD) simulations have been performed on one of these structures with a single active site occupied by propionyl-CoA and glyoxylate. No reliable parameters were available for glyoxylate in the CHARMM forcefield, thus, they have been optimised following the protocol set out for CGenFF. Molecular docking has been used to predict the conformations of the bonded substrates, followed by MD simulations to assess the stability of these complexes. To better describe the active site, the enzyme was then modelled using hybrid quantum mechanics/molecular mechanics (QM/MM) methods. This will facilitate the prediction of the intermediate states adopted over the course of the reaction.

Acknowledgements

I would like to thank my supervisor Jonathan Hirst and his research group, particularly Magnus Hansen-Heine and Ellen Guest, for all the help and support they have given me. None of this work would be possible without access to UoN HPC Augusta. Thanks, also, to Pete Licence and Peri Williams and everyone else involved in the CDT for Sustainable Chemistry as well as the EPSRC for funding my work.

Contents

List of Figures

List of Tables

Abbreviations

CETCH : Crotonyl-CoA/Ethylmalonyl-CoA/Hydroxybutyryl-CoA

CHARMM : Chemistry at HARvard Macromolecular Mechanics

CGO : Chemical Gaussian Overlay

ECR : Enoyl-CoA Carboxylase/Reductases

MCL : β-Methylmalyl Lyase

MD : Molecular Dynamics

ML : Machine Learning

MM : Molecular Mechanics

NAMD : Nanoscale Molecular Dynamics

NPT : Isobaric-Isothermal ensemble

NVT : Canonical Ensemble

PDB : Protein Data Bank

PME : Partial Mesh Ewald

QM/MM : Quantum Mechanics Molecular Mechanics

RCD : Redistributed Charge and Dipole method

RMSD : Root Mean Square Deviation

RMSF : Root Mean Square Fluctuations

TIP3P : Transferable Intermolecular Potential 3P

VMD : Visual Molecular Dynamics

1. Introduction

1.1. Carbon Fixation

Carbon fixation is the process by which organisms use atmospheric CO_2 to synthesise the many different carbon-based molecules required for growth. The Calvin Cycle is the most prevalent of the six natural fixation pathways and utilises the most abundant enzyme on Earth: RuBisCO. RuBisCO is well documented to have slow activity and is further sabotaged by the side reaction with O_2 resulting in the release of up to 50% of the previously fixed carbon.[1] While attempts have been made to improve the activity of this carboxylase, none have been particularly fruitful resulting in the speculation that RuBisCO has reached the end of its natural evolutionary progress [2] although recent theoretical research disputes this hypothesis.[3] Some scientists, therefore, have turned to other members of the carboxylase family beyond those found in autotrophic organisms and pathways.

1.2. The CETCH Cycle

Schwander et al. [4] chose enoyl-CoA carboxylase/reductases (ECRs) as their carbon fixator: a recently discovered family of carboxylases that catalyse reactions between CO_2 and CoA derivatives in bacterial metabolisms. These promiscuous enzymes are oxygen-insensitive and require less energy for catalysis than any autotrophic carboxylases. Their work spawned the CETCH cycle (figure 1.1). 17 enzymes from nine different organisms make up the first synthetic carbon fixation cycle exhibiting improved energy efficiency and a faster CO_2 turnover than the six natural predecessors, fixing two CO_2 molecules per turn. In the original paper, three of the enzymes had been engineered to improve reaction rates or modify specificity to better suit the reaction at that point.[4] Major challenges include the amount of ATP and NADPH required by the system, expensive when not generated in situ, and the output molecule, glyoxylate, has not been reported in many natural metabolic pathways suggesting accumulation issues if incorporated into living cells.[1] To solve the energy requirements, the most recent work from their lab has combined a thylakoid membrane with the CETCH cycle to demonstrate how coupling with solar energy can power the cycle cheaply.[5] Scope to convert glyoxylate into useful compounds is already apparent in this new cycle and further enzyme engineering will contribute to the realisation this goal.

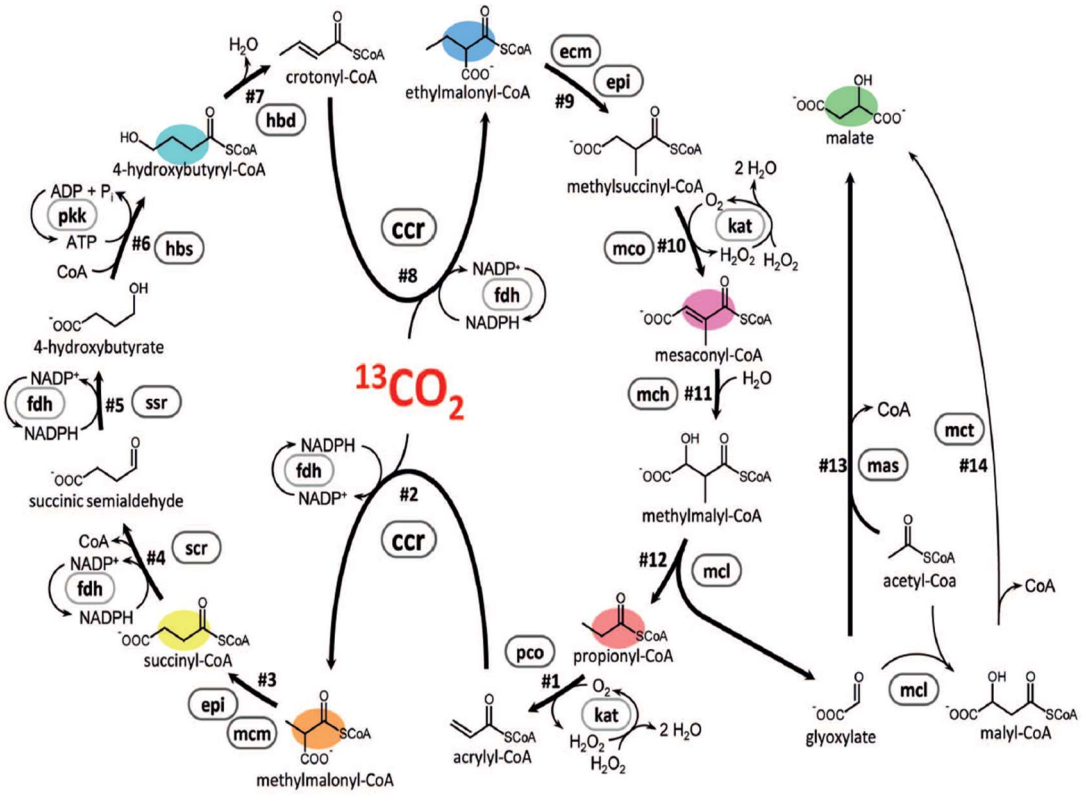

Figure 1.1: The CETCH 5.4 cycle, the first synthetic carbon fixation cycle, taken from Schwander et al.[4]

1.3. Enzyme Engineering

While there are enzymes available for the synthesis or degradation of nearly every natural product, engineering allows a scientist to improve and tweak reactions necessary for the current visions of a fully sustainable bioeconomy.[6] Enzymes offer the ability to conduct highly specific reactions in mild conditions. Engineering can be a delicate process as only a handful of modifications will produce a stable enzyme and only a fraction of these will yield the intended result.[7] If done entirely experimentally this process can be slow and laborious with intensive use of resources and power; computational methods provide likely candidates before embarking on the heavy experimental work. Rational enzyme engineering has already shown promise in recent years and the best results have originated from collaborations between experimental and computational efforts.[7] A systematic workflow to characterise enzymes and glean insights into their mechanisms will help produce educated mutations and would be a novel contribution to the computational side of rational enzyme design. This project aims to develop such a workflow by experimenting with enzymes from the CETCH cycle with a view to contribute to the field of synthetic carbon fixation.

1.4. β-methylmalyl lyase

The enzyme to be tested is β-methylmalyl lyase (MCL). In the CETCH 5.4 cycle, an MCL from *R.sphaeroids* is employed to perform the two reactions it does endogenously as part of the ethylmalonyl-CoA pathway (figure 1.2). Several crystal structures of MCL can be found in the literature from three different species: *M.extorquens* [8], *C.aurantiacus* and *R.sphaeroids* [9]. *M.extorquens* and *R.sphaeroids* are the most closely related with a sequence similarity of 57% but all structures have the same residues conserved within their active sites and the C-terminal hinge. They also all require a bivalent metal ion for functionality: magnesium or manganese. *R.sphaeroids* uses a magnesium ion.

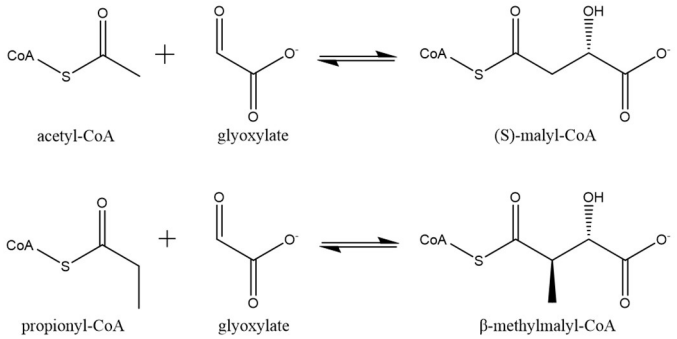

Figure 1.2: The two reversible reactions MCL is able to perform are depicted here.

Versions of the MCL enzyme, malyl lyase, can be found in different organisms performing reactions in acetyl-CoA assimilation pathways. It is a member of the CitE-like superfamily, a β-subunit of an ATP-independent citrate lyase which take acyl-thioesters as substrates and require Mg^{2+} or Mn^{2+} for catalysis. In the early days, MCL1 activity combined with MCL2 activity was mistaken as the work of a single malate synthase. Malate synthases perform similar reactions on CoA thioesters; however, they are non-reversible as the final step of their reaction is the hydrolysis of CoA. While there is low overall amino acid sequence similarity between the malate synthases and the malyl lyases, the active site residues between them are highly conserved.

Structure

There are two crystal structures available for MCL from *R.sphaeroids* are PDB:4L9Y and PDB:4L9Z with resolutions of 2.1 Å and 2.01 Å respectively.[9] 4L9Y was crystallised without substrates and has an incomplete C-termini with three of the six chains lacking the electron density to complete the lid. Of the three that were intact, two were in the closed conformation and these were modelled with glyoxylate and magnesium in the active sites thought to come from the *E.coli* they were expressed in. Mg^{2+} was fitted into the four remaining sites with four water molecules in coordination instead of glyoxylate. The crystals were then soaked in propionyl-CoA leading to 4L9Y with one out of six active sites filled with glyoxylate and propionyl-CoA. 4L9Z is derived from crystals grown in the presence of propionyl-CoA, oxalate and Mg^{2+}. The C-terminal lid domains were completely resolved in the closed conformation with each site occupied with Mg^{2+}, oxalate and free CoA.

Meister et al.[10] found MCL in various *Rhoderbacter* species and isolated it as a band at 35 kDa on an SDS-PAGE, a mass consistent with a single monomer. Using chromatography, they found a protein of around 195 kDa which corresponds to a hexamer. Erb et al.[11] found the same banding pattern for MCL in *R.sphaeroids* specifically. This evidence supports the existence of MCL as a hexamer with six active sites as has been resolved for the crystal structure grown in the physiological environment.[9] The monomer comprises an unordered

N-terminal domain of around 15 residues before turning into the first β-strand of the TIM-barrel. After the sixth strand of the TIM-barrel, there is an insertion loop of around 30 residues. Finally, the C-terminal lid is made up of two α-helices and a β-hairpin.

The hexameric protein is arranged as a dimer of trimers with a central cavity between the two trimers (figure 1.3). The trimeric structure appears to be necessary for catalysis as the C-terminal lid domain stretches over the neighbouring monomer to complete the active site when the substrate is bound. This is termed the closed conformation. Malate synthases A and G function as monomers but other members of the CitE-like superfamily exist as trimers. Tang et al[12] observed three distinct conformations of the C-terminal lid in molecular dynamics (MD) simulations using an MCL from *Roseiflexus xastenholzii.* Conformations were based on the presence of substrate in the active site: closed showed a 30° rotation to cover oxalate/glyoxylate occupied active site in adjacent monomer which returned to open when only Mg^{2+} is present and then half-closed when occupied by Mg^{2+} and Cl^-. The rotation of the lid occurs through Thr264 and Pro265 which, together, act as a hinge and these residues conserved. Closed active sites may influence substrate binding: propionyl-CoA was only found in one of two closed sites of the crystal structure[9], the other containing a glyoxylate molecule. Hersh et al.[13] determined that the glyoxylate must bind the enzyme prior to acetyl-CoA for the reaction to proceed; indeed the crystal structure was soaked in propionyl-CoA after crystallisation with glyoxylate.

