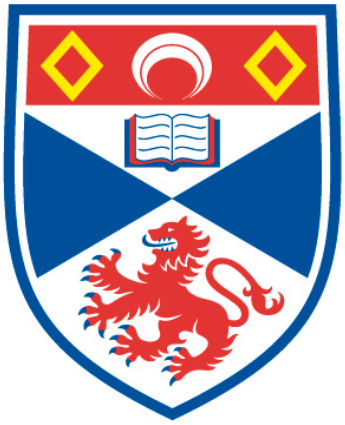
Computational insights into catalytic mechanism and thermostability of the enzyme Is-PETase

Eugene Shrimpton-Phoenix

A thesis submitted for the degree of PhD
at the
University of St Andrews



2023

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/643

July 2023

Candidate's declaration

I, Eugene Shrimpton-Phoenix, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 65,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2018.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 17th July 2023 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 17th July 2023 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Eugene Shrimpton-Phoenix, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 17th July 2023 Signature of candidate

Date 17th July 2023 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Eugene Shrimpton-Phoenix, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 17th July 2023 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date Signature of candidate

Date Signature of supervisor

Acknowledgements

I would like to thank:

- My supervisors Michael Bühl and John Mitchell for their guidance throughout my PhD.
- Jonathan Colburn for introducing me to the world of QM/MM and for sharing his scripts with me.
- EASTBIO and the University of St Andrews for providing funding for my project.
- Herbert Früchtl for keeping our various HPCs running and his assistance on technical matters.
- Tanja van Mourik and Rebecca Goss for their guidance during our yearly review meetings.
- Everyone in room 150 for their discussions of my project, and general chitchat.
- My master's projects students for doing work related to my projects: Chloe Martin, Alexandra Guerin, Lilli Kuerten, and Sylvia Kaempf.
- My partner Esmé Dove for providing so much support throughout my PhD.
- My son Ivan Dove-Phoenix. Honestly - he didn't help much in the writing of my thesis, which is understandable as he is just a baby.

Abstract

The field of study surrounding the PETase class of enzymes has gained a great deal of popularity in the past five years. PETases are enzymes capable of degrading poly(ethylene) terephthalate and presents an opportunity to bioanalytically recycle this common pollutant. The most extensively researched PETase is the enzyme *Is*-PETase, which originates from the organism *Idonella sakaiensis* and was discovered near a recycling plant in 2016. To further the utility of *Is*-PETase in the industrial recycling of poly(ethylene) terephthalate, improvements in desirable properties such as catalytic rate and thermostability must be achieved via introduction of amino-acid substitutions to the enzyme.

In order to further improve the activity of *Is*-PETase, it would be informative to have a deeper understanding of its catalytic mechanism. To this end, we have investigated the catalytic mechanism of the degradation of poly(ethylene) terephthalate via *Is*-PETase. We applied hybrid quantum mechanical/ molecular mechanical to generate several independent reaction profiles. From these reaction profiles we have concluded that this reaction proceeds with an overall activation barrier of 35.6 kJ mol^{-1}. This relatively low activation barrier suggests that the 'true' rate-limiting step for this reaction is a physical process such as substrate binding or product dissociation. Our calculations provide preliminary evidence for the product dissociation step being rate-limiting.

We have also investigated modifications to wild-type *Is*-PETase that contribute to the enzyme's thermostability. We have created our own mutant of *Is*-PETase, DISU-PETase, through the *in silico* introduction of a novel disulfide bond. We generated reaction profiles for DISU-PETase using the same quantum mechanical/ molecular mechanical techniques as applied to wild-type *Is*-PETase. We found that introduction of the novel disulfide bond in DISU-PETase had no adverse effect upon the reaction profile of the degradation of poly(ethylene) terephthalate. We have also applied molecular dynamics techniques to investigate the intramolecular interactions that contribute toward the thermostability of the most highly active variants of *Is*-PETase in the literature.

We have also used our quantum mechanical/ molecular mechanical techniques to generate reaction profiles for the degradation of an alternate substrate, poly(ethylene) furanoate via wild-type *Is*-PETase.

Table of Contents

Table of Abbreviations

DEFINITION	ABBREVIATION
BROYDEN-FLETCHER-GOLDFARB-SHANNO	BFGS
BIS(2-HYDROXYETHYL) TEREPHTHALATE	BHET
BASIC LOCAL ALIGNMENT SEARCH TOOL	BLAST
BORN-OPPENHEIMER MOLECULAR DYNAMICS	BOMD
***CANDIDA ANTARCTICA* LIPASE B**	CALB
DENSITY FUNCTIONAL THEORY	DFT
DIMETHYLSULFOXIDE	DMSO
DIFFERENTIAL SCANNING FLUORIMETRY	DSF
ETHYLENE GLYCOL	EG
FORMIC ACID	FA
2,5-FURANDICARBOXYLIC ACID	FDCA
FREE-ENERGY LANDSCAPE	FEL
GENERALISED GRADIENT APPROXIMATION	GGA
GENETICALLY MODIFIED ORGANISMS	GMOs
GAUSSIAN-TYPE ORBITALS	GTO
HOMOGENEOUS ELECTRON GAS	HEG
2-HYDROXYETHYL METHYLFURANOATE	HEMF
2-HYDROXYETHYL METHYLTEREPHTHALATE	HEMT
HARTREE-FOCK	HF
HOHENBERG-KOHN	HK
5-HYDROXYMETHYLFURFURAL	HMFF
2-HYDROXYETHYL FURANOIC ACID	HMF
HIGH-PRESSURE LIQUID CHROMATOGRAPHY	HPLC
KOHN-SHAM	KS
LEVULINIC ACID	LA
LEAF COMPOST CUTINASE	LCC
LOCAL DENSITY APPROXIMATION	LDA
MINIMUM ENERGY PATHWAY	MEP
MONO(2-HYDROXYETHYL) FURANOATE	MHEF
MONO(2-HYDROXYETHYL) TEREPHTHALIC ACID	MHET
MOLECULAR MECHANICS	MM
MULTIPLE SEQUENCE ALIGNMENT	MSA
MICROBIAL TRANSGLUTAMINASE	MTG
NUDGED-ELASTIC-BAND	NEB
NEAR-INFRARED	NIR
OXYANION HOLE	OAH
PRINCIPAL COMPONENT ANALYSIS	PCA
PEPTIDE-CHAIN-REACTION	PCR
PROTEIN DATA BANK	PDB
POLYETHYLENE GLYCOL	PEG
POTENTIAL ENERGY SURFACE	PES
POLY(ETHYLENE) TEREPHTHALATE	PET
QUANTUM MECHANICS	QM
RHIZOMUCOR MIEHEI LIPASE	RML
ROOT MEAN SQUARED FLUCTUATION	RMSF
SELF-CONSISTENT-FIELD	SCF
SCANNING ELECTRON MICROSCOPY	SEM
SLATER-TYPE ORBITAL	STO
TEREPHTHALIC ACID	TA
TRYPTOPHAN SYNTHASE ENZYMES	TRPS
WATER-MEDIATED HYDROGEN BOND	WMHB

Chapter 1: Introduction to Biocatalytic Degradation of Poly(ethylene) Terephthalate

1.1.1 Poly(ethylene) Terephthalate as a Pollutant

In the 1980s it became clear that poly(ethylene) terephthalate (PET) pollution represented a major threat to the marine environment. This was most apparent due to large marine animals and seabirds being found dead with undigested PET containers in their stomachs. Animals that consume these plastics can die through gastric blockage, starvation, perforation or rupture of the gastrointestinal tract[1]. Oily substances in the oceans are either absorbed by or adhere to PET and other plastics due to their hydrophobicity. It is thought to be this (rather than the size, shape, and colour) of these plastics that attracts the attention of marine animals[1]. While some of these oily substances are harmless, chemicals such as pesticides and industrial chemicals can cling to ocean plastics, becoming bioavailable to animals that ingest the plastics.

In recent years the more insidious threat of microplastics, produced through the photodegradation and mechanical erosion of PET and other plastics or through their use in cleaning and beauty products, has been revealed. The definition of microplastics is yet to be completely agreed on, but generally they can be understood to be small particles of plastic with sizes between 500µm and 5mm[2]. Microplastics are consumed by marine life that feeds on plankton, as they are a similar size. These contaminants are then passed to creatures that eat the plankton-feeders. Microplastics have been found at different concentrations across the entire aquatic food-chain with higher concentrations in high-level predators. Equally as concerning is that microplastics have been found to have entered our own food supply through contaminated fish stocks. The health effects of consumption of these microplastics are a topic of much research with the general consensus that exposure can lead to a wide array of conditions such as inflammation and neoplasia[3].

Perhaps the most dramatic result of oceanic plastic pollution is the so-called plastic islands located in the North Atlantic and Pacific Oceans (see Figure 1). Natural currents carry debris towards gyres which are areas where the currents move in a spiral pattern, where they accumulate. Contrary to the popular imagination, these plastic islands are large areas with high concentrations of small pieces of plastic and microplastics rather than physical floating islands of plastic refuse (although

these do still exist).

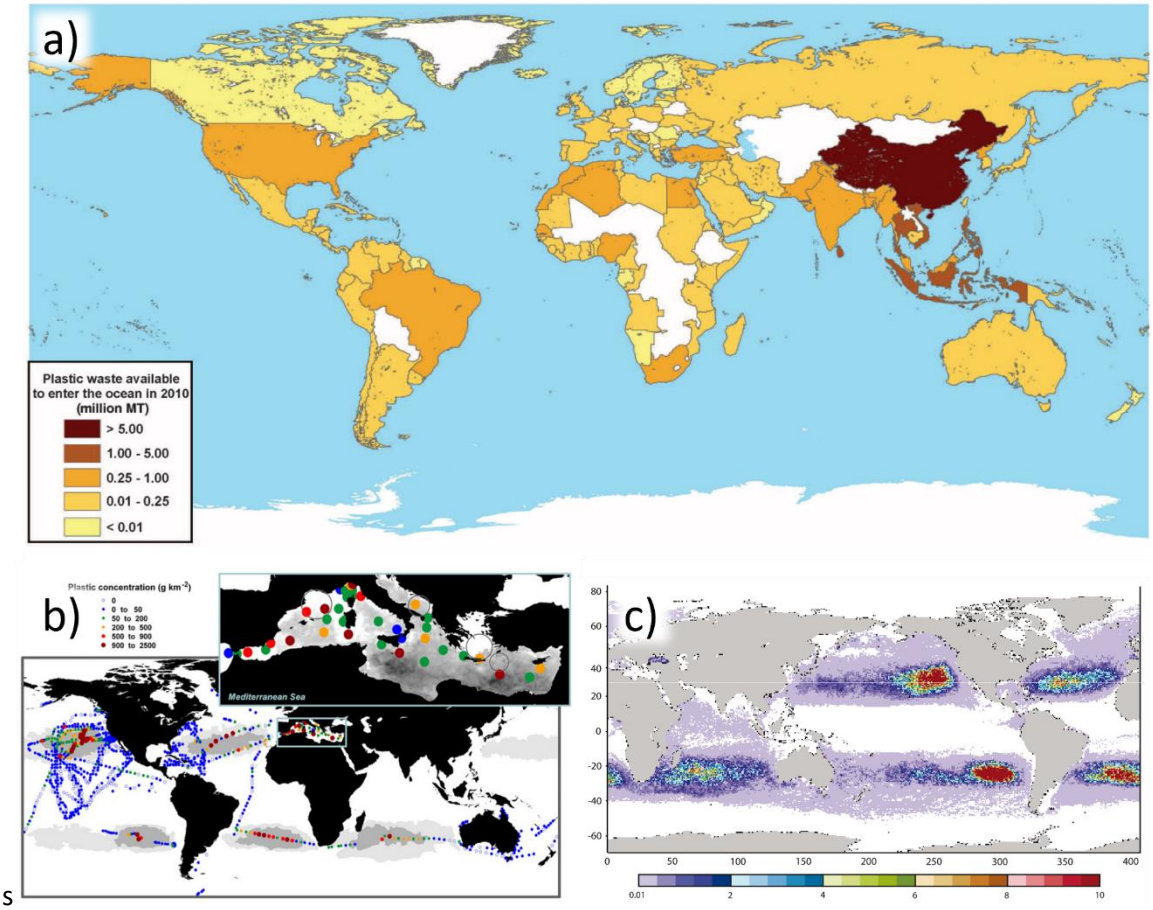


Figure 1: Maps of the world showing production of plastics waste by country (A, reprinted from Jambeck et al.[4]), measurements of concentration of plastics pollution in the ocean (B, reprinted from Cozar et. al[5]) and predicted distributions of plastic waste in the oceans based on simulations of ocean currents (C, reprinted from Kershaw et al.[6]).

PET was for many years considered to be impervious to enzymatic degradation, placing it firmly in the non-biodegradable category of plastics. Despite this obvious drawback (and in some cases because of this property) PET has been widely used since its invention in the early 1940's (see Figure 2). The primary use of PET is as a container for carbonated drinks and water, making up between 83%-84% of the global market demand[7], where the low gas permittivity of PET combined with its translucency, high chemical resistance and shatterproof properties make it an ideal material. Another common use of PET is as a fabric. A fabric that is soft, strong, and easy to wash can be produced using a series of treatments to woven PET fibres for a fraction of the cost of natural materials such as linen, cotton or wool.