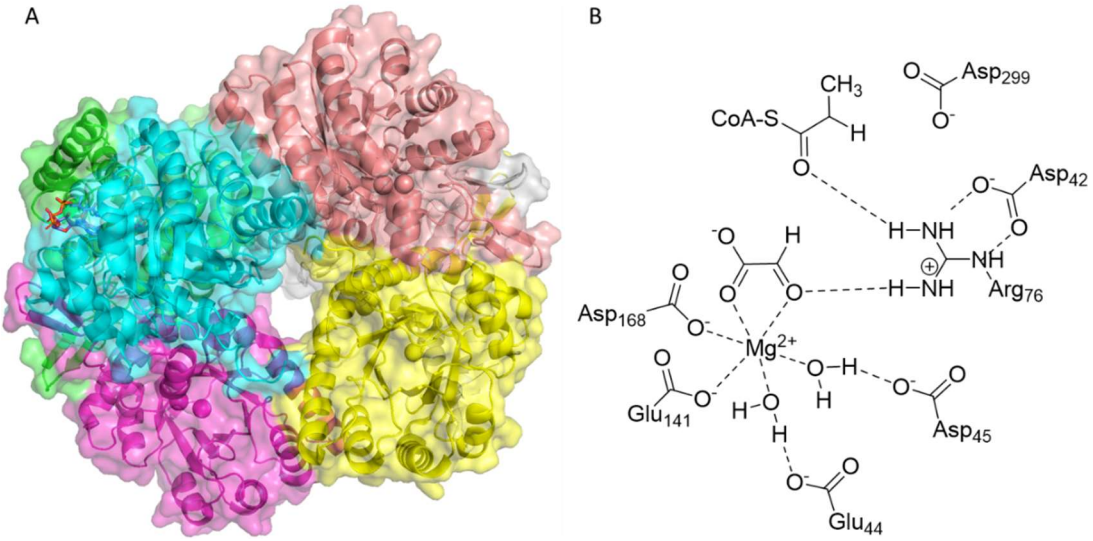

Figure 1.3: Panel A shows the structure of MCL with the phosphate group depicted as sticks where the active site is situated in the cyan monomer and the α-helix of the C-terminal lid is shown in green. The cleft between the trimeric structure can be seen in the centre. Panel B is a diagram of the conserved residues in the active site with propionyl-CoA and glyoxylate present.

The active site is situated at the bottom of tunnel with the CoA adhering to a cleft on the outside of the TIM-barrel. The adenosine moiety is fixed in a largely hydrophobic pocket on the surface of the TIM-barrel with one or two hydrogen bonds between the adenine ring and the carbonyl oxygen atoms of protein backbone. The bent conformation of CoA supports H-bond between the adenosine ring and the hydroxyl of the pantheteine. The phosphate groups coordinated by arginine, lysine or histidine residues and the pantheteine tail is inserted into the narrow tunnel leading to the active site cavity. In the active site itself, Mg^{2+} is octahedrally coordinated by Glu141 and Asp168, two oxygen atoms from glyoxylate and two water molecules (figure 1.3). In an empty active site, the glyoxylate is replaced by two

more water molecules, similar to arrangements seen in malate synthases.[14] The glyoxylate coordinates with the Mg^{2+} and Ala167 and Arg76. Asp299 also sits above the active site, donated by the C-terminal lid from the adjacent chain. Asp299 and Arg76 are conserved across the entire CitE-like superfamily and the malate synthases. Glu141 and Asp168 are also conserved for Mg^{2+} coordination. Other conserved residues include Asp42, which forms hydrogen bonds with Arg76 and is thought to orientate Arg76 into the correct position for reactive interactions with the CoA moieties. Asp45 and Glu44 which are located at the bottom of the active site tunnel where both form hydrogen bonds with the water molecules that coordinate the Mg^{2+}.

Mechanism

Zarzycki and Kerfeld[9] propose a mechanism based on experimental work done for malate synthase from *E.coli.*[15] This group of enzymes perform similar reactions with a CoA moiety and glyoxylate resulting in (S)-malate but this reaction ends with the hydrolysis of the thioester so cleaving the CoA. Quantum mechanics/molecular mechanics (QM/MM) studies of citrate synthase indicate a similar reaction and concludes with a similar mechanism.[16]

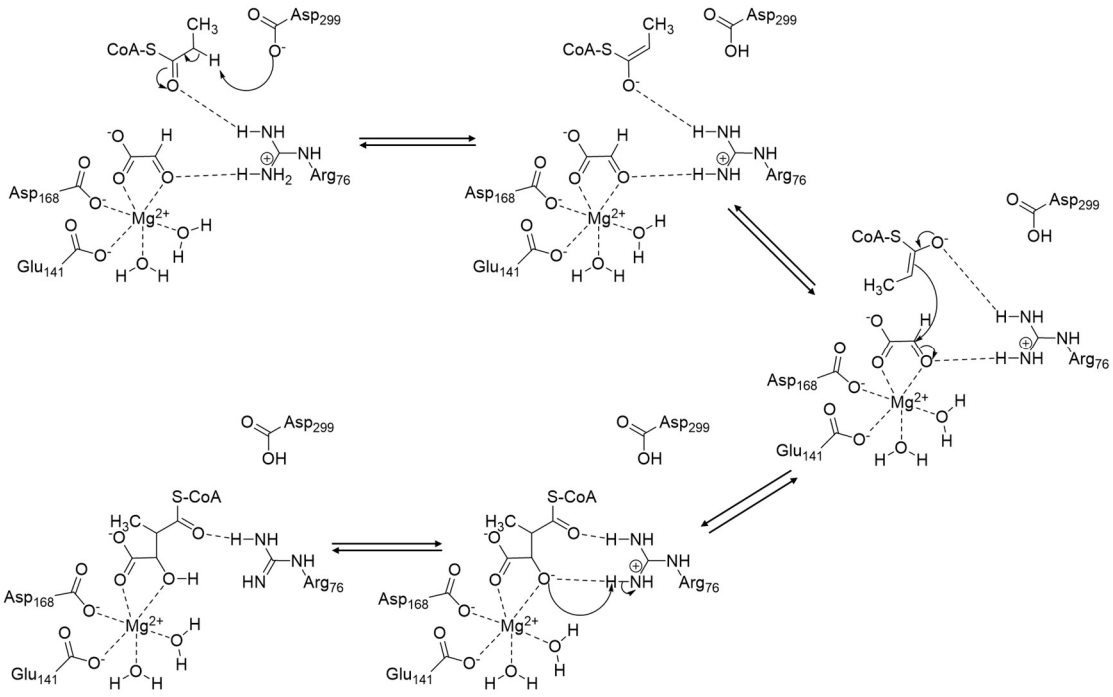

Figure 1.3: The proposed mechanism of MCL with glyoxylate and propionyl as substrates and β-methylmalyl-CoA as a product. Steps one to two is the enolization of the CoA substrate, a rotation occurs to step three where upon the C-C bond is formed between the glyoxylate and the propionyl-CoA as depicted in step for. Finally, a proton is abstracted from Arg76 leading to the formation of β-methylmalyl-CoA. For MCL, all steps are reversible.

For a reaction between propionyl-CoA and glyoxylate to proceed in the context of MCL, the enolization of the propionyl-CoA must occur. It is triggered by the deprotonated Asp299 from the adjacent protein chain, abstracting a hydrogen from the propionyl-CoA leading to the formation of an enol. To stabilise the enol, a hydrogen-bond forms between itself and the correctly orientated and protonated Arg76. Next, the enol rotates so that nucleophilic attack on the carbonyl bond of the glyoxylate is possible. The resulting oxyanion is stabilised by the again by Arg76 but also the Mg^{2+} ion.

The final step is unclear for both malyl lyases and malate synthases. Water leaving the active site has been deemed energetically unfavourable in *E.coli.* So instead there is some

discussion around whether Arg76 could be a proton donor despite an unusually high pKa value.[17] The hydrogen from the hydroxyl group of glyoxylate points away from the Mg^{2+} towards the conserved Arg76 seemingly forming a hydrogen bond and making it a likely candidate. For the reversible reaction to be considered, it has been proposed that Arg76 abstracts the proton from the hydroxyl group when β-methylmalyl-CoA or S-malyl-CoA is in the binding site meaning the arginine residue is required in its deprotonated form.[9] After C-C bond cleavage, the enolate will be neutralized by proton donation by Asp299 from the neighbouring monomer (figure 1.4). Once completed, it can be assumed that the C-terminal lid will revolve away, releasing the product.

The work presented below considers the proposed reaction of MCL from *R.sphaeroids* first through MD simulations and then moving forward to interrogate the mechanism using QM/MM methods.

2. Theory

2.1. Molecular Dynamics

Forcefield

A forcefield assumes a certain chemical bonding pattern providing all the assumed preferred bond lengths and angles so that the formula takes distortions from these preferred parameters into account. A general forcefield is one that applies a generic set of rules to a molecular mechanics (MM) energy expression to a general class of models which includes energy terms and the parameter values to be used in those energy terms (2.1).

$$V_{total} = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} + V_{van\ der\ Waals} + V_{coulomb} \tag{2.1}$$

The full equation for the CHARMM forcefield is described in equation 2.2.

$$V = \sum k_b(b - b_0)^2 + \sum k_\theta(\theta - \theta_0)^2 + \sum k_\varphi(1 + cos(n\varphi - \delta)) + \sum k_\phi(\phi - \phi_0)^2 + \sum k_u(u - u_0)^2 + \sum 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.2}$$

The first parameter for each term represents the force constant for that term: a constant value that tries to describe the potential energy changes for the term as well as possible. x_0 is the reference value for each term and corresponds to the equilibrium structure of that particular construct of atoms.

The first four terms define the bonded interactions (figure 2.1). Bonds and angles are described by Hooke's law which is used as an approximation of the Morse potential. The Morse potential represents bonds and angles as an elastic spring which eventually breaks; Hooke's law does not allow bonds to break and is therefore simpler to calculate.

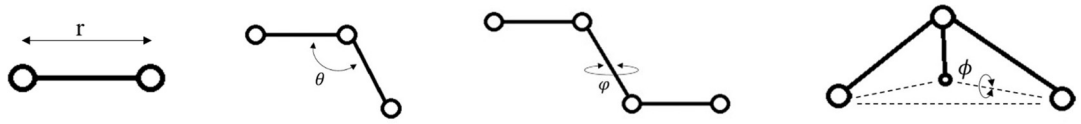

Figure 2.1: The bonded interactions that make up a forcefield where r is the bond length, θ the bond angle, φ is the dihedral angle and ϕ is the improper angle.

The final two terms represent the non-bonded interactions between atoms accounting for weak van der Waals forces and stronger electrostatic forces. These occur between every non-bonded pair of atoms in a system. So, the potentials are truncated, and a switching function is applied to smooth the truncation before it is applied.

$$V_{bond} = \sum k_b(b - b_0)^2$$

(2.3)

All bonds in a molecule are accounted for using equation 2.3, where k_b is the constant to describe the potential energy changes for bond stretching and b_0 is the reference value for stretch which equates to the equilibrium structure. The result is a harmonic potential (figure 2.2).

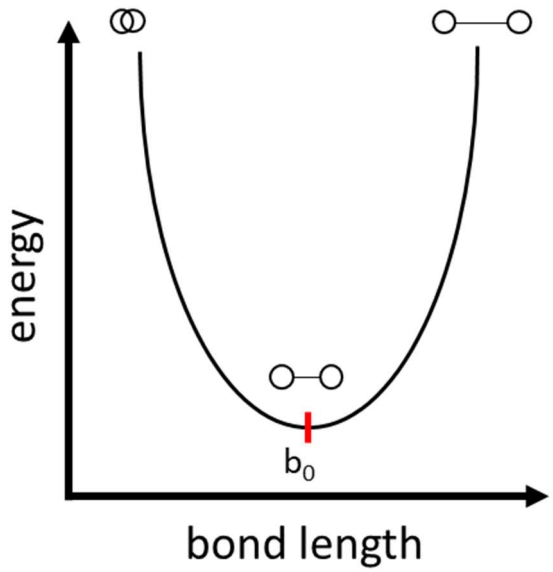

Figure 2.2: A harmonic potential is used to describe both bond and angles in the CHARMM forcefield.

The base of the parabola represents the bond length at equilibrium. In reality, the bond would break as the atoms move apart but, due to the use of Hooke's law, in MM this does not happen and instead the energy goes to infinity. k_b defines the breadth and steepness of the parabola relating to the stiffness of the spring model for bond stretching.

$$V_{angle} = \sum k_\theta(\theta - \theta_0)^2$$

(2.4)

All angles in a molecule are accounted for using equation 2.4, where k_θ is the constant to describe the potential energy changes for bond bending and θ_0 is the reference value for bend which equates to the equilibrium structure. As with bonds, the result is a harmonic potential. Again, the force constant is related to the stiffness of the angle, the larger the value of k_θ, the more energy required to deform the angle from equilibrium.

$$V_{dihedral} = \sum k_\varphi(1 + cos(n\varphi - \delta))$$

(2.5)

The dihedral term describes the torsional movement of the dihedral angle between three covalent bonds (2.5). k_φ describes the stiffness of system about the central bond, n is the multiplicity of the torsional term and δ represents the phase shift. The force constant is related to the amplitude of the curve, if you double k_φ you get the value of the peak. δ controls how far up the axis the potential is shifts do if δ = 0 then the minimum is at 0. *n*

controls the periodicity of the curve, how many stationary points there are around the 360°. If *n* = 1 then there is one minimum, if *n* = 2 then 2 minima and so on (figure 2.3).

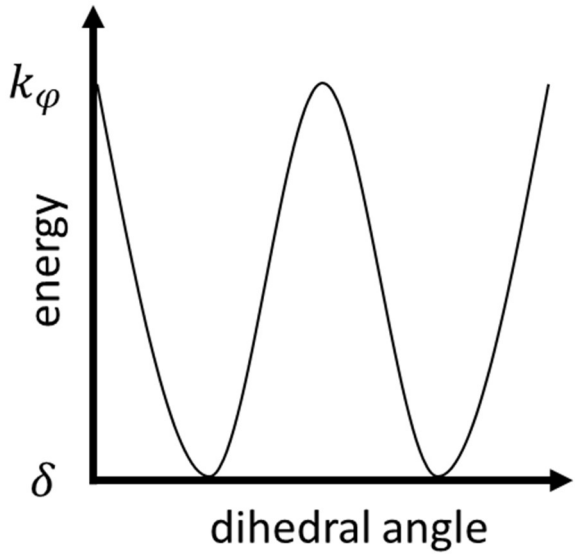

Figure 2.3: A graphical depiction of the dihedral angle model for the CHARMM forcefield.

$$V_{improper} = \sum k_\phi (\phi - \phi_0)^2 \tag{2.6}$$

Improper angles are a special type of dihedral but are treated differently to torsion angles in the CHARMM forcefield. Like bond lengths and angles, they are treated by a harmonic potential where the force constant controls the stiffness of the angle and ϕ_0 is the reference value relating to the equilibrium for those atom types (2.6).

$$V_{van\ der\ Waals} = \sum 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.7}$$

The Lennard-Jones 6-12 potential is used to approximate the weak van der Waals forces between distant atoms. This includes a repulsive force coming from the electron clouds from the two atoms overlapping due to the Pauli Principle. There is also an attractive force between non-bonded atoms due to London dispersion effects. Their addition makes up the Lennard-Jones potential (2.7). ε_{ij} describes the strength of the interaction between atoms *i* and *j* and equates to the well depth where the distance between atoms is optimal. σ_{ij} is the distance between two atoms where the energy is zero; r_{ij} is the distance between atoms *i* and *j*.

$$V_{coulomb} = \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.8}$$

Coulomb's law is used to describe the electrostatics in a system between ions and polar molecules (2.8). *q* represents the partial charge of an atom. ε_0 is the permittivity of vacuum

and r_{ij} is the distance between atoms *i* and *j*. Permittivity is a measure of the electric polarizability of a dielectric. This is an approximation as in reality the partial charges of atoms changes with the environment surrounding them plus a 3D dynamic electron cloud is not properly represented by a single number.

Energy Minimisation

A minimum is a type of stationary point where the first derivative of the potential energy surface is equal to zero and the second derivatives are all positive. Minimisation is the process of finding the conformation with the lowest possible energy on the potential energy surface. This could be a local or global minima. At the minimum point, all the atoms in the molecule are in the lowest energy state where any movement of any atom requires an input of energy. Various algorithms are used to compute this state.

The steepest descent method is one of the simplest for minimization. It moves the coordinates in the opposite direction to the energy gradient. It is generally used for the first few steps to eliminate bad van der Waals contacts as it cannot converge on local minima. The second derivative is assumed to be a constant so the gradients at each point need to be calculated but the second derivative does not. This approximation allows for a quick calculation at each step; however, more steps are required to find the minimum. The direction in which the geometry is first minimized is opposite to the direction where the gradient is largest. Once a minimum is reached in that direction, the algorithm then moves to the next direction and this process is repeated until a minimum has been reached in all directions within a tolerance.

A slightly more complex algorithm is the conjugate gradient method. This method searches for the minimum in multiple directions to eliminate minimization along the same direction. First it minimizes in the direction opposite to where the gradient is largest, in a similar fashion to Steepest Descent. When it reaches the minimum in that direction, instead of simply repeating the process, a little of the previous direction is added to the search thereby reaching the minimum more rapidly.

Classical Molecular Dynamics

In order to simulate the movement of atoms within a given complex, the parameters of a forcefield must be combined with Newton's equations of motion. In this way, interactions between atoms are accounted for by the forcefield while Newton's equations are integrated to propagate movement. NAMD uses the Verlet integration method to achieve this. [18] The force acting on a particle is related to the mass and acceleration of said particle. Further, the acceleration can be determined from the potential energy of the particle. The integration of the equation of motion is used to find the positions of the atoms which can, in turn, be derived to find first the acceleration and second the velocity. Hence, the position and velocity at the current time can be used to find the position and velocity at a new time point. The timestep is limited by the highest frequency of motion within the step and cannot exceed this. To constrain hydrogen bonds, the SHAKE algorithm can be applied thus 2 fs is an appropriate timestep.

In addition, an ensemble is applied to a system acting as a probability distribution of the structures under a given set of macroscopic states such as temperature, pressure, and volume. These simulations use the Canonical (NVT) and Isobaric-Isothermal (NPT) ensembles where NVT maintains the volume, temperature, and number of particles and NPT keeps a

constant pressure, temperature, and number of particles. These ensembles are suitable for modelling enzymatic systems where these attributes are likely to remain constant.

2.2. Docking

Molecular docking of proteins is the process of fitting a ligand into the protein complex. The ligand could be a drug, the natural ligand, DNA or RNA or even another protein. The prediction of the how the ligand orientates itself in the active site is generally based on likely interactions between the enzyme and the ligand. Most programs use a static model of the enzyme however others are capable of combining docking with MD to simulate an induced fit. [19]

Usually, the protein is prepared by giving the program the rough region where the ligand is thought to bind based on experimental evidence or an existing ligand. The ligand is prepared by simulating flexibility in the form of multiple conformers. Various scoring methods can then be used to assess which conformer and orientation fits in the protein best.

OMEGA [20] is the tool used to generate conformers in the OpenEye toolkit by random coordinate embedding and refinement followed by torsion driving resulting in a conformer ensemble. First the ligand is fragmented and matched to a fragment library embedded in OMEGA. The library is derived from commercial compounds whose atoms have been randomly placed in cartesian space and optimized to fill distance bounds and matrix and volume constraints. The MMFF94 forcefield with the non-bonded terms omitted is then used to refine the structures. The fragments of the ligand are compared to this which may result in one or a handful of starting conformers.

A torsion library is also predefined with each torsion definition associated with a list of angles that should be sampled for each torsion. All rotatable bonds in the fragments are compared to the torsion dictionary then all angles are generated leading to a large ensemble of conformers. An MMFF94 scoring system is used to score the conformers and eliminate internal clashes. Further filtering is applied using root mean square deviation (RMSD) to get structurally unique conformers within an energy threshold. [20]

HYBRID and FRED are two of the docking algorithms provided by OpenEye Toolkit. [21] FRED docks a ligand based solely on active site structure and interactions, whereas HYBRID takes the position of an existing ligand into account when docking the compound.

Inputs for both algorithms are the conformer library generated by OMEGA and the protein receptor. First, an exhaustive search is performed where each rigid conformer is translated and rotated around the active site generating a pose library. If a pose clashes with the protein or extends to far from the proposed active site then that pose is rejected. The surviving poses are scored by a scoring function which differs between algorithms: HYBRID uses Chemical Gaussian Overlay (CGO) and FRED uses ChemGauss3. [22] The top 100 poses are then optimized by a local exhaustive search. The same rotation and translation occurs but at half the resolution of the previous step and the best pose is retained for each of the 100 poses. Final scoring is done using ChemGauss4 then the ligands are ranked with the top scoring pose being the one that fits in the active site best according to these parameters. [21]

ChemGauss3 uses Gaussian smoothed potentials to measure how well a pose complements the active site. It does this by looking at the shape, hydrogen-bond interactions with residues in the active site as well as implicit solvent and metal-chelator interactions. CGO takes the existing ligand into account by measuring how well a pose overlays with the original ligand.

This means the scoring takes the overall shape into account as well as any hydrogen-bond interactions and metal chelating groups. ChemGauss4 is an improvement of ChemGauss3 where hydrogen-bond geometry and bond networks recognition are improved.

2.3. Quantum Mechanics/Molecular Mechanics

In a QM/MM simulation all non-bonded and bonded atoms in the QM region are treated by the QM program and the level of theory of choice. All MM atoms are treated by the MM forcefield of choice. The interface, including any covalent bonds, between the MM and the QM can be handled in various different ways and can dramatically affect the outcome.

Non-bonded interactions between QM and MM atoms can be manipulated to suit the purpose. Van der Waals interactions can be calculated with the default parameters but there are options to change the well depth and radii per element type if over-polarization of the QM atoms needs to be compensated for. The electrostatic interactions between the QM and MM atoms can also be modified: mechanical embedding or electrostatic embedding can be used. Mechanical embedding only considers the influence of the van der Waals forces from the MM region on the QM region and vice versa as only the element and atom position of atoms in the QM region is passed to the QM software. For electrostatic embedding, the partial charges of the atoms in the MM region surrounding the QM region are considered so that an approximate electrostatic environment can be applied. A smoothing function can be applied to avoid an abrupt decay of the electrostatic charge. The classical point charge utilization can be further modified so that the MM region can be represented by the total charge, rounded to a whole charge or to a charge complementary to that of the QM region.

To maintain the effect of a bond between the QM and MM regions, link atoms are placed along the bond. This can be any element but is usually a hydrogen atom and usually placed closer to the MM atom meaning the point charge from the MM atom can have profound effects on the link atom. If using the electrostatic embedding scheme, the surrounding partial charges need special treatment.[23] There are 5 widely used methods for embedding implemented in NAMD QM/MM software to handle the point charges surrounding the QM region. None of these methods involve passing the partial charge of the MM atom participating in the QM/MM covalent bond to the QM software. If it were passed on, there would be excessive repulsion or attraction applied to the link atom. This is the “z1” method. “z2” and “z3” are similar but with a wider circle of MM atoms not contributing their partial charges to the set passed to the QM software.

The Redistributed Charge and Dipole (RCD) method avoids obliterating the charges of the nearest MM atoms. It manages the charges by rearranging those of the MM1 and MM2 atoms where MM1 is the atom nearest the link atom and MM2 is bonded directly to MM1. Redistribution of charges means the partial charges assigned to these atoms still contribute to the overall charge. Virtual point charges are generated along the MM1-MM2 bond which are passed to the QM software with the charge coordinates. The result is that no point charge is assigned to the position that the MM1 atom was but is instead redistributed along the bond. The Charge Shifting (CS) method is similar with the MM1 partial charges equally distributed across the MM2 atoms. Virtual partial charges are placed close to the MM2 atoms thus keeping the total charge of the region the same while preserving the local dipoles formed by the MM1-MM2 bonds.

3. Methodology

3.1. Molecular Dynamics Simulations

MCL (PDB: 4L9Y [9]) is made up of six peptide chains where glyoxylate and propionyl-CoA are present in the active site of chain B, a second glyoxylate molecule is located in the active site of chain D and the remaining active sites contain only Cl^- ions. All monomers contain Mg^{2+} coordinated by four water molecules or with two water molecules plus a glyoxylate molecule.