Despite the clear threats to the environment and our health, global production of PET is steadily rising with 30.5 million metric tonnes produced in 2019 and 35.3 tonnes of forecast production in 2024 (according to statistica.com sourcing GlobalData[8]). Several waste-management solutions have

been implemented to mitigate the damage caused by PET in the environment. The most basic of these being the burning of waste plastics as fuel to generate electricity – this recovers a fraction of the energy used to manufacture the burned PET but has the major downside of releasing large amounts of CO_2 into the atmosphere, contributing to climate change. A more effective method is recycling; current industrial processes for PET recycling involve mechanical shredding followed by immersion at high temperature under either highly acidic or alkaline conditions. This process is approximately 70% less energy intensive when compared to the production of virgin PET from crude oil[9]. At each round of recycling, the physical properties of the PET produced deteriorate and the material becomes unsuitable for food packaging.

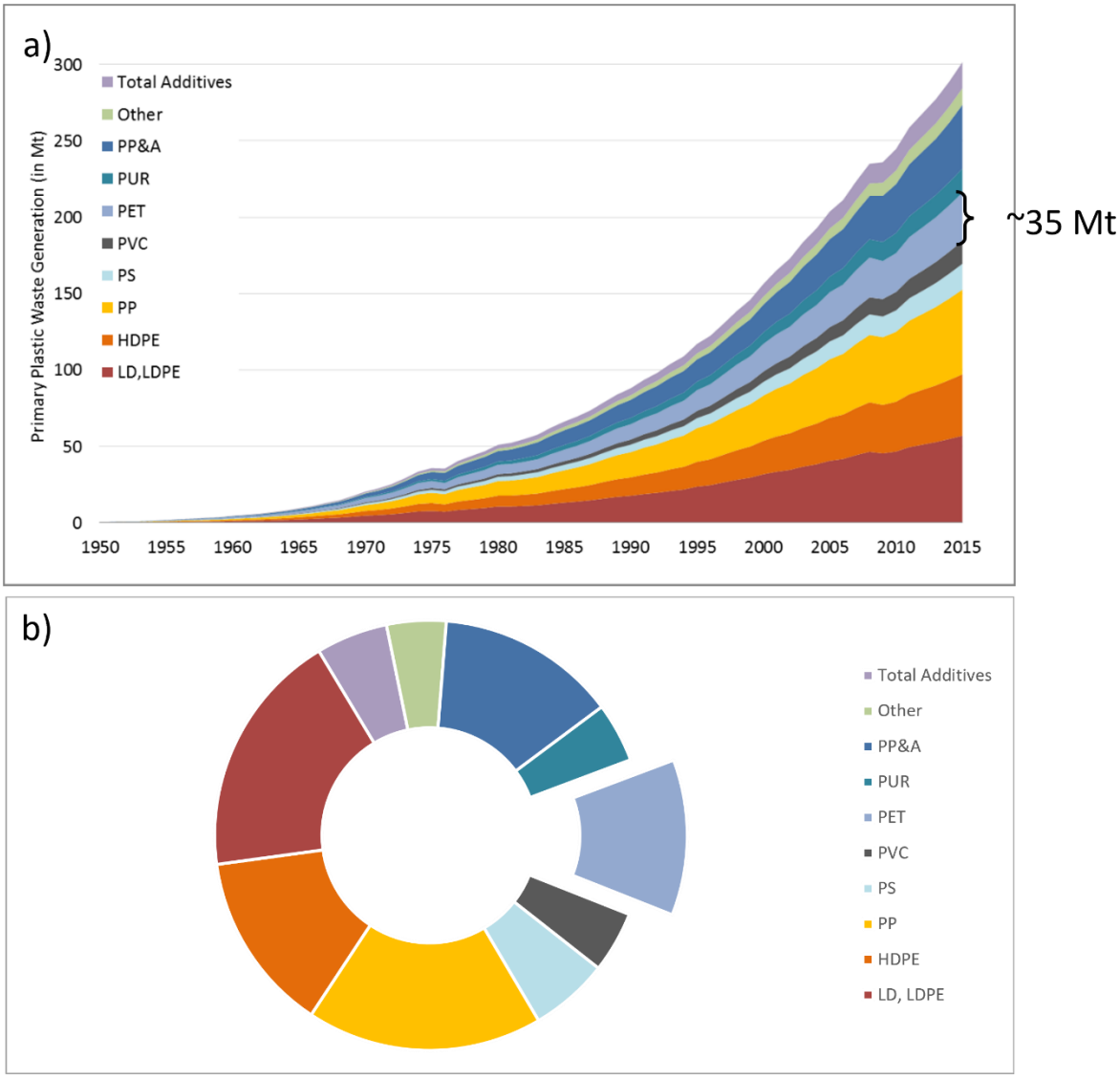


Figure 2: The increase in plastic waste production globally over time, by material type (A). The composition of the plastic waste produced globally in 2015 by material type (B), our material of interest PET has been separated for clearer viewing.

1.1.2 PET Degrading Enzymes

Most polyester plastics have some degree of biodegradability; this is due to the ester bonds that join their monomers being susceptible to hydrolysis by naturally occurring lipases. The resistance to biodegradation of PET however is due to its terephthalic acid subunits (see Figure 3). The high aromaticity and steric bulk makes binding to the active site of enzymes difficult. Additionally, inter-strand π-stacking interactions between aromatic rings allow PET to form highly ordered crystalline structures. Polyesterases have been found to prefer to hydrolyse amorphous regions of polymers. In cases like PET high crystallinity can restrict access to these amorphous regions making PET highly resistant to enzymatic degradation.

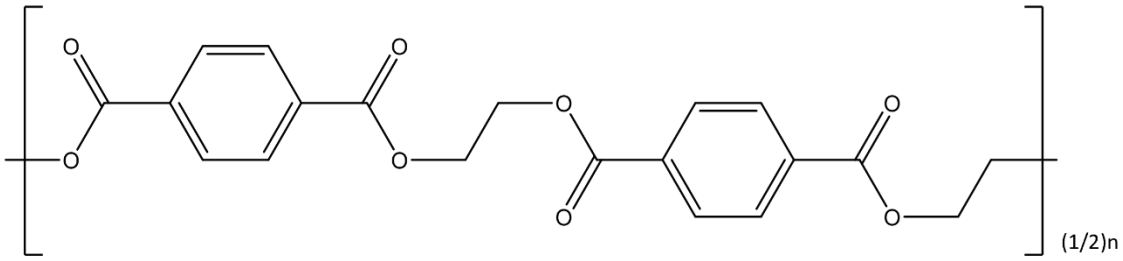


Figure 3: The chemical structure of PET, shown as its dimeric unit.

1.1.2.1 PET degrading enzyme TfH

It was thought for many years that no natural enzyme could degrade PET. This changed in 2005 when Müller et al.[10] published their findings on TfH, an enzyme derived from the actinomycete bacterium *Thermobifida fusca* that was capable of PET hydrolysis, albeit at a fairly sluggish rate. *Thermobifida fusca* is a thermophilic organism most commonly found in compost heaps or other self-heating sources of decaying organic matter.

Upon characterisation TfH was classified as a serine hydrolase belonging to the α/β hydrolase family of enzymes, with activity somewhere between that of an esterase and a lipase. This led to the classification of TfH as a cutinase. Cutinases naturally are used by fungi and bacteria to degrade cutin (see Figure 4), a fatty acid derived substance that binds to the primary cell wall of leaves[11]. Cutinase enzymes have characteristic solvent exposed active sites with shallow grooves as binding pockets, allowing the extended chains of cutin to bind.

The active site of TfH contains a catalytic triad of Ser130, His208 and Asp17. This is a conserved feature of α/β hydrolases and was thought to confer TfH's catalytic activity. Analysis of the X-ray crystal structure of TfH revealed that it has an enlarged binding pocket relative to related cutinases (see Figure 4). By widening and increasing the lipophilicity of the active site of a cutinase through the introduction of aromatic residues, its activity towards PET can be greatly increased[12]. The expansion of TfH's binding pocket is thought to have occurred naturally through an evolutionary process. For

this to have occurred *T. fusca* must receive an evolutionary advantage from its ability to degrade PET, for the time being that benefit remains a mystery. The speed at which natural enzymes have been adapted to exploit this new material in their environment is remarkable.

As its name suggests, *Thermobifida fusca* is a thermophilic organism. As a results of its host's natural environment, TfH has a high degree of thermostability (with an enzymatic melting point of 70°C[13]) and an optimum reaction temperature of 60°C[13]. These factors combined with the enzyme's tolerances to organic solvents[13] make TfH an attractive candidate for industrial processes.

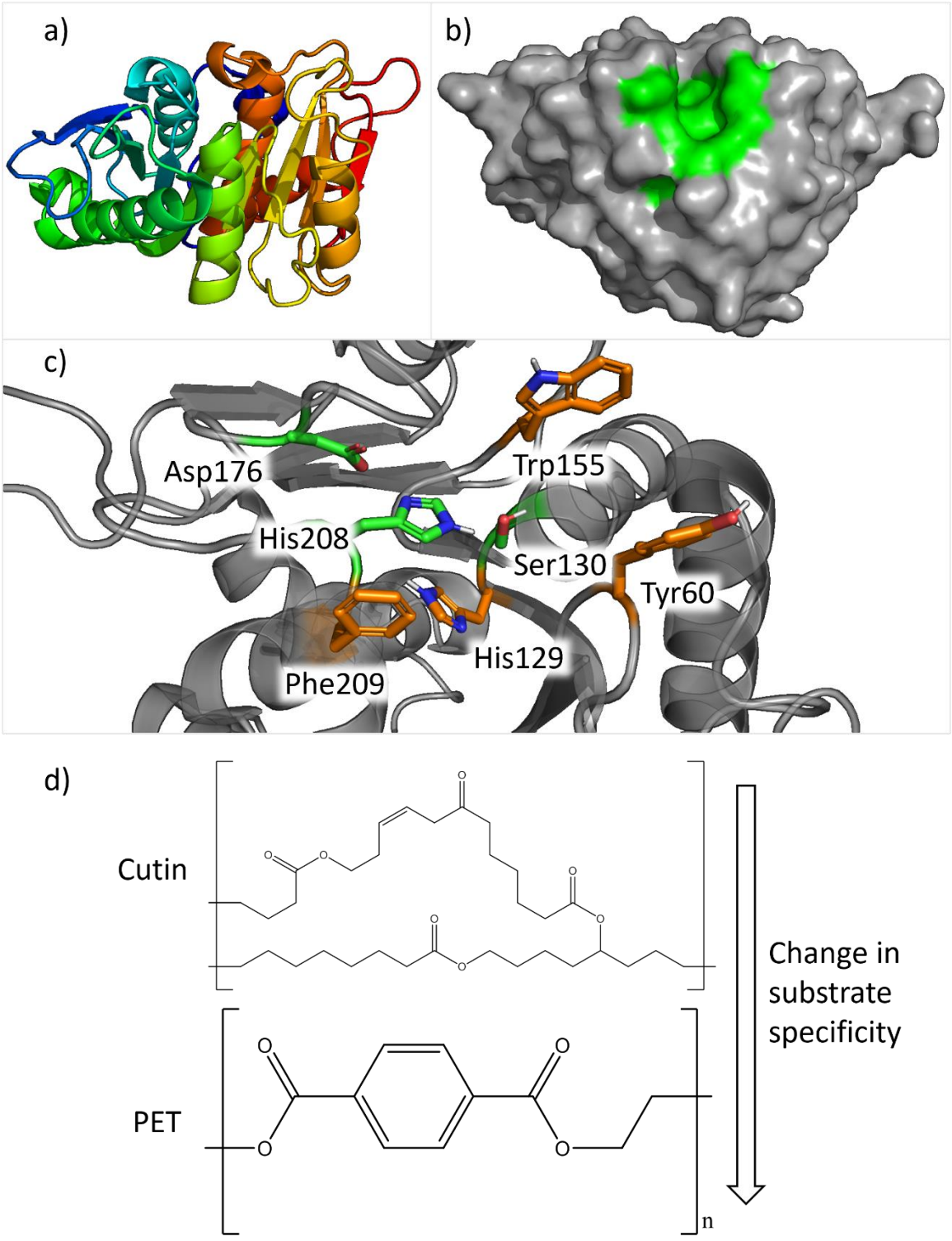


Figure 4: The tertiary structure (A) and binding pocket (B) of the PETase TfH. Key residues in the active site of TfH, including the catalytic triad S130-H208-D177 and aromatic residues Y60, H129, W155 and F209 (C). Scheme showing change in substrate specificity from cutin to PET (D). Structure taken from PDB entry 4CG3[13] and prepared in Pymol[14].

1.1.2.2 PET degrading enzyme LCC

Since the discovery of TfH several other PET degrading enzymes (now called PETases) have been found. These enzymes all adopt the α/β-hydrolase fold and use a Ser-His-Asp catalytic triad to

catalyse PET hydrolysis. The active sites of PETases are characteristically long, solvent exposed grooves which are necessary to accommodate their polymeric PET substrate. To this date the highest activity against PET for a naturally occurring enzyme belongs to the enzyme leaf compost cutinase (LCC). This enzyme was discovered through a metagenomic screening of DNA found in a leaf and branch compost[15]. The organism that produced LCC is currently unknown although it is likely to be a thermophilic bacterium due to LCC having a high sequence similarity (46%) with the bacterial TfH enzyme.

Analysis of LCC through X-ray crystallography reveals its characteristic open groove binding pocket as well as its catalytic triad made up of Ser165, His242 and Asp210 (see Figure 5). A comparison with the highly related cutinase originating from the organism *Thermobifida alba* revealed a similar hydrophobic region in LCC's binding pocket: this hints at the mechanism for binding the hydrophobic substrate PET.