The simulations were run maintaining all crystallographic water molecules and ions. PROPKA was used to assess the protonation states of the amino acids in the active site [24, 25] and to confirm the predicted states based on the postulated mechanism. For the left-hand side of the reaction, Glu141, Asp168 and Asp299 are deprotonated and Arg76 is protonated and for the right-hand side: Glu141, Asp168 and Arg76 are deprotonated and Asp299 is protonated. Neutral histidine is used throughout with the hydrogen placed on the nitrogen atom furthest from the backbone.

Using CHARMM GUI [26], a truncated octahedral periodic boundary cell was fitted with edge distances of 12 Å comprising around 42,000 TIP3P water molecules[27] (1VU: 42000; ACO: 42001; BMM: 41994, MML: 42006) and 43 neutralising potassium ions to give a net charge of zero. Protein and metal parameters are from the C36 CHARMM forcefield. Parameters for Mg^{2+} were developed by Beglov and Roux [28] and are commonly used in biomolecular dynamics. CoA parameters were adapted from Aleksandrov and Field [29] to be compatible with CHARMM36 parameters for proteins and ADP. The CoA structure is made up of phosphorylate ADP, pantothenic acid and a cysteine group. CGenFF [30] was used to generate an initial set of parameters for glyoxylate which were then optimized following the protocol set out by Vanommeslaeghe et al.[30] where required. This is discussed further below.

The NAMD software was used for all MD simulations. [18] The systems were minimised using the conjugate gradient and line search algorithms for 2000 or 3000 steps until the potential energy of the system stabilised. Next, the systems were heated up to 303 K to reflect the conditions of the CETCH cycle. [4] This was done in increments of 5 K per 1 ps. Each system was equilibrated for 100 ps under the NVT ensemble. The final equilibration step involved running the systems under the NPT ensemble until the RMSD levelled off; this usually took around 5 ns. Production dynamics were done using the NPT ensemble for 10 ns with five replicates for each system giving a total of 50 ns of simulation per system. Temperature and pressure controlled using the Langevin dynamics parameters. [31] The cut-off distance for van der Waals pairs was 12 Å and a switching distance of 10 Å was applied. Particle Mesh Ewald method was used for the electrostatics calculations. [32] Energies and trajectories were sampled every 40 ps where each time step was of 2 fs.

3.2. Parameter Optimization

Glyoxylate parameters were optimized following the procedure outlined by Vanommeslaeghe et al.[30] Initial parameters for glyoxylate were determined using the CGenFF web-based tool. Such parameters derived by analogy come with an associated penalty score where a score over 50, as was seen for glyoxylate, is likely to require extensive

optimization. Only the dihedral angle O2-C2-C1-H01 gave a penalty score above 50. So only the parameters for this angle required optimisation.

Target data for torsion angle parameter optimization were generated using a developmental version of QChem software. [33] Complexes were built to mimic ideal hydrogen bonds using TIP3P water molecules [27] and an MP2/6-31G(d) optimized glyoxylate molecule. Interaction distances were optimized using HF/6-31G(d) and interaction energies were calculated from the resulting complexes. These QM data were compared to interaction distances and energies calculated with the CHARMM forcefield then partial charge values were modified from the initial values guessed by CGenFF to reproduce interaction distances to within 0.1 Å and 0.2 kcal mol^{-1} of the target data.

The dihedral angle potential energy surface (PES) was calculated from single point energies at MP2/6-31G* optimization by scanning along the dihedral angle coordinate over a 360° rotation in steps of 5°. Relaxed scans of the dihedral angle were similarly performed in CHARMM with the force constant, K_Φ, the multiplicity, n, and the phase, δ, set to zero. Monte Carlo simulated annealing protocol [34] was not used to minimise the root mean square error between the QM and MM energies. The MacKerell group recommend that, for small molecules, the dihedral parameters should be fitted manually. These parameters were optimized to match the QM target data to the data derived from the CHARMM scans (figure 3.1).

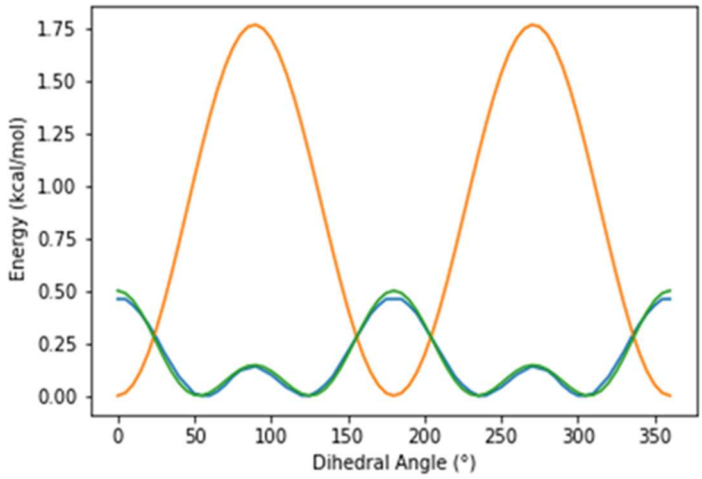

Figure 3.1: Dihedral scan around the O2-C2-C1-H01 angle with the MP2/6-31(d) surface in blue, the CGENFF surface in orange and the optimized CHARMM surface in green.

Partial charges were then verified following the same protocol but using the optimized dihedral parameters and the MM minimised structures rather than the QM target data. The resultant parameters were used for the MD simulations of MCL where necessary.

3.3. Docking

To model the enzyme with acetyl-CoA in the active site, the propionyl-CoA in the original structure was simply cropped by a methyl group. (S)-malyl-CoA and β-methylmalyl-CoA were docked into the active site using the Openeye docking toolkit. The X-ray crystal structure was used for docking and the propionyl-CoA and glyoxylate present were designated as ligands. Their coordinates were used to define a box of 23.3 Å x 31.0 Å x 18.7 Å around the binding site with a total receptor volume of 8027 $Å^3$. A single constraint was applied ensuring a metal chelate interaction between Mg^{2+} and the (S)-malyl-CoA or β-methylmalyl-CoA.

The ligands were protonated giving a net charge of -5 arising predominantly from the phosphate groups. They were then prepared using OMEGA [20], generating a maximum of 10000 conformers with an energy window of 10 kcal mol^{-1}. Furthermore, any conformers within 0.5 Å RMSD were discarded as duplicates. Conformers were generated using MMFF94s forcefield. Docking was performed using OpenEye HYBRID [21] where scoring of conformers is done using the Chemical Gaussian Overlay function which assesses how well a conformer matches the chemistry and shape of the original ligand. The top 10 scoring poses were inspected and the conformation with the most similar orientation to the template ligands was used for MD simulations. Once simulations had been analysed, it became clear that the complex containing β-methylmalyl-CoA was not stable and did not maintain the expected interactions after equilibration. Instead, the complex bound with (S)-malyl-CoA was modified by the addition of a methyl group to create the β-methylmalyl-CoA complex. The equilibrated structure maintained all expected interactions with residues from the active site.

3.4. Quantum Mechanics/Molecular Mechanics Simulations

Protein structures were derived from the MD simulations of all four substrate sets from PDB:4L9Y. Atomic positions from the final equilibration frame were used as the initial coordinates. For each complex, two simulations were run including a minimal set of atoms and a more extensive set of atoms in the QM region. The smaller atom set contains the portion of the CoA moiety after the cysteine group, glyoxylate, Mg^{2+}, coordinated water molecules and the residues implicated in the predicted mechanism: Arg76 and Asp299 (figure 1.4). The larger group of atoms also includes residues conserved across this enzyme family: Asp42, Glu141 and Asp168 (figure 3.2). Both regions have an overall charge of -1. The numbers of atoms in each complex changes where complexes including propionyl-CoA and β-methylmalyl-CoA contained either 72 or 138 atoms in the smaller and larger regions respectively. For acetyl-CoA and (S)-malyl-CoA containing complexes, the number of atoms was 69 or 135.

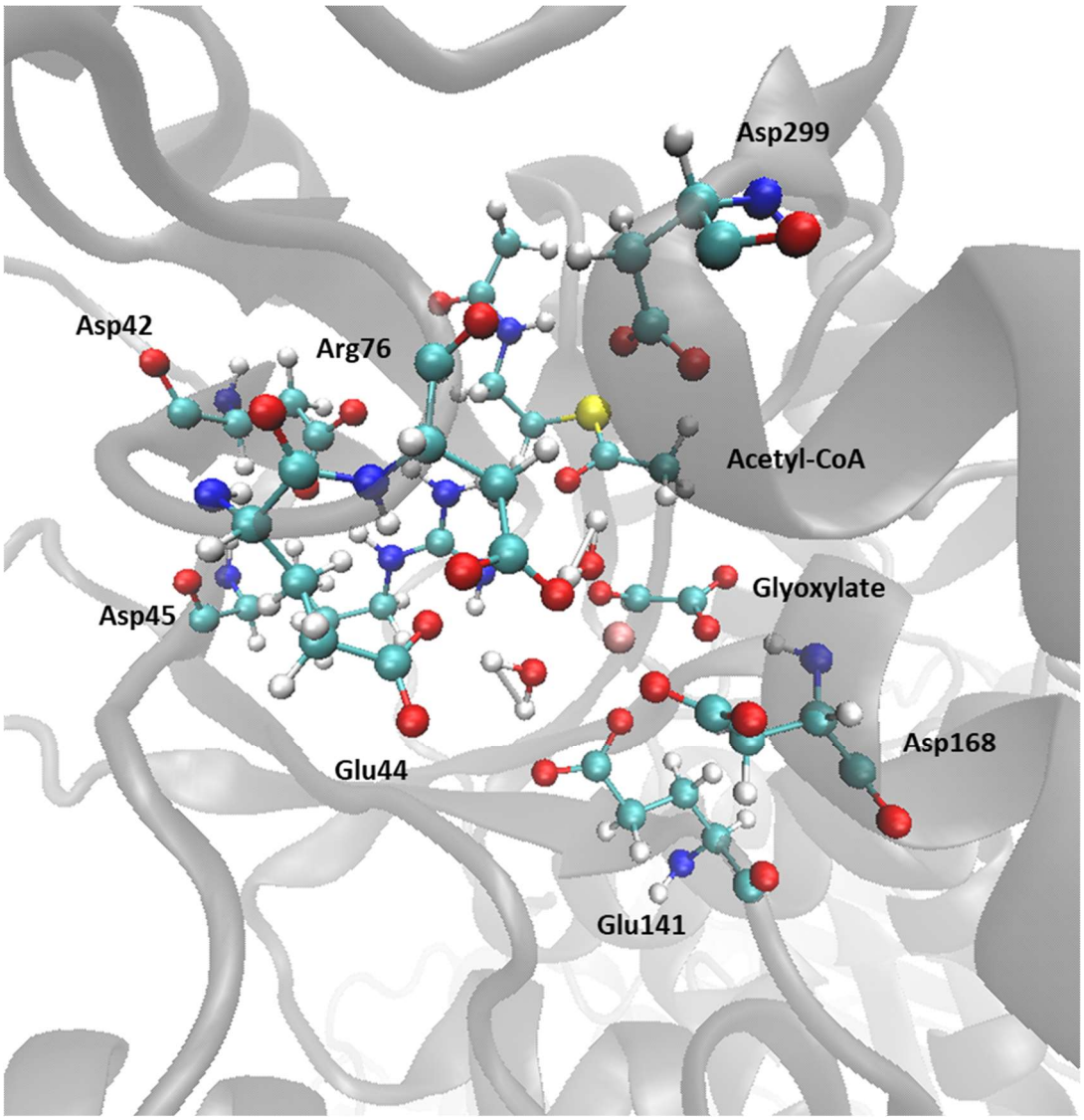

Figure 3.2: The enzyme is depicted using a ribbon representation while the atomic model represents those atoms included in the QM region of the QM/MM model.

All QM/MM simulations were conducted using NAMD. The MM system was described using the CHARMM forcefield, afore mentioned optimized parameters for glyoxylate and parameters from Aleksandrov and Field [29] for the CoA substrates. The NPT ensemble was used to maintain the temperature at 303 K and the pressure at 1.0 bar. The QM region was modelled using MOPAC2016 [35] using the semiempirical method AM1 to describe these atoms. More accurate methods can be used but in the interest of computational time, AM1 was deemed acceptable. To model the interface between the QM and MM regions, the electrostatic embedding procedure was used using a shift function to smooth the electrostatic forces between the QM and MM regions. The total charge from the surrounding point charges was rounded to the nearest whole charge. Hydrogen link atoms were implemented applying the Redistributed Charge and Dipole (RCD) method. Five replicate simulations were performed with time steps of 0.5 fs to a total of 50 ps per simulation.

Visualisation of the trajectories was done using VMD. [36] All distances and RMSD calculations were computed using the python package MDTraj. [37]

4. Results & Discussion

4.1. Molecular Dynamics Simulations

MD simulations were done on the four complexes of the MCL crystal structure to investigate both the reversible reactions it is able to perform. Complexes were set up containing acetyl-CoA with glyoxylate, (S)-malyl-CoA, propionyl-CoA with glyoxylate and β-methylmalyl-CoA in a single active site of the six available to the enzyme.

The crystal structure, PDB:4L9Y, was chosen because the structure was refined with propionyl-CoA and glyoxylate in one of the active sites. The coordination sphere of the Mg^{2+} was observed to include two residues from the protein: Asp168 and Glu141, two water molecules and two oxygen atoms from the glyoxylate. [38] The glyoxylate, in turn, forms hydrogen bonds with Ala167 and protonated Arg76. The relevant portion of the propionyl-CoA is largely free to move within the active site but exhibits some transient interaction between its carbonyl group and Arg76. This fits with the initial structure described by the proposed mechanism.

The terminal methyl group was clipped from propionyl-CoA within the crystal structure to produce a complex containing acetyl-CoA and glyoxylate thus providing the starting structure for the second reaction of MCL. Complexes containing the products were generated using docking techniques as described in the methods with structures chosen based on the interactions seen in the active site thus the starting and equilibrated structures containing (S)-malyl-CoA and β-methylmalyl-CoA interact with the Mg^{2+} by two oxygen atoms reflecting the glyoxylate molecule orientation in the original crystal and the hydroxyl groups are stable to within 2 Å of the deprotonated Arg76. [39]

Propionyl-CoA

Within the active site, three water molecules maintain their position throughout the simulations. First, are the two water molecules (510, 724) that coordinate with the Mg^{2+} and secondly, a single molecule (572) which interacts with Asp299 from chain A, the neighbouring monomer, and the conserved Asp45 residue. The coordination sphere around the Mg^{2+} is maintained throughout all simulations (table 4.1). It is made up of two water molecules, two oxygen atoms from glyoxylate and two oxygen atoms from Glu141 and Asp168 from chain B of the protein.

Other distances look at those of potential importance for the mechanism. None of these distances become short enough to be described as an interaction. However, the glyoxylate is fairly close to Arg76 and the propionyl-CoA ligand also moves towards the same residue and the glyoxylate as would be expected based on the mechanism. Nevertheless, the propionyl-CoA is too far away from Asp299 for a proton to be abstracted and for the reaction to be initiated. The propionyl portion of the propionyl-CoA is free and flexible and does not interact with any of the residues in the active site tunnel for long.

Table 4.1: The average distance and standard deviation was calculated for all 50 ns of simulation of the MCL complexed with propionyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and 1VU is propionyl-CoA.

Interaction	Average distance (Å)
Mg – TIP510	2.02 ± 0.07
Mg – TIP724	1.99 ± 0.07
Mg -GLV (O1)	2.11 ± 0.11
Mg – GLV (O2)	1.98 ± 0.08
Mg – Glu141 (OE1)	1.82 ± 0.04
Mg – Asp168 (OD2)	1.82 ± 0.04
GLV (O1) – Arg76 (NH1)	3.49 ± 0.24
GLV (O1) – Arg76 (NH2)	4.83 ± 0.25
1VU (O) – Arg76 (NH1)	6.17 ± 1.17
1VU (O) – Arg76 (NH2)	5.24 ± 1.03
1VU (O) – Asp299 (OD1)	6.95 ± 0.84
1VU (O) – Asp299 (OD2)	5.74 ± 0.78
1VU (C1) – GLV (C1)	4.71 ± 0.75

β-methylmalyl-CoA

To get a reasonable starting structure, the complex containing (S)-malyl-CoA was modified by adding a methyl group so the substrate. Coordination around the magnesium was maintained for all interactions except the hydroxyl oxygen of the β-methylmalyl-CoA (table 4.2).

Table 4.2: The average distance and standard deviation was calculated for all 50 ns of simulation of the MCL complexed with β-methylmalyl-CoA. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, and BMM is β-methylmalyl-CoA.

Interaction	Average distance (Å)
Mg – TIP510	2.10 ± 0.12
Mg – TIP724	2.03 ± 0.10
Mg – BMM (O41)	1.82 ± 0.04

Mg - BMM (O49)	2.83 ± 1.21
Mg – Glu141 (OE1)	1.84 ± 0.05
Mg – Asp168 (OD2)	1.85 ± 0.05
TIP510 – GLU44 (OE2)	4.14 ± 2.13
TIP724 - GLU44 (OE1)	3.72 ± 1.85
BMM C25 – ASP299 (OD2)	4.78 ± 0.45
BMM O42 – ARG76 (NH2)	5.13 ± 0.68
BMM O49 – ARG76 (NH1)	4.38 ± 1.13

Out of five simulations, three demonstrated that the hydroxyl group on the β-methylmalyl-CoA moves away from the Mg^{2+} while the water molecules rearranged themselves. At the same time, Glu44 also rotates away from the active site seen at 4, 6 and 9 ns (figure 4.1). Also at these times, the β-methylmalyl-CoA is observed to move closer to Asp299 donated by chain A and further from Arg76. This could either suggest movement to begin the reverse reaction or the beginning of the release mechanism.

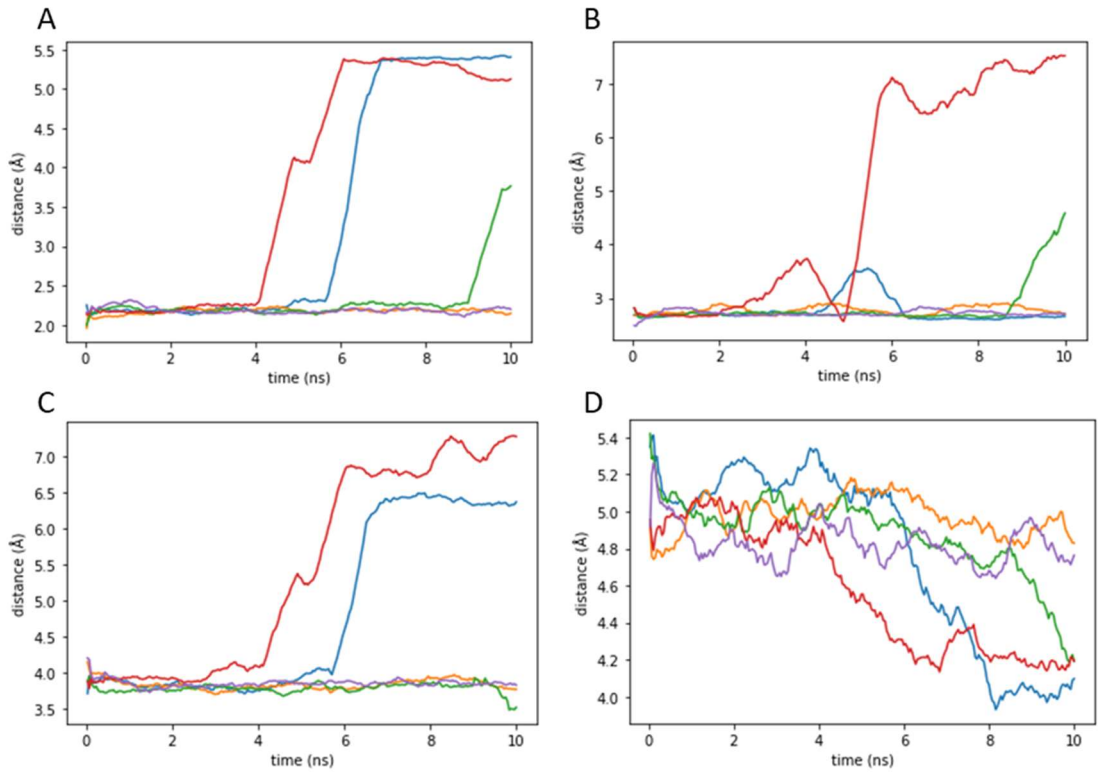

Figure 4.1: The distance between atoms as the moving average with a window size of 40 ps for all 5 replicate simulations. Panel A: is the distance between the Mg2+ and the hydroxyl oxygen of β-methylmalyl-CoA; panel B: is the distance between a coordinating water molecule and Glu44; panel C: β-methylmalyl-CoA hydroxyl oxygen and Arg76; panel D: the carboxyl carbon of the β-methylmalyl-CoA and Asp299.

Acetyl-CoA

Within the active site four water molecules are found to remain for the majority of the simulation. First, the two water molecules (510, 724) that coordinate with the Mg^{2+}, then a third molecule (641) that maintains its interactions with Asp299 and Arg76. Finally, a fourth molecule (572) interacts with Asp45 and Asp168 but soon escapes in all but one of the simulations. The coordination sphere surrounding Mg^{2+} is maintained throughout the simulations (table 4.3). As expected, the system with the acetyl-CoA ligand behaves similarly to the system with propionyl-CoA.

Table 4.3: The average distance and standard deviation was calculated for all 50 ns of simulation of the MCL complexed with acetyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and ACO is acetyl-CoA.

Interaction	Average distance (Å)
Mg – TIP510	1.98 ± 0.06
Mg – TIP724	1.98 ± 0.06
Mg -GLV (O1)	2.09 ± 0.10
Mg – GLV (O2)	1.99 ± 0.08
Mg – Glu141 (OE1)	1.82 ± 0.04
Mg – Asp168 (OD2)	1.83 ± 0.04
GLV (O1) – Arg76 (NH1)	3.33 ± 0.28
GLV (O1) – Arg76 (NH2)	5.19 ± 0.33
ACO (O) – Arg76 (NH1)	5.58 ± 0.96
ACO (O) – Arg76 (NH2)	5.62 ± 0.79
ACO (O) – Asp299 (OD1)	5.62 ± 0.66
ACO (O) – Asp299 (OD2)	4.30 ± 0.53
ACO (C1) – GLV (C1)	6.16 ± 0.93

(S)-Malyl-CoA

Out of the top 10 poses derived from the docking, the third top scoring conformation of (S)-malyl-CoA appeared to be closest to the expected coordination based on the complex containing glyoxylate and acetyl-CoA. In four of the simulations, distances between atoms remain relatively static however for the final simulation changes can be seen at around 6 ns (figure 4.2).

Table 4.4: The average distance and standard deviation was calculated for all 50 ns of simulation of the MCL complexed with (S)-malyl-CoA. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule and MML is (S)-malyl-CoA.

Interaction	Average distance (Å)
Mg – TIP510	2.11 ± 0.12
Mg – TIP724	2.01 ± 0.08
Mg – MML (O33)	1.83 ± 0.05
Mg - MML (O48)	2.33 ± 0.44
Mg – Glu141 (OE1)	1.84 ± 0.05
Mg – Asp168 (OD2)	1.85 ± 0.05
TIP510 – GLU44 (OE1)	4.44 ± 1.10
TIP724 - GLU44 (OE1)	3.21 ± 1.58
MML C17 – ASP299 (OD2)	4.50 ± 0.46
MML O48 – ARG76 (NH1)	2.95 ± 0.36

For all but one of the simulations, the coordination sphere around the Mg^{2+} is maintained (table 4.4). During this simulation, the hydroxyl group on the (S)-malyl-CoA loses its interaction with the Mg^{2+} and moves away from Arg76 and the water molecules rearrange as their interaction with Glu44 is broken at around 6 ns (figure 4.2). At the same time, the (S)-malyl-CoA carbon atom thought to be deprotonated by Asp299 moves towards this residue and away from Arg76.