LCC is able to outperform other natural PETases due to its high thermodynamic stability, with protein unfolding occurring at 86.2°C[16]. This allows enzymatic degradation of PET to occur at temperatures closer to the glass-transition point of PET (the importance of which is detailed in Chapter 1.1.3). Despite LCC's impressive thermodynamic stability it tends to agglomerate at higher temperatures, significantly reducing activity and making the optimum temperature for the enzyme's activity 50°C. Additives such as polyethylene glycol (PEG) and salts have been found to reduce this agglomeration effect but would significantly increase the cost of a scaled-up industrial process if used. An alternative to additives has been explored through the expression of LCC in a fungal host; this allows for N-glycosylation and O-glycosylation (attachment of sugar molecules through Asn, Ser and Thr residues) to occur at surface residues of the enzyme. This has been shown to reduce agglomeration and to increase the optimal temperature for LCC's activity by 10°C.

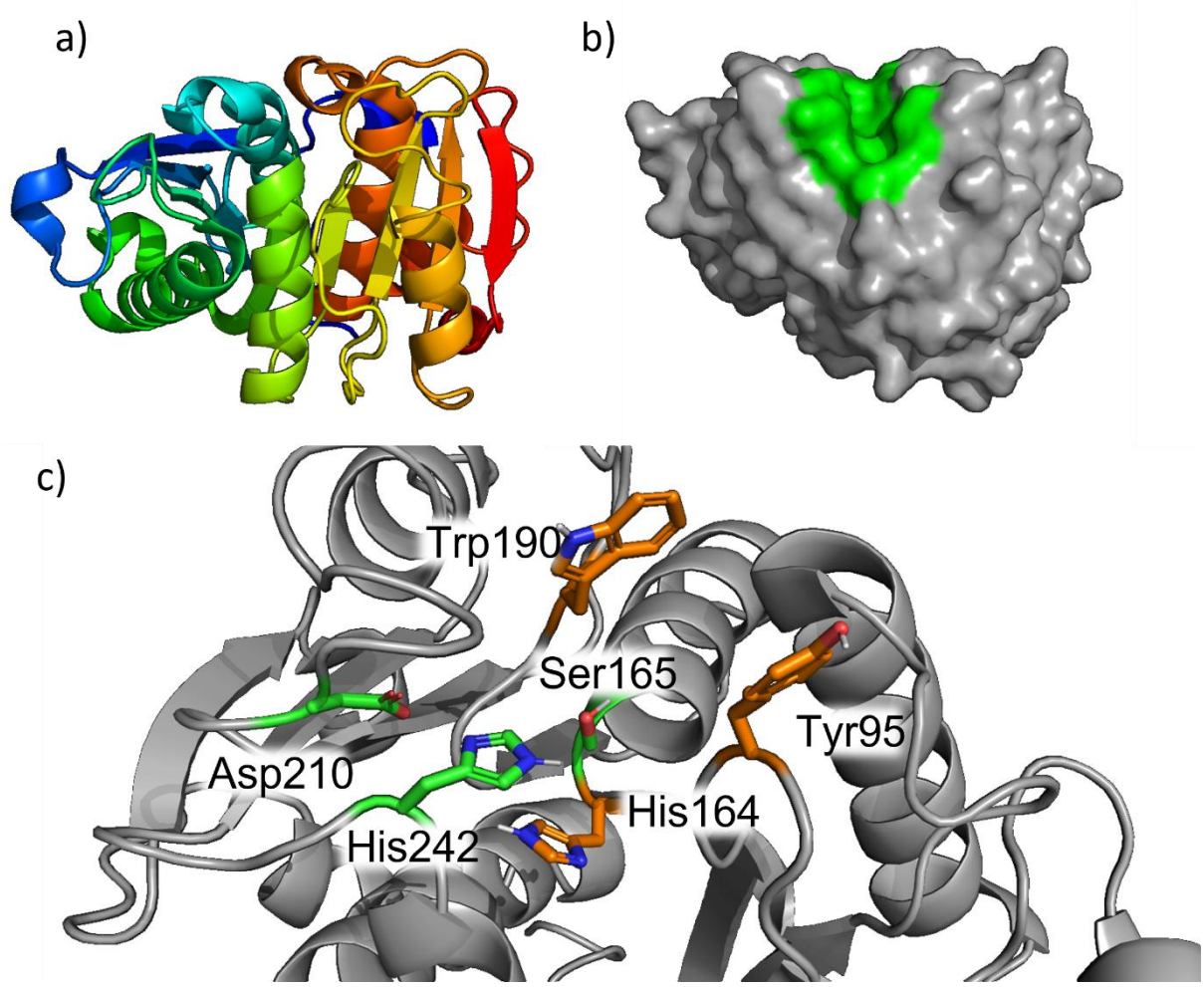


Figure 5: The tertiary structure (A) and binding pocket (B) of the PETase enzyme LCC. The key residues in the active site of LCC including the catalytic triad S165-H242-D210 and aromatic residues Y95, H164 and W190. Structures taken from PDB entry 4EBO[16] and prepared in Pymol[14].

1.1.2.3 Idonella Sakaiensis Discovery and Features

In 2016, a bacterium named *Idonella sakaiensis* was discovered by Yoshida *et al.* growing in a recycling plant in Sakai City, Japan[17]. *I. sakaiensis* was found to not only degrade PET but used PET as its primary carbon source. Through genomic analysis the enzymes responsible for this novel ability were found to be a cutinase-like enzyme *is*-PETase and an esterase MHETase[17]. The enzyme *Is*-PETase is secreted extra-cellularly[18] and is responsible for hydrolysing the amorphous regions of PET into the products mono(2-hydroxyethyl) terephthalic acid (MHET) and ethylene glycol (EG) with a trace product of terephthalic acid (TA). The MHET is then transported across the cell membrane of *I. sakaiensis* where it is further hydrolysed by MHETase into TA and EG (see Figure 6). The TA unit is then further broken down by as-of-yet unidentified enzymes and used as the primary carbon source for *I. sakaiensis*. Although it is possible that bacteria that produce cutinases such as TfH and LCC receive some evolutionary advantage for their PET degradation abilities, *I. sakaiensis* is the first organism identified that is clearly exploiting PET as a resource.

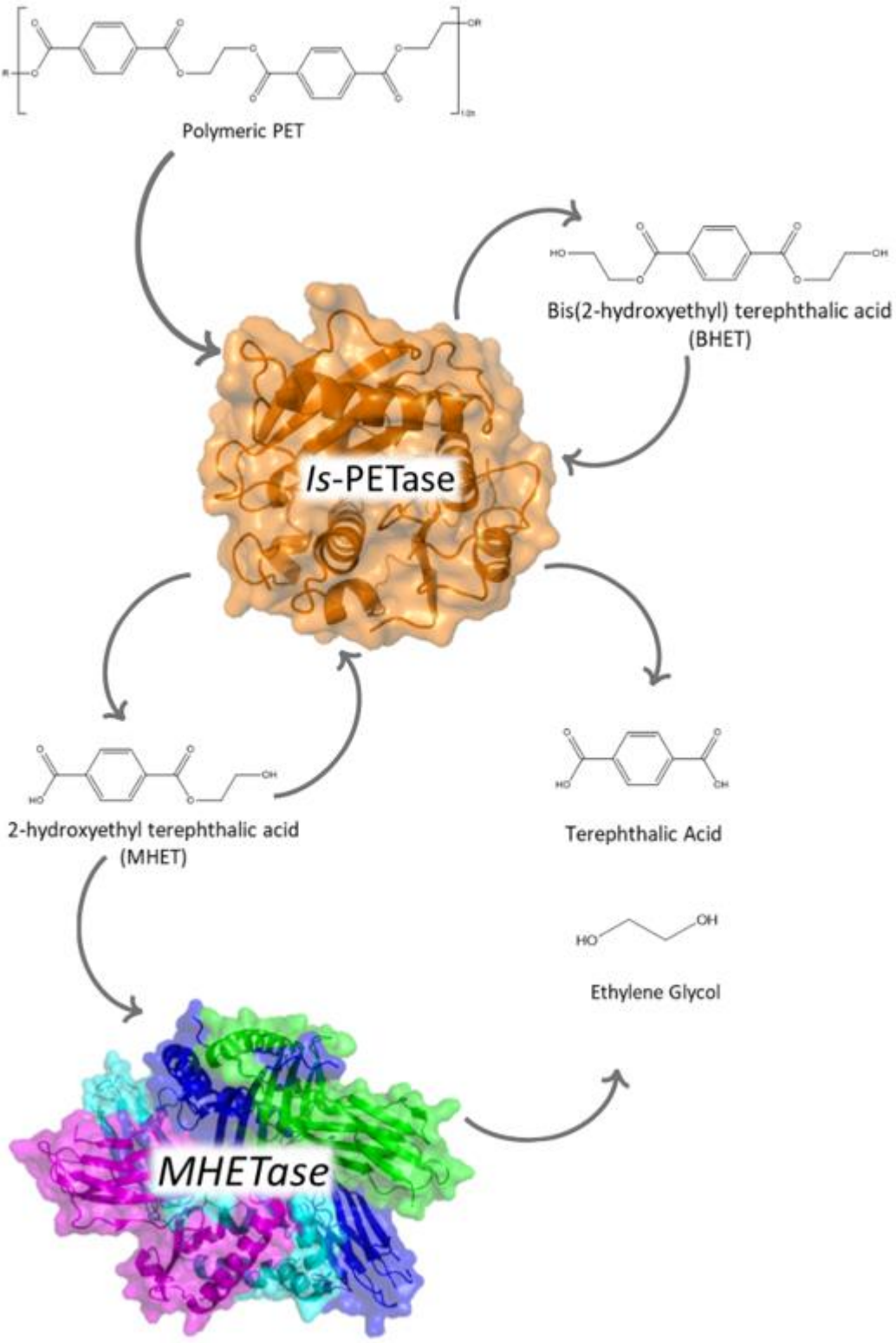


Figure 6: Scheme showing the degradation of polymeric PET by the enzyme pair Is-PETase and MHETase. The degradation of MHET via MHETase is faster than the various reactions catalysed by Is-PETase. This makes the presence of MHETase as well as Is-PETase accelerate the degradation of PET to TA relative to Is-PETase alone.

X-ray crystallography has revealed that *Is*-PETase belongs to the α/ß-hydrolase family[19], retaining the distinct tertiary structure of nine central β-sheets flanked by six α-helices and a series of loops[18,20,21] (see Figure 7). *Is*-PETase also retains the catalytic residues of the ancestral α/β-hydrolases: A catalytic triad made up of a Ser131, His208 and Asp177 provide a nucleophile in the form of the activated hydroxyl group of Ser131 (see Figure 7). This is achieved via the deprotonation of Ser131 by His208. His208 then acts as a proton shuttle, transferring its original proton to Asp177. Each of these catalytic residues resides on loops that connect an α-helix to a β-sheet; this feature is highly conserved throughout the α/β-hydrolase family[19]. Mutagenesis studies have confirmed the central importance of the residues in the catalytic triad by performing 'knock-out' experiments[17]; the replacement of any residue in the triad with an inert alanine residue resulted in near total abolition of activity towards PET.

The backbone nitrogen atoms of Met130 and Tyr58 act as an oxyanion hole stabilising negative charges that build up throughout catalysis through hydrogen bonding interactions. The oxyanion hole is a conserved characteristic of all cutinases. By forming a pair of hydrogen bonds with the ester carbonyl oxygen of the substrate, electron density is reduced in the ester carbonyl carbon making it more electrophilic, thus activating the substrate for nucleophilic attack via the hydroxyl group of Ser131. The oxyanion hole is also thought to function by lowering the energy of negatively charged transition states throughout the hydrolysis reaction, lowering the overall activation energy of the process.