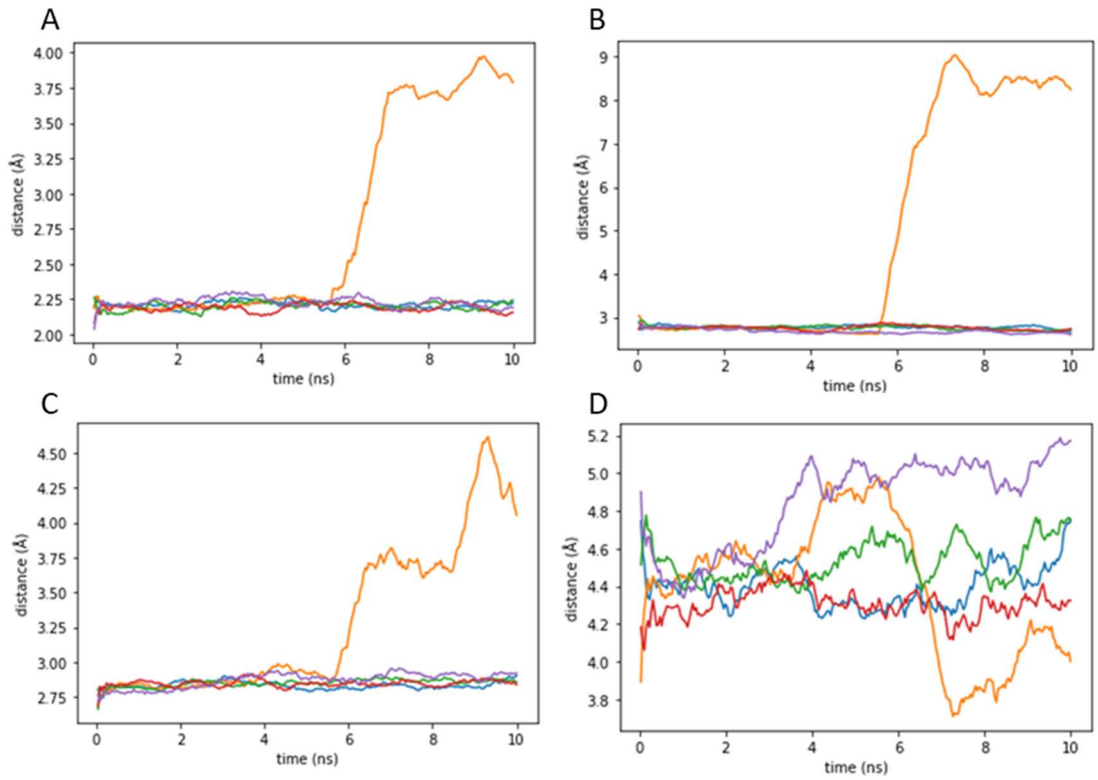

Figure 4.2: The distance between atoms are plotted as the moving average with a window size of 40 ps for all 5 replicate simulations. A: is the distance between the Mg^{2+} and the hydroxyl oxygen of (S)-malyl-CoA; B: is the distance between a coordinating water molecule and Glu44; C: (S)-malyl-CoA hydroxyl oxygen and Arg76; D: the carboxyl carbon of the (S)-malyl-CoA and Asp299.

C-terminal Lids

This crystal structure of MCL contains three near complete C-terminal domains which are known to act as a lid for the active site required for completion as it donates the aspartate residue necessary to trigger the reaction.[12] RMSD plots of these domains demonstrate this flexibility as the RMSD is higher in the C-terminal domain that is not associated with an active site containing substrates. The C-terminal domain of chain A covers the filled active site, chain C is associated with an empty active site and chain F links with a partial filled active site where only glyoxylate is present (figure 4.3). Chloride ions are present in the empty active sites so it is possible that more movement would be seen from chain C if they were omitted from the structure as these are known to induce the half closed conformation.[12]

This observation is backed up by root mean squared fluctuation (RMSF) plots where the most movement per amino acid corresponds to the C-terminal domains (figure 4.3). The spikes in the RMSF plot correspond to those amino acids at the end of each chain.

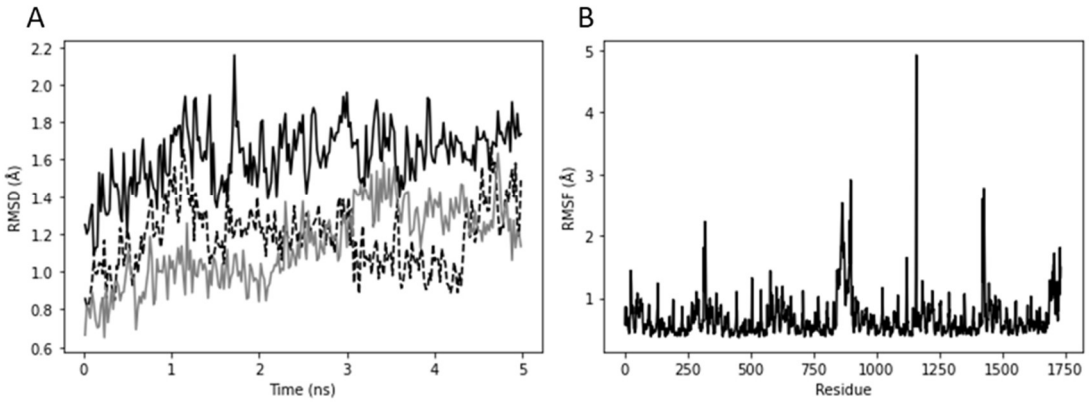

Figure 4.3: Panel A: RMSD of the C-terminal domains in the crystal structure. The dashed line represents chain A covering the active site containing propionyl-CoA and glyoxylate. The black line is the C-terminal from chain C is associated with the empty active site. The grey line is the RMSD of chain F covering the active site containing only glyoxylate. Panel B: RMSF per residue where the spikes correspond to the C-terminal domain amino acids.

Clustering

For each complex, all five trajectories were combined resulting in 50 ns of simulation per complex for each cluster analysis equating to 1250 frames. The analysis was performed by including every fifth frame in the calculation yielding approximately 20 clusters per simulation. For all complexes, the cluster with the most frames included was a structure similar to or the same as the initial coordinates given to the production dynamics after equilibration. As such, the fully equilibrated structure was used as the initial structure for the QM/MM simulations.

4.2. QM/MM Simulations

Modelling the atoms within the active site with a semi-empirical method will give the coordinates of atoms a more accurate level of theory while maintaining a reasonable calculation time. Similar characterisation of the active site was seen in both the larger and smaller QM regions modelled. No unusual behaviour was noted in any of the simulations suggesting that AM1 is a suitable model for the active site of MCL. Alternatively, longer simulation times may result in similar behaviour to some of the MD simulations. All distances quoted are from the models containing the larger number of atoms included in the QM regions however they are reflected in the models with the smaller QM regions.

Propionyl-CoA

Using the AM1 semi-empirical method, the magnesium coordination sphere was maintained between the Mg^{2+} and the oxygen atoms donated by the glyoxylate substrate, water molecules in the active site and Alu141 and Asp168 (table 4.5). The overall position of the propionyl-CoA does not vary. However, the simulation only lasts 250 ps overall, which is significantly shorter (by necessity) than the MD simulations.

Table 4.5: The average distance and standard deviation was calculated for all 250 ps of simulation of the MCL complexed with propionyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and 1VU is propionyl-CoA. All distances measured are within the QM region of the model.

Interaction	Average distance (Å)
Mg – TIP510	2.07 ± 0.07
Mg – TIP724	2.05 ± 0.07
Mg -GLV (O1)	2.21 ± 0.09
Mg – GLV (O2)	2.04 ± 0.07
Mg – Glu141 (OE1)	2.04 ± 0.09
Mg – Asp168 (OD2)	1.93 ± 0.06
GLV (O1) – Arg76 (NH1)	3.45 ± 0.26
GLV (O1) – Arg76 (NH2)	4.84 ± 0.33
1VU (O) – Arg76 (NH1)	6.68 ± 0.41
1VU (O) – Arg76 (NH2)	5.03 ± 0.42
1VU (O) – Asp299 (OD1)	6.09 ± 0.51
1VU (O) – Asp299 (OD2)	4.58 ± 0.52
1VU (C1) – GLV (C1)	5.05 ± 0.55

β-methylmalyl-CoA

In the enzyme complex containing β-methylmalyl-CoA, the magnesium coordination sphere is maintained as expected. Unlike in the MD simulation, none of the production runs display any disruption of the forces between the metal ion and the substrate (table 4.6). This could either be a lack of sufficient sampling or the QM methods model the region in such a way that the magnesium is more likely to maintain interactions with β-methylmalyl-CoA.

Table 4.6: The average distance and standard deviation was calculated for all 250 ps of simulation of the MCL complexed with β-methylmalyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and BMM is β-methylmalyl-CoA. All distances measured are within the QM region of the model.

Interaction	Average distance (Å)
Mg – TIP510	2.06 ± 0.07
Mg – TIP724	2.12 ± 0.08
Mg – BMM (O41)	2.01 ± 0.06

Mg - BMM (O49)	2.16 ± 0.09
Mg – Glu141 (OE1)	2.01 ± 0.10
Mg – Asp168 (OD2)	1.94 ± 0.06
TIP510 – GLU44 (OE2)	3.76 ± 0.34
TIP724 - GLU44 (OE1)	3.09 ± 0.48
BMM C25 – ASP299 (OD2)	5.02 ± 0.44
BMM O42 – ARG76 (NH2)	5.61 ± 0.55
BMM O49 – ARG76 (NH1)	3.14 ± 0.20

Acetyl-CoA

The magnesium coordination sphere is maintained throughout the simulation, reflecting the results of the MD simulation. As expected, there is some flexibility in the acetyl-CoA substrate, while it is an unexciting distance from the reactive residues the standard deviation is also relatively high (table 4.7). The mobile nature of the substrate is possibly necessary for the mechanism to proceed as expected.

Table 4.7: The average distance and standard deviation was calculated for all 250 ps of simulation of the MCL complexed with acetyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and ACO is acetyl-CoA. All distances measured are within the QM region of the model.

Interaction	Average distance (Å)
Mg – TIP510	2.09 ± 0.12
Mg – TIP724	2.11 ± 0.08
Mg -GLV (O1)	2.66 ± 0.85
Mg – GLV (O2)	2.05 ± 0.07
Mg – Glu141 (OE1)	2.37 ± 0.09
Mg – Asp168 (OD2)	1.94 ± 0.06
GLV (O1) – Arg76 (NH1)	3.85 ± 0.85
GLV (O1) – Arg76 (NH2)	5.52 ± 0.72
ACO (O) – Arg76 (NH1)	6.42 ± 0.83
ACO (O) – Arg76 (NH2)	6.11 ± 0.90

ACO (O) – Asp299 (OD1)	4.36 ± 0.60
ACO (O) – Asp299 (OD2)	4.09 ± 0.57
ACO (C1) – GLV (C1)	6.80 ± 0.88

(S)-malyl-CoA

The complex containing (S)-malyl-CoA in the active site of MCL exhibits similar behaviour to the other simulations that have been described. The coordination sphere is kept intact throughout the entire 250 ps of run time and the water molecules involved maintain their interactions with the residues around the active site (table 4.8). There is little flexibility in the moiety but the reactive residues are positioned in the way expected from the proposed mechanism.

Table 4.8: The average distance and standard deviation was calculated for all 250 ps of simulation of the MCL complexed with (S)-malyl-CoA and glyoxylate. Mg represents the magnesium ion in the active site, TIP is the oxygen atom of a TIP3P water molecule, GLV is glyoxylate and MML is (S)-malyl-CoA. All distances measured are within the QM region of the model.

Interaction	Average distance (Å)
Mg – TIP510	2.08 ± 0.07
Mg – TIP724	2.08 ± 0.07
Mg – MML (O33)	2.00 ± 0.06
Mg - MML (O48)	2.17 ± 0.09
Mg – Glu141 (OE1)	1.96 ± 0.06
Mg – Asp168 (OD2)	1.95 ± 0.06
TIP510 – GLU44 (OE1)	2.81 ± 0.19
TIP724 - GLU44 (OE1)	2.97 ± 0.24
MML C17 – ASP299 (OD2)	5.60 ± 0.49
MML O48 – ARG76 (NH1)	3.39 ± 0.25

The QM/MM results have demonstrated that using the CHARMM forcefield with NAMD to model the MM portion of the structure and AM1 to model the QM region of the model gives and acceptable simulation that can be calculated on a reasonable timescale. The stability of the structures provides a good starting point with which to perform experiments to investigate the proposed mechanism and the likelihood that the expected transition states are correct.

5. Conclusions

5.1. Summary

β-methylmalyl lyase is an interesting enzyme to study with implications in the CETCH cycle for synthetic carbon fixation and in its own endogenous pathways as it has important roles in bacterial metabolism in both the 3-hydrocypropionate cycle [40] and the ethylmalonyl-CoA pathway. [41] Mechanisms for family members have been studied and the primary steps well characterised. However, characterisation studies of the reversible reactions performed by malyl lyases while using the same active site residues as malate and citrate synthases which hydrolyse the CoA from the product irreversibly will result in an interesting investigation. The MD simulations performed in this work demonstrate that the enzyme is able to remain as a stable complex with all substrates. Subsequent QM/MM models and simulations along the reaction coordinates may yet elucidate the details of this mechanism and may also help to explain where the differences lie between the enzymes in this family.