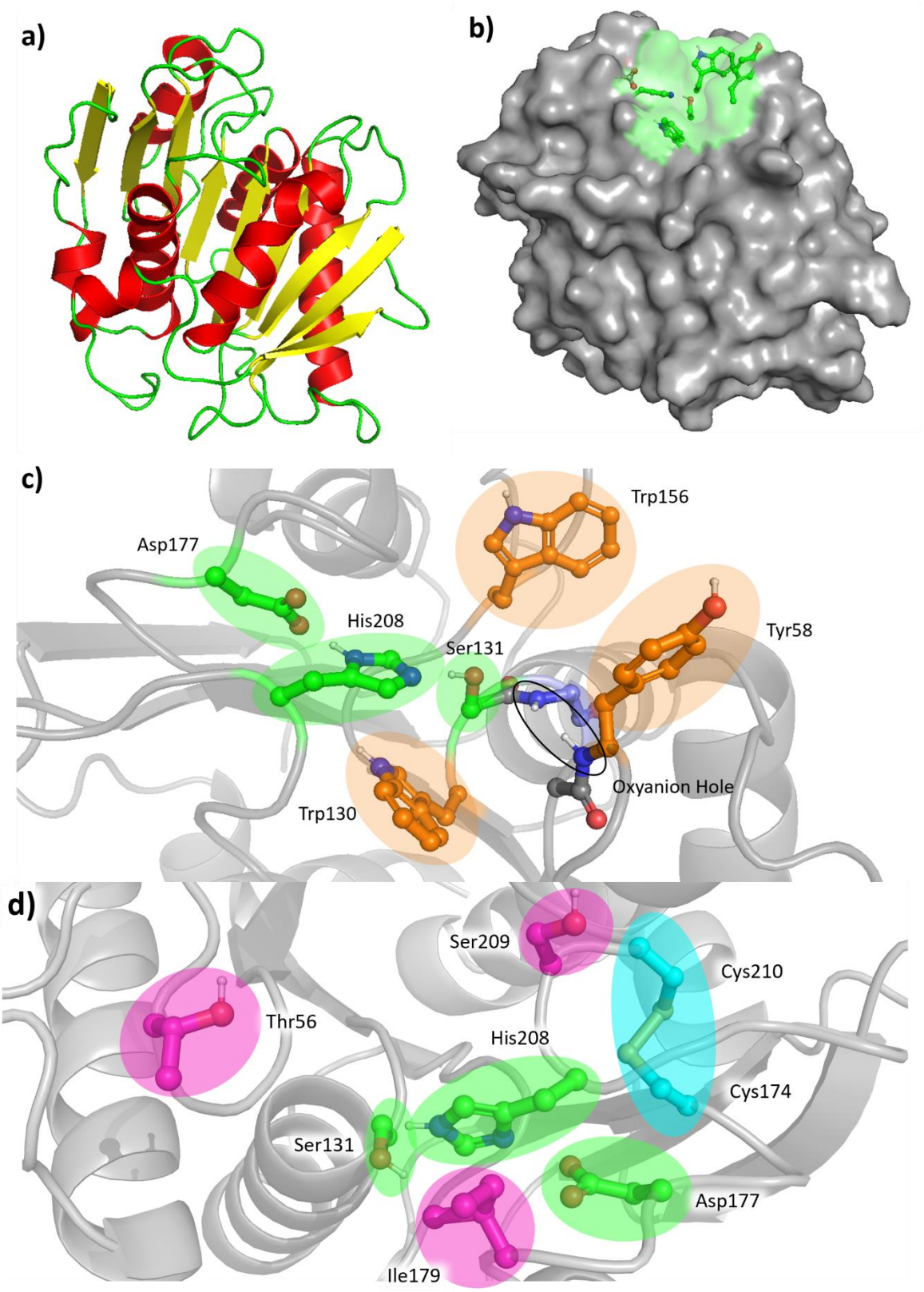


Figure 7: Key features of the enzyme Is-PETase. The tertiary structure (a) and location of the active site (b). Key amino acid residues in the active site (c and d) are shown, with the catalytic triad shown in green and key aromatic residues in orange. The oxyanion hole is identified by a black oval. Other non-catalytic residues that have been shown to be important for activity have been shown in magenta. The disulfide bond novel to Is-PETase is shown in cyan.

Sequence analysis revealed that *Is*-PETase has evolved from a cutinase-like enzyme and is related to TfH, the first discovered PET-degrading enzyme (with a sequence similarity of 52%)[17]. *Is*-PETase shares a similar solvent exposed binding pocket with TfH but has a significantly enlarged active site (up to threefold in some dimensions). This has been theorised to be essential for the binding of semicrystalline regions of PET[22]. Two amino acid substitutions seem to be responsible for this: A large, lipophilic residue Trp130 sits at the base of the active site. In cutinases this position is occupied by a smaller, more hydrophilic histidine residue. To accommodate Trp130 an additional mutation has occurred in an adjacent residue; the replacement of a phenylalanine (which is conserved throughout the cutinase family) with a smaller Ser209 allows Trp130 to rotate such that its aromatic side-chain lies 'flat' against the base of the active site. This serves the dual purpose of enlarging the active site of the enzyme and permitting the formation of T-stacking interactions with the aromatic rings of the PET substrate.

A key feature of the active site of *Is*-PETase is the aromatic clamp made up of Tyr58 and Trp156. These residues sit either side of the active site, adjacent to the catalytic Ser131 residue. The side-chains of Tyr58 and Trp156 form π-π stacking interactions with the substrate's terephthalic ring. This fixes in place the adjacent ester bond in the substrate and orients it towards the catalytic Ser131 residue. While the aromatic clamp feature is not uncommon in related cutinases, the replacement of a histidine residue with Ser185 in *Is*-PETase allows the side-chain of Trp156 greater rotational freedom. It has been observed through X-ray crystallography[18,23] and molecular dynamics[21] simulations that Trp156 adopts a different conformation when a PET substrate is bound than when the active site is empty. The ability for Trp156 to exhibit this 'wobbling' behaviour may therefore be crucial for substrate binding[18,23]. It has also been theorised that a further conformational change occurs in Trp156 during the product release step; the side chain rotates to form a weaker T-stacking interaction with the product's aromatic ring reducing the energy barrier for product dissociation.

The active site of *Is*-PETase is lined with lipophilic residues; these form hydrophobic interactions with the largely hydrophobic PET substrate and assist binding. One of these residues is Ile179 which is found in the equivalent position in the TfH enzyme but is not found in many other cutinases. Thr59 is also found at the edge of the active site. It was found that replacement of Thr59 with an alanine residue did not affect the production of MHET, but greatly reduced the amount of TA produced as a product. As of now, there has been no mechanism proposed to explain this result.

Another key difference between *Is*-PETase and other related cutinases is the presence of a novel disulfide bond adjacent to the active site of *is*-PETase. This bond between Cys174 and Cys210 connects the ß7-sheet (which is below the active site and contains the catalytic residue His208) and

a loop connecting the ß8-sheet and the α5-helix. A series of molecular docking studies and molecular dynamics simulations revealed that a three-residue insertion in the α5-ß8 loop creates a more flexible active site. This flexible loop is in turn stabilised by the Cys174-Cys210 disulfide bond. This is supported by work performed on a cutinase found in the bacterium *Fusarium solani* which contains a disulfide bond at a similar position. Removal of this disulfide bond significantly changed the hydrogen bonding network of the active site and resulted in greater fluctuations of atoms in the active site pocket[24]. This shows the importance of these type of proximal disulfide bonds to the activity of their respective enzymes. The Cys174-Cys210 disulfide bond has been removed *in vivo* by creating C174S and C210S mutants; this resulted in the near total loss of the enzyme's activity[25].The flexibility and integrity of its active site provided by this Cys174-Cys210 disulfide is therefore likely to play a crucial role in the binding of a large substrate such as PET[21].

Is-PETase exhibits a significant dipole moment across the entire macromolecule. This dipole moment is caused primarily by the localisation of charged residues at the surface of *Is*-PETase, with most of the positively charged surface residues located in one hemisphere of the enzyme and most of the negatively charged surface residues on the other. The related cutinase TfH was found to have a more even distribution of acidic and basic residues on its surface[18]. So far, the advantage of this whole-enzyme dipole has not been proposed. It is worth noting that the binding pocket is flanked by areas of opposite overall charge while the binding pocket itself has an essentially neutral charge overall. This feature may serve to guide the lipophilic PET substrate to the binding pocket and help prevent it from binding to other areas on the enzyme's surface.

1.1.3 Importance of Thermostability in PETase Enzymes

Thermostability is an important factor in the industrial application of PETases. The rate of PET-degradation in highly crystalline PET has been shown to be dependent on the accessibility of amorphous regions of PET to the PETase. By heating PET to its glass transition temperature (T_g) the proportion of amorphous to crystalline PET is dramatically increased. The T_g of PET is between 76°C and 86°C depending on the age and degradation of the polymer. In an aqueous environment the T_g of PET is lowered to approximately 65°C; this is due to water molecules interfering with the inter-chain interactions. It has been argued that it is desirable to engineer PETases that can function at temperatures above the T_g of PET. These engineered PETases could replace conventional PET recycling methods by their use bioreactors containing waste PET. By heating these bioreactors past the T_g of PET, the efficiency of the PET degradation could be greatly improved. This would require the optimisation of two parameters: the enzymes must have optimal temperatures for their activity (T_{opt}) above T_g and the enzymes must have sufficient thermostability to maintain their structure at these high temperatures.

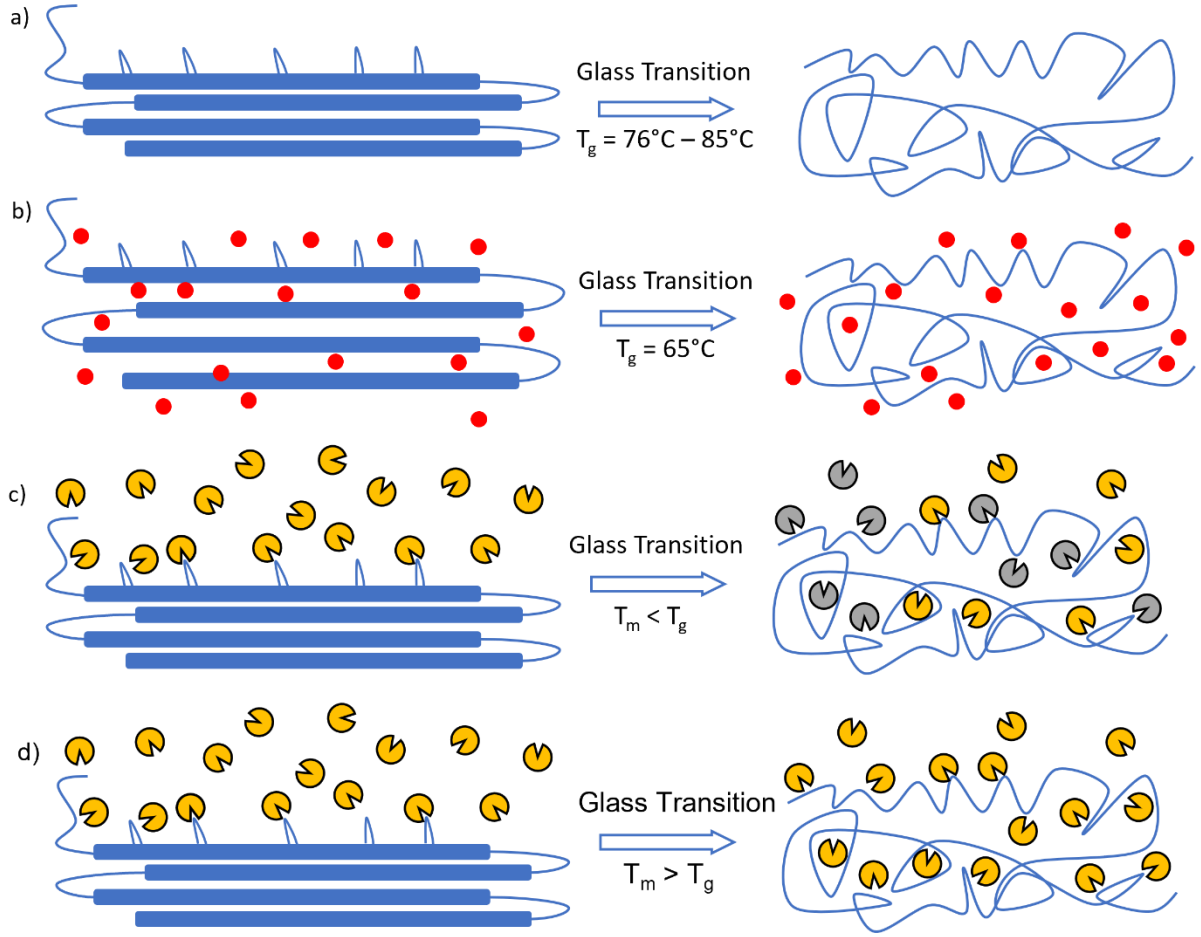


Figure 8: Scheme showing importance of thermostability to PET-degrading enzymes. The glass transition temperature in non-aqueous conditions (a) is significantly higher than in aqueous conditions (b). A PET-degrading enzyme with a T_m value lower than the T_g value of PET will be denatured at the T_g temperature (c). A PET-degrading enzyme with a T_m value above the T_g value of PET will not be denatured at these temperatures. Water molecules are represented by red dots. Active and denatured enzymes are represented by yellow and grey "pacmen" respectively.

The majority of PETases have been discovered in thermophilic microorganisms (such as TfH[26]) or in high temperature environments (such as LCC[15]). PETases with thermophilic host organisms exhibit high T_{opt} and high thermostability making these PETases the most promising candidates for use in high temperature bioreactors.

Unlike TfH and LCC, *Is*-PETase has evolved in a mesophilic environment. Consequently, *Is*-PETase has a low T_{opt} value of only 30 °C and an enzymatic melting point (T_m) of 48.81 °C[27]. It has been argued that *Is*-PETase's intolerance to high temperatures makes it unsuitable for industrial applications[28]. On the other hand, the mesophilic nature of *Is*-PETase makes it far more active at room temperature than its thermophilic counterparts. This means that industrial applications using *Is*-PETase could be performed without requiring additional heating, making them cheaper, more environmentally

friendly, and more scalable. The low T_{opt} of *Is*-PETase also increases its potential to be used in the environment, where it is impossible to supply additional heat.

Based on the results of QM/MM studies[29–33] of the mechanism of *Is*-PETase (discussed later in section 3.1) the degradation of PET to TA and EG is exothermic. This means that the reaction itself will provide heat to a bioreactor. Even without external heating it is possible that the bioreactor could reach temperatures exceeding the T_{opt} of *Is*-PETase. This makes the enhancement of *Is*-PETase's thermostability desirable, even for non-heated bioreactors.

Even when the low T_{opt} of *Is*-PETase is considered to be desirable, increasing its thermostability has been shown to be important. An increase in *Is*-PETase's thermostability has been achieved by introducing an additional hydrogen bonding interaction at the exterior of the protein; this was achieved through the S121E mutation. The new Glu121 forms a hydrogen bond with the residue Asn172. This modification increased the T_m by 8.81 °C and resulted in a 14-fold increase in activity at 40°C over a 72 h period[27]. Another study introduced a novel disulfide bond to connecting two exterior loops of the enzyme, resulting in an increase in T_m by 21.3 °C and a 7-fold increase in activity at 30°C over a 72 h period. Continuous UV-absorbance measurements taken throughout this rection suggest that increased thermostability extends the life-time of each enzyme, even at mesophilic temperatures.[34]

1.1.4 Successful Modifications to Is-PETase

As *Is*-PETase is a remarkably young enzyme (as its substrate PET has only been introduced to the environment in the last century) it is understandable that it is not fully optimised for PET degradation. Over time we could expect *Is*-PETase efficiency to improve through natural selection, if provided with an ample supply of PET for its host organism *I. sakaiensis* to feed on. Unfortunately, the catalytic efficiency of *Is*-PETase is not currently enough for it to be useful in industrial processes and we do not have the time to wait for this to be improved by natural selection. For this reason, several mutagenesis studies have been conducted to improve the catalytic abilities of the enzyme. These can be split into two categories; active site modification that improves substrate binding and modifications that improve the enzyme's thermostability.

1.1.4.1 Initial Single-point Mutations in Is-PETase Resulting in Increased Activity

The catalytic abilities of *Is*-PETase and its mutants have been assessed using several different methods. The most common measurement is through concentration of the products MHET and TA released over time. This is done by taking measurements at set points throughout the reaction using liquid chromatography[17,23,27]. Another common assay technique is to measure the weight-loss of the initial substrate at the end of the reaction[25]. A less common method has been to measure the

crystallinity loss of the substate through scanning electron microscopy[18]. All these methods are non-continuous measurements and give little kinetic information about the enzyme's activity over time. A novel UV-absorbance method has been developed that provides the ability to measure PET-degradation activity in a continuous manner. This technique provides additional information regarding enzyme stability and is useful for the development of more active enzyme mutants. It has been observed however that these measurements are only pseudo-quantitative due to the overlap in UV absorbance of the products MHET and TA. The substrate itself varies from study to study; plastic drinking bottles[20], PET films of various thicknesses[25,27], biaxially oriented PET films[35] and PET pellets[34]. Due to this high variance in measurement techniques across studies it is difficult to compare their results directly. There are several cases (described below) where different studies have made the same single-point mutation but have wildly different results when it came to the mutant's activity. Due to this high variance in measurement techniques, this field of research is in dire need of standardisation. My recommendation would be to use a combination of continuous UV-absorbance measurements and non-continuous liquid chromatography measurements. These techniques complement each other as the former produces detailed kinetic data while the latter is highly quantitative.

Single-point mutations to *Is*-PETase have been made to explore the role of key residues in the active site. Generally, a reduction in PET degradation activity will accompany these mutations, confirming the importance of the original residue to the activity of the enzyme. In some cases, residues that have been thought to play important roles in substrate binding have been substituted for a different residue resulting in an increase in activity. One such case is the replacement of Tyr58 with a small, non-aromatic alanine residue; this resulted in a 3.1-fold increase in activity verses the wild-type against a drinking bottle substrate. This is a surprising result as Tyr58 is part of the aromatic clamp (see section 2.1.2). An explanation for this increase in activity could be that the Y58A substitution widened the active site which facilitated greater binding of the highly crystalline PET chains in plastic bottles[20]. Our knowledge of the role of Tyr58 is further complicated by other studies which observed reduced activity against PET films in the Y58A mutant[25,36]. Discrepancies such as this further highlight the need for standardised assays in this field. A similar pattern of results can be found with the W130A mutant; a 1.3-fold increase in activity was observed against the plastic bottle substrate[20]

while activity was greatly reduced against PET film substrates[25,36].

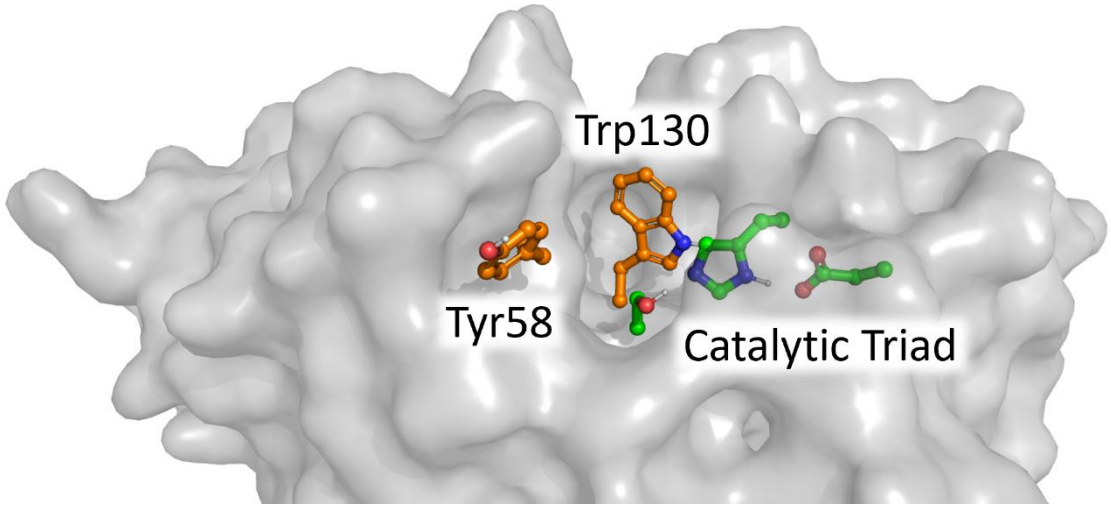


Figure 9: The location of residues Tyr58 and Trp130, relative to the catalytic triad in Is-PETase. The mutation of these residues to alanine resulted in a marginal increase in activity.

Modifications to increase the hydrophobicity or aromaticity of the binding pocket of *Is*-PETase have been made. The enhanced hydrophobicity is thought to increase the binding pocket's affinity towards the lipophilic chains of PET while the increased aromaticity increases the likelihood of favourable π-stacking interactions with the TA rings of PET. These effects combined reduce the energy requirement for substrate binding, hence increasing the reaction rate. Examples of these modifications are: R61A[35], L88F[35], I179F[35], A180I[20], S185H[20,23] and S209F[18] (see Figure 10). These mutations have resulted in modest improvements in activity with increases of up to 2.5-fold versus wild-type.

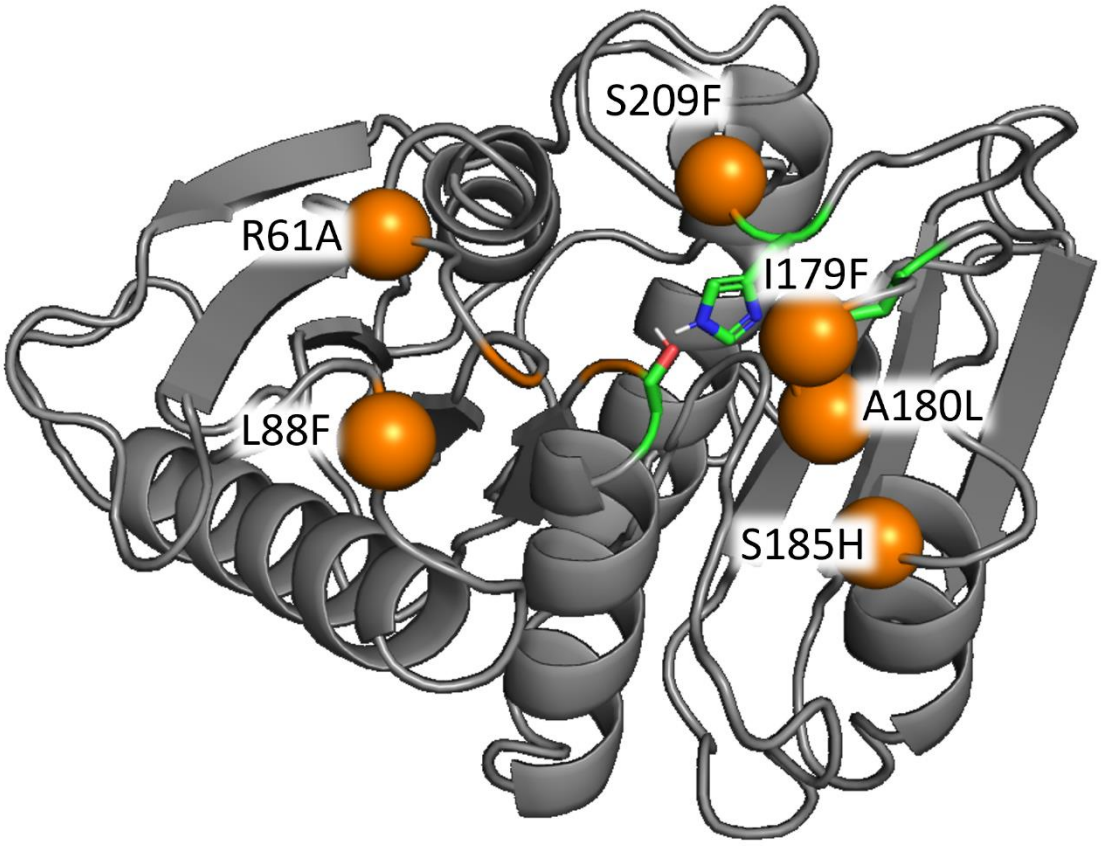


Figure 10: Location of early mutations made to Is-PETase that resulted in marginal increases in activity.

1.1.4.2 Rational design through results in improved Is-PETase variant

In enzyme engineering it is common for successful single-point mutations to be combined in the same protein, producing an enzyme with multiple single-point mutations. Additional successful mutations can then be added to this enzyme in a process that mimics natural selection. The resulting enzyme from several rounds of this process can have a wide array of mutations and greatly improved activities. An *Is*-PETase mutant with five substitutions was produced: S92E D157H N204C R251A S253C. The R251A mutation was introduced as it has found to increase enzymatic activity in previous studies. The S92E D157H mutations introduced a novel hydrogen bonding interaction between the new residues while the N204C S253C mutations introduced an additional disulfide bond between the two new residues. These new intramolecular interactions served to increase the enzyme's thermostability, resulting in an approximately 5-fold increase in activity. The triple mutant S92E D157H R251A (dubbed TS-PETase) was produced resulting in an approximately 5-fold increase in activity versus wild-type against a PET pellet[34]. The R251A mutation was included as previous studies had found that it was beneficial to the enzyme's activity. The S92E D157H substitution was made in an attempt to introduce a novel hydrogen bond between the two residues; this was done to

improve the enzyme's thermostability. It was found through X-ray crystallography that S92E formed a water mediated hydrogen bond with Asn172 instead. This accidental hydrogen bonding interaction had the intended effect, increasing the enzyme's activity by up to 13.9-fold. This highlights the difficulty in engineering stabilising interactions into enzymes using rational design techniques alone.

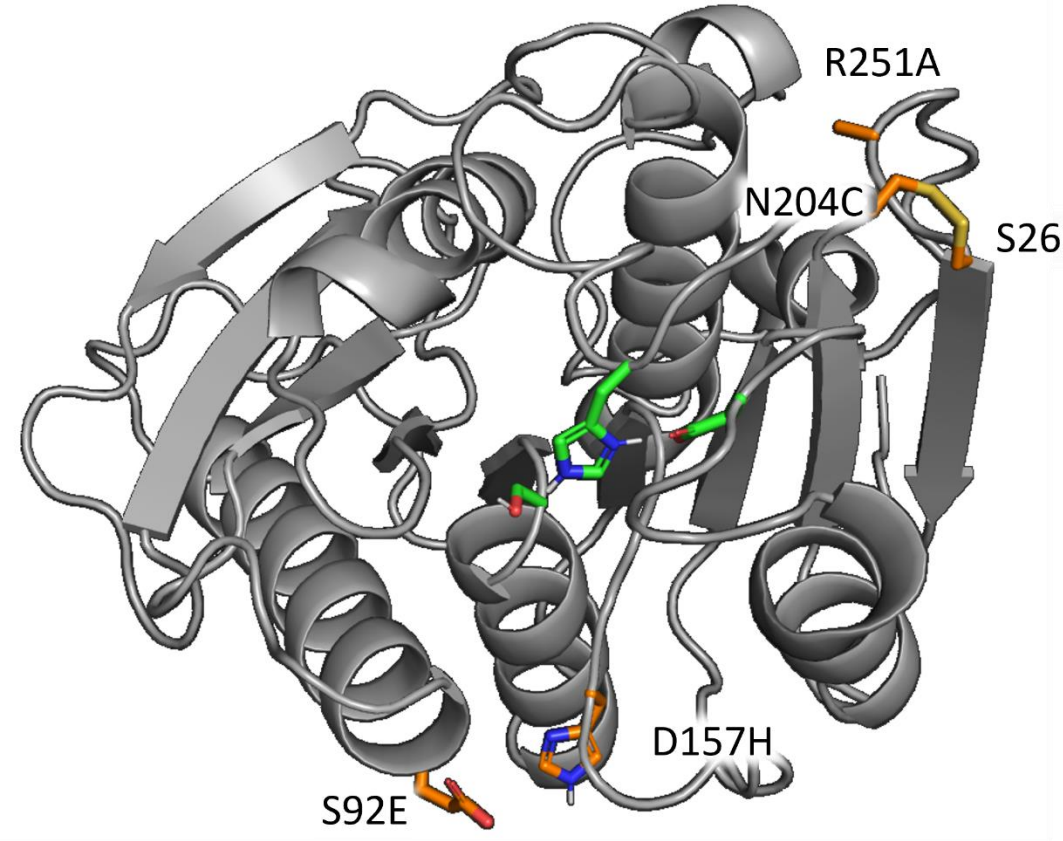


Figure 11: Location of mutations in TS-PETase, introduced through rational design methods. The mutant residues are displayed in orange, while the catalytic triad of the enzyme is displayed in green.