5.2. Future Work

The immediate next steps of this project are to continue with the QM/MM studies with some further, directed simulations. The intention is to run umbrella sampling along the reaction coordinates of each step of the mechanism to calculate the free energy potential. Reaction coordinates have been derived from the predicted mechanism based on malate synthase [14, 15] and previous QM/MM studies on citrate synthase. [42] Three coordinates have been defined:

$r_1 = d(C_{\text{propionyl-CoA}}H) - d(O_{\text{Asp299}}H)$
$r_2 = d(C_{\text{propionyl-CoA}}C_{\text{glyoxylate}})$
$r_3 = d(N_{\text{Arg76}}H) - d(O_{\text{glyoxylate}}H)$

r_1 describes the enolization step, r_2 the condensation step of the mechanism and r_3 the final proton transfer to the substrate. These will be simulated in both the forward and backward directions to simulate the reversible properties of this enzyme. This will give an indication of the feasibility of proposed mechanism *in vivo*.

Further ahead, machine learning (ML) could be applied to β-methylmalyl-lyase and potential mutations suggested by this study can then be assessed using similar QM/MM and MD methods. ML is a tool that has regained popularity in recent years due to increased computational power allowing the manipulation of big data sets and the improved accessibility of ML algorithms. The direction for this study is to apply these algorithms to enzyme itself by generating relevant descriptors from the active site of the enzyme. The implementation of ML should produce novel candidate models.[43]

Grillo et al.[44] describe a QM/MM method which defines reactivity descriptors to illustrate enzyme activity. Various theoretical properties can be derived from the electronic structure of a molecule using DFT and those related to reactivity are particularly useful when calculated from time-step conformations of enzymes as a reaction progresses. These descriptors use the concept of electronic chemical potentials to map the movement of electrons over time by quantifying the hardness and softness of the system. These can be combined with Fukui functions to define the electron density of the system to indicate where the reaction takes place and in what orientation. [45]

The QM/MM model will provide the necessary information about the active site so the descriptors can be calculated for each frame of the simulation. This should provide a dynamic map of the electrons as the reaction occurs showing the atom movements. Using these reactivity descriptors, models can be built to predict changes in the enzyme structure to improve or change the reactivity of the active site. The mutations the algorithm suggests will need to be assessed for stability and whether they are able to improve the activity of the enzyme.

The conclusion of this workflow would be to test these mutations through the MD and QM/MM simulations described above. Viable candidates could then be passed on for verification in experiment. A systematic workflow such as this will be a novel contribution to the field of rational design and could be applied to any enzymatic system.

References

1. M. Naseem, Ö. Osmanoglu and T. Dandekar, *Trends Biotechnol.*, 2020, **38**, 354-359.
2. A. Bar-Even, E. Noor, N. E. Lewis and R. Milo, *PNAS*, 2010, **107**, 8889.
3. P. L. Cummins, B. Kannappan and J. E. Gready, *Front. Plant Sci.*, 2018, **9**.
4. T. Schwander, L. Schada von Borzyskowski, S. Burgener, N. S. Cortina and T. J. Erb, *Science*, 2016, **354**, 900.
5. T. E. Miller, T. Beneyton, T. Schwander, C. Diehl, M. Girault, R. McLean, T. Chotel, P. Claus, N. S. Cortina, J.-C. Baret and T. J. Erb, *Science*, 2020, **368**, 649.
6. R. A. Sheldon, D. Brady and M. L. Bode, *Chem.*, 2020, **11**, 2587-2605.
7. R. Chowdhury and C. D. Maranas, *AIChE Journal*, 2020, **66**, e16847.
8. J. M. Gonzalez, R. Marti-Arbona, J. C. H. Chen and C. J. Unkefer, *ACSFEN*, 2017, **73**, 79-85.
9. J. Zarzycki and C. A. Kerfeld, *BMC Struct.l Bio.*, 2013, **13**, 28.
10. M. Meister, S. Saum, B. E. Alber and G. Fuchs, *J. Bacteriol.*, 2005, **187**, 1415.
11. T. J. Erb, L. Frerichs-Revermann, G. Fuchs and B. E. Alber, *J. Bacteriol.*, 2010, **192**, 1249.
12. W. Tang, Z. Wang, C. Zhang, C. Wang, Z. Min, X. Zhang, D. Liu, J. Shen and X. Xu, *Biochem. Biophys. Res. Commun.*, 2019, **518**, 72-79.
13. L. B. Hersh, *J. Biol. Chem.*, 1974, **249**, 5208-5212.
14. C. E. Quartararo and J. S. Blanchard, *Biochemistry*, 2011, **50**, 6879-6887.
15. B. R. Howard, J. A. Endrizzi and S. J. Remington, *Biochemistry*, 2000, **39**, 3156-3168.
16. M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708-2728.
17. M. W. van der Kamp, F. Perruccio and A. J. Mulholland, *Chem Commun (Camb)*, 2008, 1874-1876.
18. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
19. E. Paquet and H. L. Viktor, *BioMed Res. Int.*, 2015, **2015**, 183918.
20. P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson and M. T. Stahl, *J. Chem. Inf. Model.*, 2010, **50**, 572-584.
21. M. McGann, *J. Comput. Aided Mol. Des.*, 2012, **26**, 897-906.
22. M. McGann, *J. Chem. Inf. Model.*, 2011, **51**, 578-596.
23. M. C. R. Melo, R. C. Bernardi, T. Rudack, M. Scheurer, C. Riplinger, J. C. Phillips, J. D. C. Maia, G. B. Rocha, J. V. Ribeiro, J. E. Stone, F. Neese, K. Schulten and Z. Luthey-Schulten, *Nat. Methods*, 2018, **15**, 351-354.

24. M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525-537.
25. C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comp.*, 2011, **7**, 2284-2295.
26. S. Jo, T. Kim, V. G. Iyer and W. Im, *J Comput. Chem.*, 2008, **29**, 1859-1865.
27. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
28. D. Beglov and B. Roux, *J. Chem. Phys.*, 1994, **100**, 9050-9063.
29. A. Aleksandrov and M. Field, *PCCP*, 2011, **13**, 10503-10509.
30. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell Jr, *J. Comput. Chem.*, 2010, **31**, 671-690.
31. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613.
32. S. E. Feller, R. W. Pastor, A. Rojnuckarin, S. Bogusz and B. R. Brooks, *J. Phys. Chem.*, 1996, **100**, 17011-17020.
33. Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S.

R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184-215.

34. O. Guvench and A. D. MacKerell, Jr., *J. Mol. Model.*, 2008, **14**, 667-679.
35. J. J. P. Stewart, *Journal*, 2016.
36. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.
37. Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, C. Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, L.-P. Wang, Thomas J. Lane and Vijay S. Pande, *Biophys. J.*, 2015, **109**, 1528-1532.
38. A. C. Wallace, R. A. Laskowski and J. M. Thornton, *Protein Eng.*, 1995, **8**, 127-134.
39. L. Schrodinger, *Journal*, 2015.
40. S. Herter, A. Busch and G. Fuchs, *J. Bacteriol.*, 2002, **184**, 5999-6006.
41. Y. Okubo, S. Yang, L. Chistoserdova and M. E. Lidstrom, *J. Bacteriol.*, 2010, **192**, 1813.
42. M. W. van der Kamp, J. Żurek, F. R. Manby, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. B*, 2010, **114**, 11303-11314.
43. S. Mazurenko, Z. Prokop and J. Damborsky, *ACS Catal.*, 2020, **10**, 1210-1223.
44. I. B. Grillo, G. A. Urquiza-Carvalho, J. F. R. Bachega and G. B. Rocha, *J. Chem. Inf. Model.*, 2020, **60**, 578-591.
45. J. Oller, D. A. Saez and E. Vöhringer-Martinez, *J. Phys. Chem. A*, 2020, **124**, 849-857.

6. Appendices

6.1. MD

All the figures below are representative of all four enzyme complexes and all repeated simulations. Minimisation was done for 2000 or 3000 steps until the energy of the system decreased no further to ensure a truly minimised protein structure (figure 6.1).

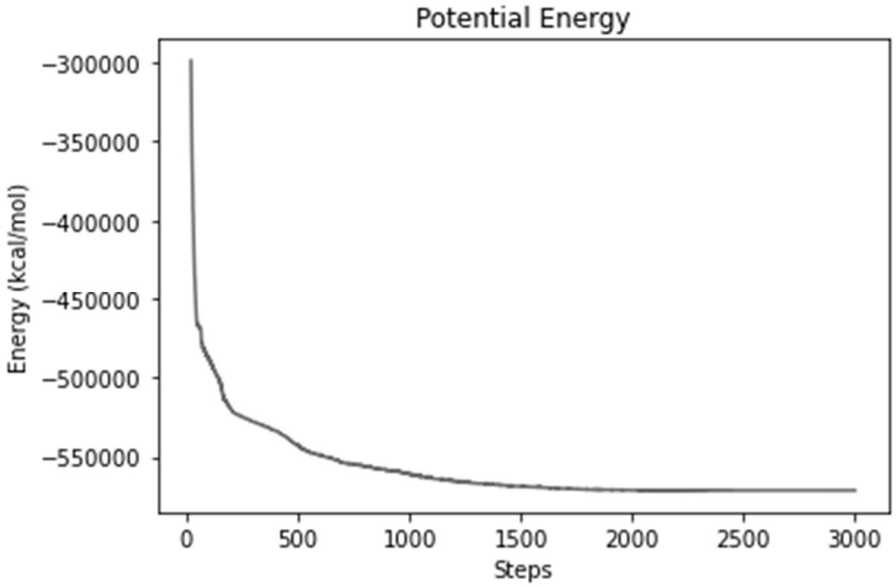

Figure 6.1: The potential energy of the system showing that 3000 steps of minimisation is enough for the energy to plateaux and thus the system is minimised

Equilibration of the MD simulation was monitored in order to produce a reliable system. After 0.1 ns of equilibration under the NVT ensemble, a further 4.9 ns were simulated using the NPT ensemble. RMSD was calculated for all atoms except hydrogen atoms over that time to assess the amount of movement between timesteps (figure 6.2). If RMSD were to rise above 2 Å there would be cause for concern as it may indicate that the protein is beginning to fall apart and denature.

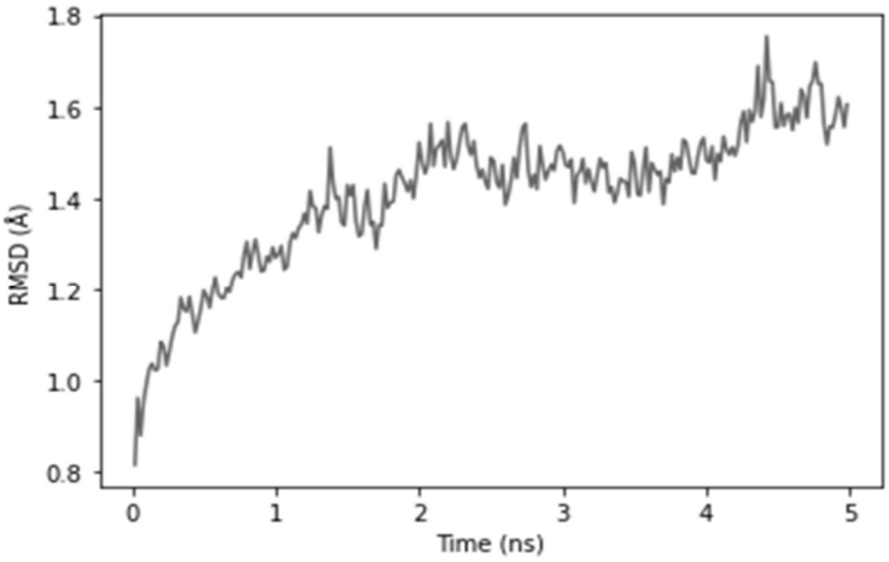

Figure 6.2: The RMSD of the system during equilibration using the NPT ensemble. The system seems to stabilise after 2 ns.

The final coordinates for each structure were selected after the system had stabilised as seen by the calculated RMSD. These structures were used as the starting point for both the MD and the QM/MM simulations.

6.2. QM/MM

Two QM regions were set up for simulation. One including just the amino acids involved in the mechanism (table 6.1) and a second including all amino acids known to be conserved in the wider family CitE enzyme family that are present in the active site (table 6.2).