1.1.4.3 Computationally Designed Variant of Is-PETase Displays Improved Activity

Recently, machine learning methods were applied to design mutants of *Is*-PETase with greatly improved thermostability[37]. A three-dimensional, self-supervised, convolutional neural-network was trained using over 19,000 X-ray crystal structures to predict the extent to which a residue is optimal (in terms of forming stabilising interactions) for its local environment. This network was used to predict single-point mutations to *Is*-PETase that would result in optimised interactions with the residue's local environment. This optimisation usually is in the form of an additional hydrogen bond or the removal of steric clashes with adjacent residues. Structures generated by the machine learning model were ranked by a predicted fold change score; those with the lowest predicted fold change were expressed experimentally and assayed for their PET-degradation abilities. A combinatorial approach was then applied whereby the most successful single-point mutations were combined to make double, triple, and quadruple mutants. The most successful mutant enzyme from this process had the following mutations: S92E D157H R195Q N204K R251A. Interestingly, this method has produced similar mutations to the methods previously discussed, with the R195Q and N204K mutations being novel. The activity of this new variant (named FAST-PETase) was assessed at various temperatures with its highest activity at 50°C, releasing 38.8 mM of PET monomers in 96 h. Under the same conditions the S92E D157H R251A mutant produced approximately 10 mM of PET monomers in 96 h. The viability of the use of PETases in industrial applications was demonstrated by scaling up the reaction to degrade 6.4 g piece of untreated PET. After a 96 h incubation period using FAST-PETase, near-total depolymerisation of the PET was observed. The same process was then performed upon a post-consumer plastic bottle; it was found that due to the higher crystallinity of the bottle FAST-PETase was unable to fully degrade it, particularly in the neck regions of the bottle. By a simple thermal pre-treatment, the crystallinity of the bottle was greatly reduced. FAST-PETase was able to efficiently degrade this pre-treated bottle. This shows that through a combination of pre-treatment and enzymatic degradation, it is possible to effectively depolymerise PET.[37]

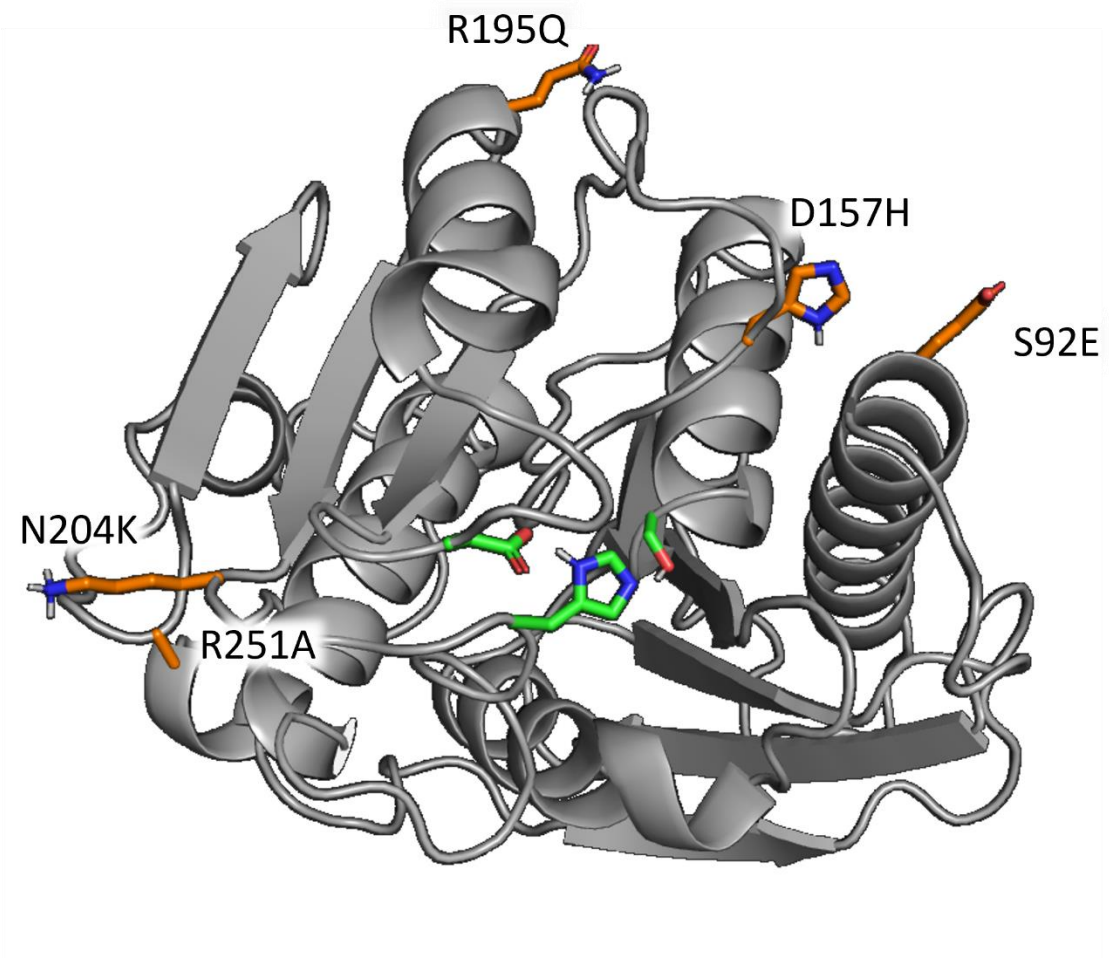


Figure 12: Location of mutant residues in the computationally designed FAST-PETase (shown in orange). The catalytic triad of FAST-PETase is shown in green.

1.1.4.4 Directed Evolution Produces Highly Thermostable and Efficient Variant of Is-PETase

Directed evolution is an enzyme engineering technique that has had an astounding amount of success in the past couple of decades. Its importance was recognised by the awarding of the Nobel Prize for Chemistry to Frances Arnold in 2018[38]. This technique aims to mimic the natural selection process over a greatly accelerated timeframe. Control of the selection pressures applied is used to create enzymes with desired properties such as enhanced catalytic rates, altered or improved substrate specificities or increased tolerance to heat or organic solvents.

The directed evolution approach has been applied to the task of improving the thermostability of *Is*-PETase with characteristically impressive results[39]. A variant (dubbed HOT-PETase) with 21 single-point mutations (see Figure 13) relative to wild-type *Is*-PETase (WT-PETase) was produced. HOT-PETases exhibits a T_m value of 82.5°C, compared to the T_m value of 56.8°C WT-PETase. This allows HOT-PETase to function at temperatures above the T_g of PET, improving the overall rate of PET degradation (for a fuller explanation of this effect, see Chapter 1.1.3 and Figure 8). The improved thermostability of HOT-PETase led to greatly increased rates relative to WT-PETase. The relationship between catalytic rate and reaction temperature was found to be non-linear for HOT-PETase, with the optimum reaction temperature being 65°C. At this temperature HOT-PETase had a catalytic rate (measured as production of soluble product per hour) roughly four-fold higher than the enzyme TS-PETase[27] at its optimum temperature of 40°C, which was measured to have a 13.9-fold increase in activity relative to WT-PETase[27]. The catalytic rate of the PET-degrading cutinase LCC[40] was measured under the same conditions. It was found that HOT-PETase had a roughly 4.5-fold higher catalytic rate at 65°C relative to LCC at 65°C (which is also LCC's optimum reaction temperature). Interestingly though, LCC performed favourably against HOT-PETase over a long assay (50 h), due to its higher stability (see Figure 13).

The generation of HOT-PETase started with the mutant enzyme TS-PETase[27], which has the stabilising mutations S121E, D186H and R280A. Four rounds of directed evolution were performed, each showing a significant increase in activity and thermostability relative to the last (see Figure 13). In each round of directed evolution, a library of mutant enzymes was produced through saturation mutagenesis, with target residues being selected through a combination of online protein stability-enhancing tools and human expertise. These mutant enzymes were expressed in *E. coli*. The resulting cell lysate was then assayed for activity against an amorphous PET film with a crystallinity of 6.7%. Mutant enzymes which showed increased activity were then selected as starting enzymes for subsequent rounds of directed evolution. In between rounds two and three of the directed evolution process, a disulfide bond was introduced through the mutations N233C and S282C. The rationale behind the introduction of this disulfide bond was that it is present in the homologous PET-

degrading cutinase LCC (discussed in Chapter 1.1.2). This novel disulfide bond was responsible for an increase of 5°C in the enzyme's T_m value.

At time of writing, HOT-PETase remains the *Is*-PETase variant with the greatest thermostability and potentially the highest catalytic rate (although differences in measurement technique and reaction conditions preclude direct comparison). This not only shows the power of the directed evolution process but highlights the utility of combing rational design elements into the directed evolution process.

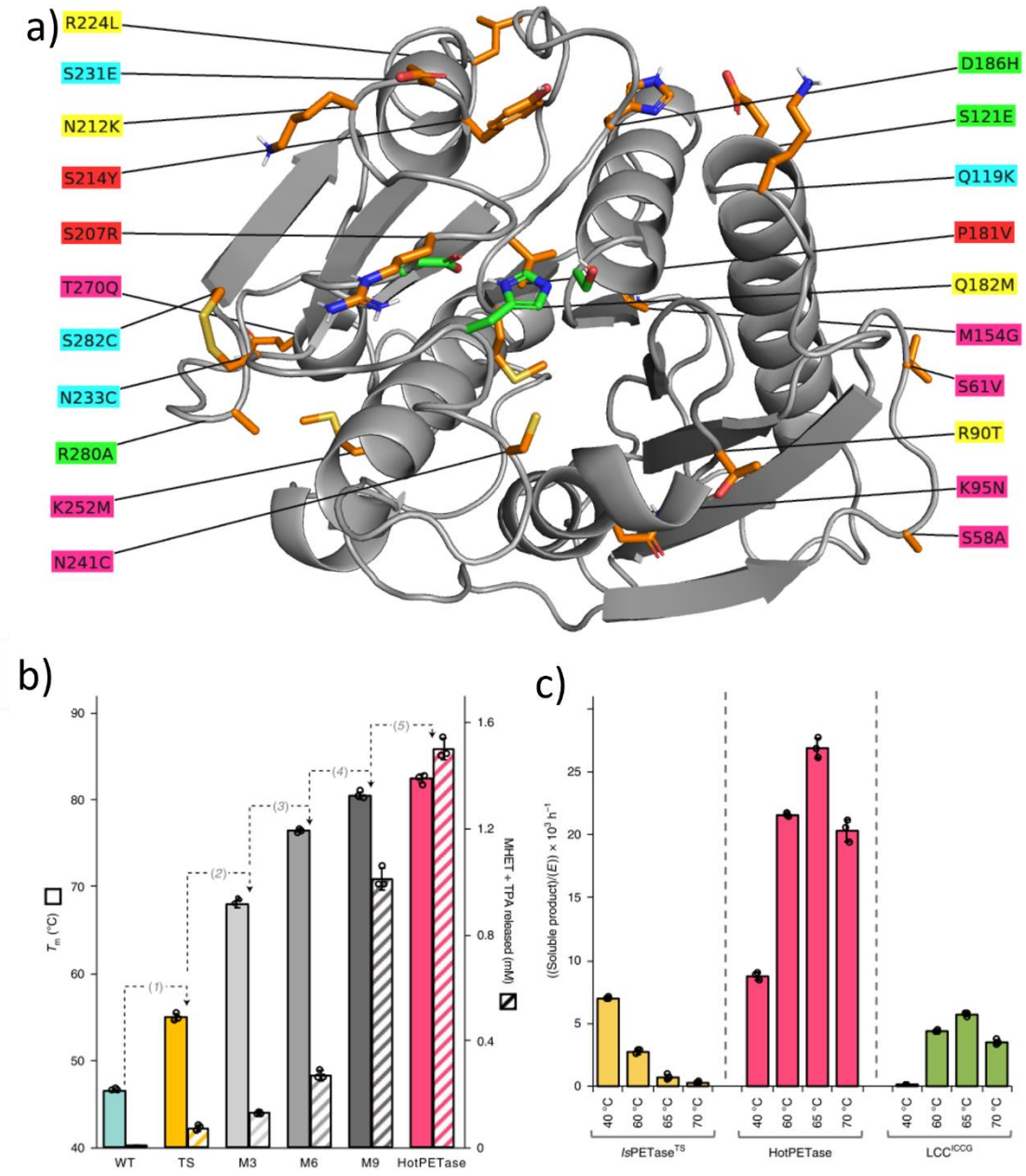


Figure 13: a): The locations of mutant residues (orange sticks) in HOT-PETase relative to the catalytic triad (green sticks). The label colours denote the round of directed evolution that the mutation was introduced: mutations (relative to wild-type Is-PETase) present in the starting enzyme TS-PETase (green labels) and those generated in directed evolution rounds 1,2,3 and 4 are shown with green, red, cyan, yellow, and pink labels respectively. Note that the numbering of these residues is different from the numbering shown in the rest of this thesis (residue numbers are 29 higher). b): The T_m (solid bars) and concentration of monomers produced in assay for WT-PETase, TS-PETase and the most successful variant from each round of directed evolution (x-axis labels show number of mutations introduced during each round of directed evolution). c): The catalytic rate of WT-PETase, HOT-PETase and the cutinase LCC at various reaction temperatures.

1.1.4.5 Summary of Is-PETase mutants and a note on assay standardisation

Across the studies discussed in the previous sections, a wide array of mutants of *Is*-PETase have been produced. From the results of these studies, it can be observed that mutations that focus upon increasing the thermal stability of *Is*-PETase have had a greater effect on the overall catalytic efficiency of the enzyme than mutations that directly modify the enzyme's active site.

As can be seen in Table 1, a plethora of different assay methods have been used to determine the catalytic rate of *Is*-PETase mutants. As discussed later in section 3.3.4, determining kinetic parameters is difficult. Instead, previous studies have focused on the rate-improvements of their mutants relative to wild-type *Is*-PETase. The diversity of methods applied makes comparison between the result of different studies very difficult. To further overall field of *Is*-PETase research, we believe it is important that the assay to determine catalytic ability of *Is*-PETase variants be standardised. We suggest that this standardised assay be conducted using liquid-chromatography methods, as they produce highly quantitative results. A common pH and reaction temperature should also be used. The most common pH used in previous studies has been pH 9; we suggest this be used in all further studies. Given the importance of thermostability to *Is*-PETase's catalytic efficiency (discussed in section 1.1.3), we recommend that assays be performed at both 40°C (the T_{opt} for wild-type *Is*-PETase) and 65°C (the T_m for PET under aqueous conditions). Regarding the type of PET substrate that is used in this standardised assay, we have no strong feelings as to what would be best, so long as it is consistent across different studies.

Table 1: A comparison of mutants of Is-PETase that have been created in previous studies. Their rate-improvements relative to wild-type Is-PETase as well as key methodological features of the assays used in each study have been included.

Mutations	Improvement in Activity	Conditions	Measurement	Substrate	Ref
Y58A	3.1 (48 h)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[23]
	0.4 (18 h), 0.5 (36 h)	30°C, pH 9	Liquid Chromatography	Commercial PET film	[27]
	(4 h) (TA production) 0.8 (42 h) (MHET production)	30°C, pH 9	Liquid Chromatography	PET film	[25]
	~4 (20 h) (TA production) ~2.8 (20 h) (MHET production)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[20]
R61A	1.4 (48h)	30°C, pH 8.5	Weight loss	Biaxial PET film	[35]
L88F	2.1 (48h)	30°C, pH 8.5	Weight loss	Biaxial PET film	[35]
W130A	1.3 (48h)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[23]
	0.2(18h) 0.2(36h)	30°C, pH 9	Liquid Chromatography	Commercial PET film	[36]
	0.1(42 h) (TA production) 0.2 (42 h) (MHET production)	30°C, pH 9	Liquid Chromatography	PET film	[25]

Table 1 continued

Mutations	Improvement in Activity	Conditions	Measurement	Substrate	Ref
I179F	2.5	30°C, pH 8.5	Weight loss	Biaxial PET film	[35]
A180I	1.3	30°C, pH 9	Liquid Chromatography	Drinking bottle	[23]
S185H	1.9	30°C, pH 9	Liquid Chromatography	Drinking bottle	[23]
	~2.5 (20 h) (TA production) ~1.5 (20 h) (MHET production)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[20]
R251A	1.2 (18 h) 1.3 (36 h)	30°C, pH 9	Liquid Chromatography	Commercial PET film	[36]
W130H	~4 (20 h) (TA production) ~1.7 (20 h) (MHET production)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[20]
A180I	~2 (20 h) (TA production) ~1 (20 h) (MHET production)	30°C, pH 9	Liquid Chromatography	Drinking bottle	[20]
S92E D157H R251A	~5 (72 h)	30°C, pH 9	UV absorbance	PET pellet	[34]
S92E D157H N204C R251A S253C	5-7 (24 h, 72 h)	30°C, pH 9	UV absorbance	PET pellet	[34]

Table 1 continued

Mutations	Improvement in Activity	Conditions	Measurement	Substrate	Ref
TS-PETase	13.9 (72 h)	40°C, pH 9	Liquid Chromatography	PET film	[27]
	5.2 (72 h)	30°C, pH 9	Liquid Chromatography	PET film	[27]
S92D D157H R251A	10.8 (72 h)	40°C, pH 9	Liquid Chromatography	PET film	[27]
	5.0 (72 h)	30°C, pH 9	Liquid Chromatography	PET film	[27]
FAST-PETase	142 (24 h)	50°C, pH 8	Liquid Chromatography	Plastic Container	[37]
HOT-PETase	~95 (50 h)	65°C, pH 8	Liquid Chromatography	Amorphous PET film	[39]

1.1.5 Potential Application of PETase Enzymes

The mere existence of PET degrading enzymes less than a century since the invention of PET itself would be enough to warrant its extensive study. Particular interest in PETases has been paid due to the industrial processes that they could be used in. PET is the most extensively used polyester in the world with approximately 19.1 million tonnes produced in 2018 alone[7]. By replacing our current chemical-based industrial processes for the processing and recycling of PET with more efficient and less energy demanding enzyme-catalysed reactions enormous gains will be made both financially and environmentally. There are two main industrial applications for PETases being explored: Surface modification of PET fibres for improved material properties in the textile industry (discussed in detail in Chapter 1.1.5.2) and full degradation of PET to facilitate recycling (discussed in detail in Chapter 1.1.5.1). Several enzymes capable of performing the former have been discovered, while enzymes capable of the latter have proven to be rarer.

1.1.5.1 PET recycling

The most widely discussed use for PET-degrading enzymes is in context of industrial PET recycling. As discussed in Chapter 1:, discarded PET poses a serious threat to the biodiversity of the oceans and our own health. The current effort to recycle PET is currently woeful with a global recycling rate of

just 10%. This poor rate of recycling is due to several factors, the main problem being collection of post-consumer PET products: The majority of plastic waste is disposed of incorrectly and does not reach the intended recycling plants. Instead the waste plastic is either incinerated (24%) or placed in landfill or the natural environment (58-62%)[41]. Unfortunately, this issue cannot be solved enzymatically and requires a combination of social change and new infrastructure to solve.

The primary industrial process for PET recycling is called mechanical recycling and has the following steps: sorting, shredding, washing, and reprocessing[42] (see Figure 14) . This process is highly energy intensive, especially during the agglomeration and extrusion stages in the reprocessing step where the washed PET flakes are converted to molten plastic through heating to >200°C. PETase enzymes function at far lower temperatures with 30°C-70°C being optimum for their activity. By replacing the agglomeration and extrusion steps with enzymatic degradation, the energy cost of PET recycling would be reduced dramatically. This would make the process cheaper and more environmentally friendly.

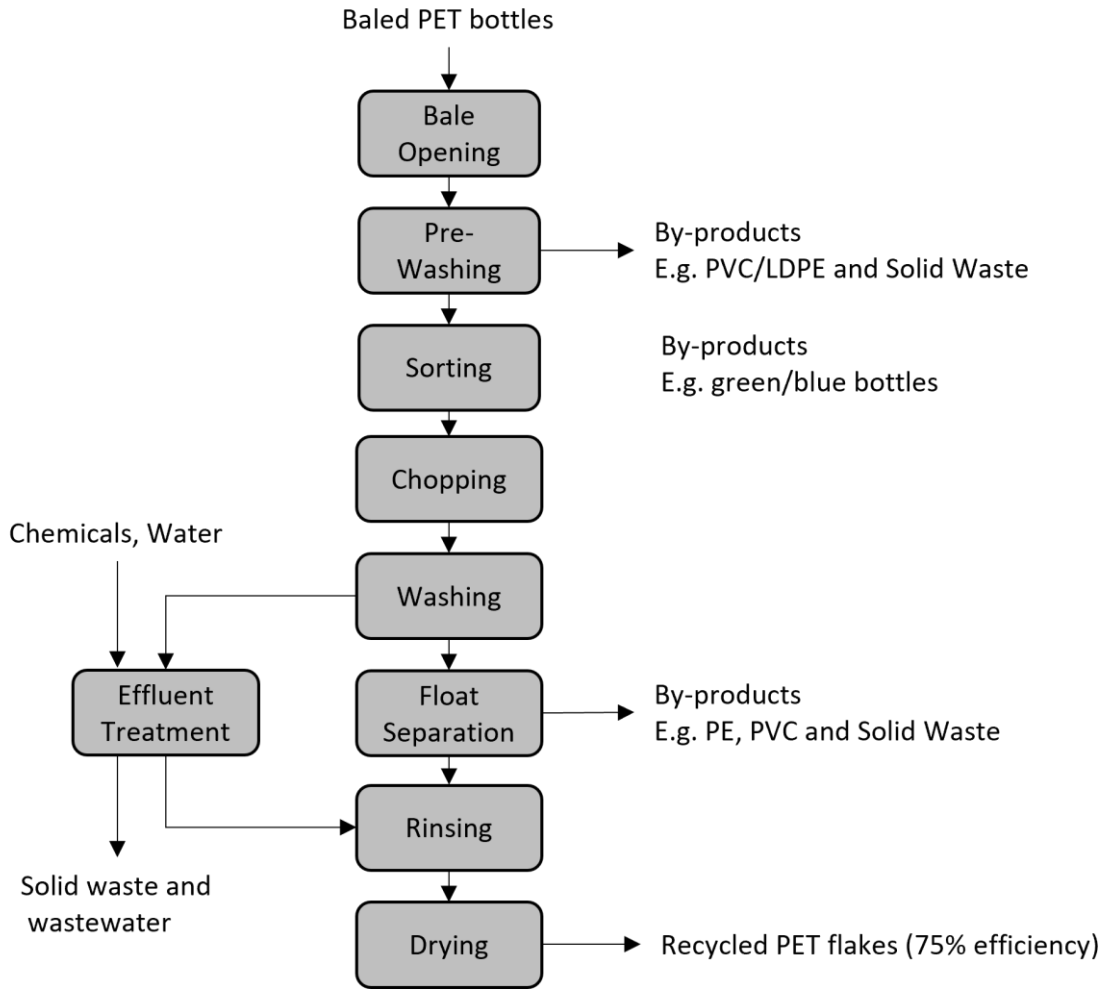


Figure 14: Scheme showing the mechanical recycling process of baled PET water into PET flakes.

A major issue that occurs during the sorting and washing steps in PET recycling is contamination: Compounds from bottle-caps, labels, glue, food, and metals that have passed through sorting processes[42] can reduce the material properties of the recycled PET (rPET) produced and can leach into foodstuffs. For this reason, the European Union prohibits the use of rPET in food packaging. Use of enzymatic degradation processes could provide a solution to this problem. Due to the substrate specificity of PETases only PET will be broken down into its monomers TA and EG, leaving any contaminants unchanged. Use of PETases would also alleviate the need for the washing step which uses either hot water or sodium hydroxide to remove oils from the surface of the PET flakes. This would further reduce the cost and environmental impact of the PET recycling process.

An alternative to mechanical recycling of PET is chemical recycling where different chemolysis reactions are used to break PET into a small-molecule feedstocks for the chemicals industry. Acid or base catalysed hydrolysis is the only reaction used that regenerates TA and EG monomers which can

then be used to resynthesise PET. This process is not widely used however as it produces TA with poor purity, reducing the usefulness of the product. The oldest and most common chemolysis reaction used in industry is glycolysis; this produces a mixture of MHET and polyols. The MHET can be reprocessed into polymeric PET while the polyols are used in the synthesis of various materials such as polyurethanes. The glycolysis process is unfortunately rather slow and can only be used for relatively pure feedstocks, making it unsuitable for recycling of consumer-grade PET. Reactions such as PET methanolysis and PET aminolysis afford products that cannot be repolymerised into PET. While these products are useful in the production a wide array of chemicals and materials, this does not achieve the highly desirable circular life-cycle of PET as a material. All these chemolysis reactions are performed at high temperatures and often require high pressure, making them highly energy intensive. Chemolysis of PET also requires the use of harsh chemicals, the production and disposal of which can be costly and environmentally damaging. As with mechanical recycling, replacement of chemical PET recycling with an enzymatic would greatly reduce the process' environmental impact and could even improve its efficiency.

The feasibility of enzymatic PET recycling has been demonstrated by the creation of lab-scale reactors (see Figure 15). PET degradation has been achieved in a one litre tank using the cutinase found in the thermophilic bacterium *Thermomonospora curvata*, achieving degradation of above 50% over a period of 24 h at 70°C[43]. Reaction rates after this 24 h period dramatically dropped; this is thought to be due to thermal denaturation of the enzyme. To make this a viable industrial process the effective lifetime of the enzyme needs to be improved by either increasing its thermostability through protein engineering or by finding an alternative enzyme with high thermostability and a lower optimal temperature for its activity. Once a suitable enzyme has been produced, a scaled-up version of this type of reactor could provide a viable alternative to mechanical or chemical degradation of PET during the recycling process. The resultant TA from these reactors can be recovered by acid precipitation and crystallisation[40]; this could then be repolymerised with EG (which can be bio-derived) to produce virgin PET. This would create a circular life-cycle for PET products where the material can be recycled indefinitely without any loss of material properties or the threat of contamination.