Table 6.1: A list of the atoms included in the smaller QM region for the complex containing propionyl-CoA. 1VU represents propionyl-CoA, GLV is glyoxylate, TIP3 corresponds to the water molecules, MG the magnesium atom, VAL – valine, ASP – aspartic acid, and ARG – arginine.

Atom Number	Segment	Residue Number	Residue	Atom
1	PROA	298	VAL	C
2	PROA	298	VAL	O
3	PROA	299	ASP	N
4	PROA	299	ASP	HN
5	PROA	299	ASP	CA
6	PROA	299	ASP	HA
7	PROA	299	ASP	CB
8	PROA	299	ASP	HB1
9	PROA	299	ASP	HB2
10	PROA	299	ASP	CG
11	PROA	299	ASP	OD1
12	PROA	299	ASP	OD2
13	PROB	75	VAL	C
14	PROB	75	VAL	O
15	PROB	76	ARG	N
16	PROB	76	ARG	HN
17	PROB	76	ARG	CA

18	PROB	76	ARG	HA
19	PROB	76	ARG	CB
20	PROB	76	ARG	HB1
21	PROB	76	ARG	HB2
22	PROB	76	ARG	CG
23	PROB	76	ARG	HG1
24	PROB	76	ARG	HG2
25	PROB	76	ARG	CD
26	PROB	76	ARG	HD1
27	PROB	76	ARG	HD2
28	PROB	76	ARG	NE
29	PROB	76	ARG	HE
30	PROB	76	ARG	CZ
31	PROB	76	ARG	NH1
32	PROB	76	ARG	HH11
33	PROB	76	ARG	HH12
34	PROB	76	ARG	NH2
35	PROB	76	ARG	HH21
36	PROB	76	ARG	HH22
37	GLVB	2	GLV	C1
38	GLVB	2	GLV	O1
39	GLVB	2	GLV	C2
40	GLVB	2	GLV	O2
41	GLVB	2	GLV	O3
42	GLVB	2	GLV	H01
43	1VU	1	1VU	C01
44	1VU	1	1VU	H36
45	1VU	1	1VU	H37
46	1VU	1	1VU	H03
47	1VU	1	1VU	C1
48	1VU	1	1VU	H01
49	1VU	1	1VU	H02
50	1VU	1	1VU	C2
51	1VU	1	1VU	O
52	1VU	1	1VU	S
53	1VU	1	1VU	C3
54	1VU	1	1VU	H06
55	1VU	1	1VU	H07
56	1VU	1	1VU	C4
57	1VU	1	1VU	H09
58	1VU	1	1VU	H10
59	1VU	1	1VU	N
60	1VU	1	1VU	H04
61	1VU	1	1VU	C5
62	1VU	1	1VU	O1
63	1VU	1	1VU	C6
64	1VU	1	1VU	H13
65	1VU	1	1VU	H14

66	HETE	403	MG	MG
67	WATB	510	TIP3	OH2
68	WATB	510	TIP3	H1
69	WATB	510	TIP3	H2
70	WATB	724	TIP3	OH2
71	WATB	724	TIP3	H1
72	WATB	724	TIP3	H2

Table 6.2: A list of the atoms included in the smaller QM region for the complex containing propionyl-CoA. 1VU represents propionyl-CoA, GLV is glyoxylate, TIP3 corresponds to the water molecules, MG the magnesium atom, VAL – valine, ASP – aspartic acid, LEU - leucine, GLU – glutamic acid, ARG – arginine, and ILE - isoleucine.

Atom Number	Segment	Residue Number	Residue	Atom
1	PROA	298	VAL	C
2	PROA	298	VAL	O
3	PROA	299	ASP	N
4	PROA	299	ASP	HN
5	PROA	299	ASP	CA
6	PROA	299	ASP	HA
7	PROA	299	ASP	CB
8	PROA	299	ASP	HB1
9	PROA	299	ASP	HB2
10	PROA	299	ASP	CG
11	PROA	299	ASP	OD1
12	PROA	299	ASP	OD2
13	PROB	41	LEU	C
14	PROB	41	LEU	O
15	PROB	42	ASP	N
16	PROB	42	ASP	HN
17	PROB	42	ASP	CA
18	PROB	42	ASP	HA
19	PROB	42	ASP	CB
20	PROB	42	ASP	HB1
21	PROB	42	ASP	HB2
22	PROB	42	ASP	CG
23	PROB	42	ASP	OD1
24	PROB	42	ASP	OD2
25	PROB	43	LEU	C
26	PROB	43	LEU	O
27	PROB	44	GLU	N
28	PROB	44	GLU	HN
29	PROB	44	GLU	CA
30	PROB	44	GLU	HA
31	PROB	44	GLU	CB
32	PROB	44	GLU	HB1
33	PROB	44	GLU	HB2
34	PROB	44	GLU	CG

35	PROB	44	GLU	HG1
36	PROB	44	GLU	HG2
37	PROB	44	GLU	CD
38	PROB	44	GLU	OE1
39	PROB	44	GLU	OE2
40	PROB	44	GLU	C
41	PROB	44	GLU	O
42	PROB	45	ASP	N
43	PROB	45	ASP	HN
44	PROB	45	ASP	CA
45	PROB	45	ASP	HA
46	PROB	45	ASP	CB
47	PROB	45	ASP	HB1
48	PROB	45	ASP	HB2
49	PROB	45	ASP	CG
50	PROB	45	ASP	OD1
51	PROB	45	ASP	OD2
52	PROB	75	VAL	C
53	PROB	75	VAL	O
54	PROB	76	ARG	N
55	PROB	76	ARG	HN
56	PROB	76	ARG	CA
57	PROB	76	ARG	HA
58	PROB	76	ARG	CB
59	PROB	76	ARG	HB1
60	PROB	76	ARG	HB2
61	PROB	76	ARG	CG
62	PROB	76	ARG	HG1
63	PROB	76	ARG	HG2
64	PROB	76	ARG	CD
65	PROB	76	ARG	HD1
66	PROB	76	ARG	HD2
67	PROB	76	ARG	NE
68	PROB	76	ARG	HE
69	PROB	76	ARG	CZ
70	PROB	76	ARG	NH1
71	PROB	76	ARG	HH11
72	PROB	76	ARG	HH12
73	PROB	76	ARG	NH2
74	PROB	76	ARG	HH21
75	PROB	76	ARG	HH22
76	PROB	140	ILE	C
77	PROB	140	ILE	O
78	PROB	141	GLU	N
79	PROB	141	GLU	HN
80	PROB	141	GLU	CA
81	PROB	141	GLU	HA
82	PROB	141	GLU	CB

83	PROB	141	GLU	HB1
84	PROB	141	GLU	HB2
85	PROB	141	GLU	CG
86	PROB	141	GLU	HG1
87	PROB	141	GLU	HG2
88	PROB	141	GLU	CD
89	PROB	141	GLU	OE1
90	PROB	141	GLU	OE2
91	PROB	167	ALA	C
92	PROB	167	ALA	O
93	PROB	168	ASP	N
94	PROB	168	ASP	HN
95	PROB	168	ASP	CA
96	PROB	168	ASP	HA
97	PROB	168	ASP	CB
98	PROB	168	ASP	HB1
99	PROB	168	ASP	HB2
100	PROB	168	ASP	CG
101	PROB	168	ASP	OD1
102	PROB	168	ASP	OD2
103	GLVB	2	GLV	C1
104	GLVB	2	GLV	O1
105	GLVB	2	GLV	C2
106	GLVB	2	GLV	O2
107	GLVB	2	GLV	O3
108	GLVB	2	GLV	H01
109	1VU	1	1VU	C01
110	1VU	1	1VU	H36
111	1VU	1	1VU	H37
112	1VU	1	1VU	H03
113	1VU	1	1VU	C1
114	1VU	1	1VU	H01
115	1VU	1	1VU	H02
116	1VU	1	1VU	C2
117	1VU	1	1VU	O
118	1VU	1	1VU	S
119	1VU	1	1VU	C3
120	1VU	1	1VU	H06
121	1VU	1	1VU	H07
122	1VU	1	1VU	C4
123	1VU	1	1VU	H09
124	1VU	1	1VU	H10
125	1VU	1	1VU	N
126	1VU	1	1VU	H04
127	1VU	1	1VU	C5
128	1VU	1	1VU	O1
129	1VU	1	1VU	C6
130	1VU	1	1VU	H13

131	1VU	1	1VU	H14
132	HETE	403	MG	MG
133	WATB	510	TIP3	OH2
134	WATB	510	TIP3	H1
135	WATB	510	TIP3	H2
136	WATB	724	TIP3	OH2
137	WATB	724	TIP3	H1
138	WATB	724	TIP3	H2

The RMSD was calculated for the QM/MM simulations and all the figures below are representative of all four enzyme complexes and all five repeated simulations (figures 6.3 & 6.4).

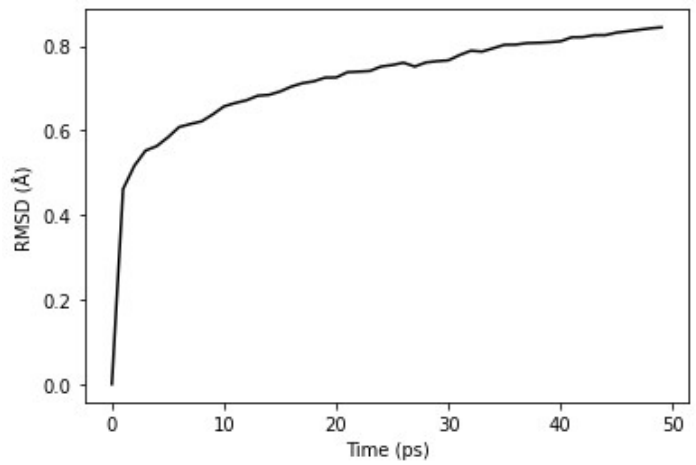

Figure 6.3: RMSD of the MM partition of the QM/MM simulation; the complex stabilises early in the simulations at around 10 ps.

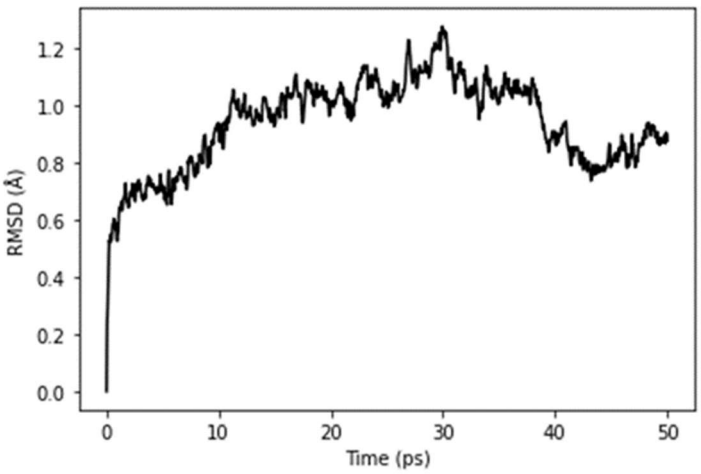

Figure 6.4: RMSD of the atoms only modelled by QM semiempirical method.

The energy in the system is also a good indicator of the stability of the system and large transgressions from the norm can indicate abnormal behaviour (figure 6.5).

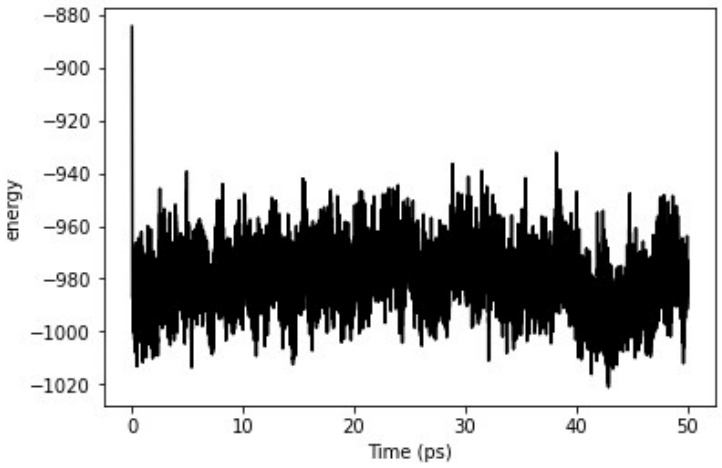

Figure 6.5: The energy in the QM region of the QM/MM model remains steady throughout all the simulations.