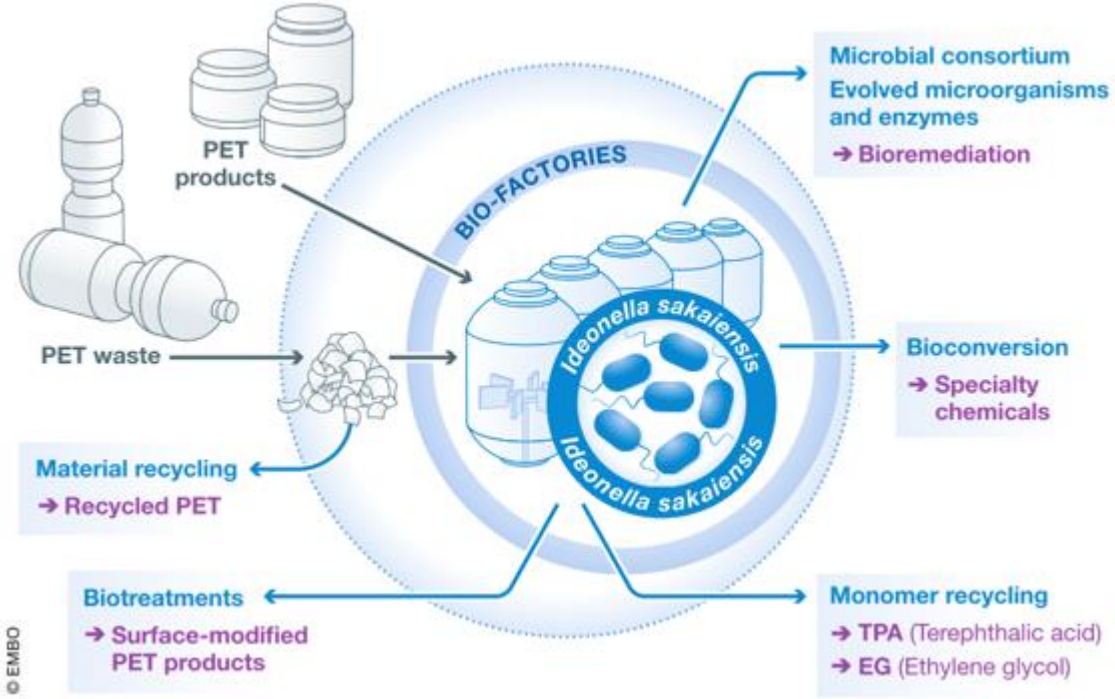


Figure 15: Scheme showing potential utility of a bioreactor-based method for recycling PET. Figure reprinted from Taniguchi et.al [44].

1.1.5.2 Fabric Treatment

Polyester (which is made from PET) is by far the most common fabric used globally making up 52% of global production of fibres in 2020. The demand for polyester fibre is also the fastest growing in the textile industry[45]. The popularity of this material can be explained as it is durable, lightweight, and inexpensive. The fibres also have high chemical resistance and low abrasion and shrinking properties[46]. The hydrophobicity of PET fibres makes them excellent for waterproof garments. To use PET fibres to produce more comfortable clothing or for improved dyability with cationic or reactive water-soluble dyes their hydrophilicity must be increased. This is achieved by hydrolysing surface-exposed amorphous regions of the PET fibres; this exposes a series of charged carboxylic acid and polar hydroxyl functional groups which have more favourable interactions with water than non-hydrolysed PET chains (see Figure 16). Ideally this surface hydrolysis would occur whilst leaving the core region of PET chains intact as this would increase the fabrics hydrophilicity without reducing its strength. Modern chemical treatments use either mild acids or bases at high temperatures to perform this hydrolysis reaction, which can be enhanced through hydrogen peroxide catalysis[47]. The use of these harsh chemicals results in partial hydrolysis in the PET core of the fibres which reduces the strength of the resulting fabric and causes weight loss of the overall fabric.

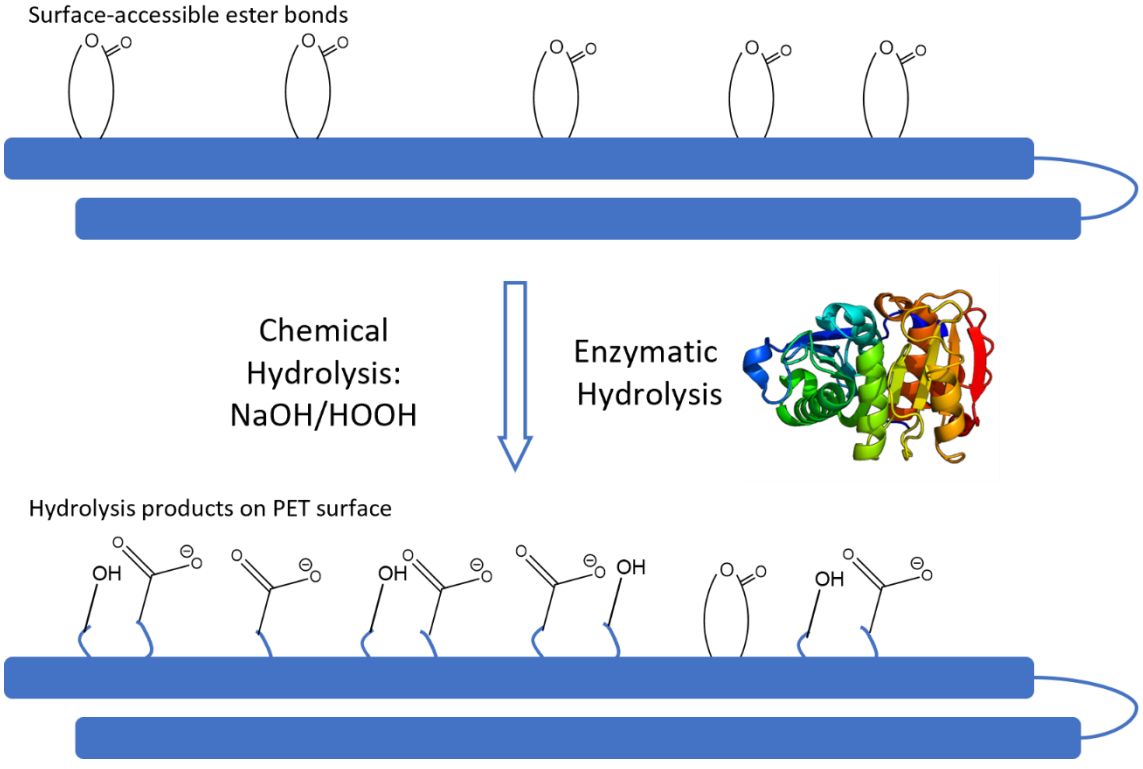


Figure 16: Scheme showing surface-hydrolysis treatment performed upon PET fibres. This can be performed through chemical or enzymatic processes.

PETases provide an enzymatic alternative to chemical surface hydrolysis of PET fibres. Their utility has been demonstrated using PETases from *Thermobifida fusca* (TfH) and *Fusarium solani*: treatment

of knitted PET swatches with these enzymes under mild conditions (50°C and pH 7.2) resulted in increases in hydrophilicity comparable to treatment with 2 molar sodium hydroxide at the same temperature. The mild conditions used in the enzymatic tests are unlikely to have damaged the polyester fibres; however this was not investigated in this study. Replacement of our current industrial processes for PET surface treatment with an enzymatic method would therefore eliminate the need for the use of harsh chemicals. This would make the process more environmentally friendly as their production and disposal pose risks to the natural environment.

1.1.6 Layout of this Thesis

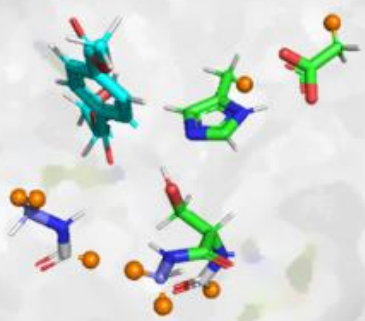
In the remainder of this thesis, we will discuss computational approaches that we have applied to the exploration of the catalytic mechanism and thermostability of the enzyme *Is*-PETase and its variants. In Chapter 2: we will discuss the theoretical background underpinning the computational techniques that we have used in the remaining chapters. In Chapter 3: we will discuss our application of primarily QM/MM techniques to the exploration of the catalytic mechanism of wild-type *Is*-PETase against poly(ethylene) terephthalate. The results from this chapter have been published in *Chemistry A European Journal*[48]. In Chapter 4:, we discuss our application of the same QM/MM methods to explore the catalytic mechanism of our own *Is*-PETase variant, DISU-PETase. Chapter 3: also discusses our use of molecular dynamics simulations to probe the thermostability of the most thermally stable (and catalytically active) *Is*-PETase variants[37,39] from the literature. Chapter 5: discusses our exploration of the catalytic mechanism of wild-type *Is*-PETase against an alternate substrate, poly(ethylene) furanoate, using the same QM/MM techniques described in Chapter 3:

Chapter 2: Theoretical Background

$$E_{torsions} = \sum_{torsions} \sum_{n=1}^{6} \frac{1}{2} V_n [1 + \cos(n\phi - \gamma_n)] .$$

$$E_{bonds} = \sum_{bonds} K_b (r - r_{eq})^2 .$$

$$E_{angles} = \sum_{angles} K_a (\theta - \theta_{eq})^2 .$$



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

2.1 Calculating Energies of Chemical Systems Ab Initio

Much of the following discussion has been adapted from the textbook "Introduction to Computational Chemistry" by Frank Jensen[49].

2.1.1 Hartree-Fock Methods

2.1.1.1 The Schrödinger Equation

The particles of interest (electrons and nuclei) in a chemical system are considered small (in comparison to for example, macromolecules or nanoparticles) and slow moving (in comparison to the speed of light). Thus, the Schrödinger Equation (which accounts for electrostatic effects while neglecting gravitational forces) is used with a non-relativistic treatment (as the masses of particles in chemical systems are virtually unaffected by their velocities) to calculate the system's properties. The motion of particles in a chemical system is therefore best described using the time-dependent non-relativistic Schrödinger Equation:

$$ih\frac{\delta}{\delta t}|\Psi(\mathrm{r},\mathrm{t})\rangle = \hat{H}|\Psi(r,t)\rangle.$$

Equation 1: The time-dependent, non-relativistic Schrödinger equation.

In this equation, $\hat{H}$ is the Hamiltonian or Hamiltonian Operator and Ψ is the wavefunction dependent on the position of particles (denoted by r) and time (denoted by t). A spin component may be introduced into the r term; this is described in more detail in section 2.1.1.3. Using the Born-Oppenheimer approximation (see Equation 10) of electrons affected by stationary nuclei, the Schrödinger equation may be made time-independent:

$$\hat{\mathrm{H}}|\Psi(r)\rangle = E|\Psi(r)\rangle.$$

Equation 2: The time-independent, non-relativistic Schrödinger equation.

Strictly speaking, time-dependent phenomena such as interactions with fluctuating fields (e.g., light) are not removed using the Born-Oppenheimer approximation. For such cases, the so-called time-independent Schrödinger Equation is not appropriate. The energy of the system can thus be calculated as the expectation value of the Hamiltonian.

$$E = \langle\Psi|\hat{H}|\Psi\rangle.$$

Equation 3: Energy expressed as an expectation value of the Hamiltonian operator.

This requires knowledge of the system's wavefunction and a correct formulation of the Hamiltonian. The rest of this chapter will be an explanation of the approximations required to approach the solution to the Schrödinger Equation.

The Molecular Hamiltonian

The Hamiltonian Operator $\hat{\mathrm{H}}$ can be decomposed to:

$$\hat{\mathrm{H}}_{tot} = T(r) + V(r).$$

Equation 4: The Hamiltonian operator expressed as the sum of kinetic energy and potential energy.

In Equation 4, $T(r)$ is the kinetic energy operator and represents the total kinetic energy of the particles in the system. $V(r)$ is the potential energy operator and represents the potential energy of the particles interacting with each other. In a chemical system where the particles of interest are electrons and nuclei, Equation 4 can be further decomposed to the molecular Hamiltonian:

$$\hat{\mathrm{H}}_{tot} = T_n + T_e + V_{nn} + V_{ee} + V_{ne}.$$

Equation 5: The molecular Hamiltonian.

In Equation 5, T_n and T_e represents the kinetic energy of the system's nuclei and electrons respectively. V_{nn} represents the potential energy of nuclei-nuclei interactions. V_{ee} represents the potential energy of electron-electron interactions and V_{ne} represents the potential energy of interactions between nuclei and electrons. Fully expanded, the Molecular Hamiltonian can be expressed in atomic units:

$$\hat{\mathrm{H}}_{tot} = -\sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 + \sum_{A=1}^{M} \sum_{B=A+1}^{M} \frac{Z_A Z_B}{R_{AB}} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} - \sum_{A=1}^{M} \sum_{i=1}^{N} \frac{Z_A}{r_{Ai}}.$$

Equation 6: The molecular Hamiltonian expanded and expressed in atomic units.

In Equation 6, the index M represents the total number of nuclei with indices A and B. Meanwhile, the index N represents the total number of electrons with indices i and j. The nuclear charges of nuclei A and B are represented by Z_A and Z_B while R_{AB}, r_{ij} and r_{Ai} represent inter-nuclear, inter-electronic and nuclei-electron distances respectively. The mass of nuclei A is represented as M_A. The Laplace Operator ∇_i^2 is expressed as the sum of the second derivatives of the position of particle i:

$$\nabla_i^2 = \frac{\delta^2}{\delta {x_i}^2} + \frac{\delta^2}{\delta {y_i}^2} + \frac{\delta^2}{\delta {z_i}^2}$$

Equation 7: The Laplace operator.

In Equation 7 x_i, y_i and z_i represents the Cartesian coordinates of particle i. It is worth noting that in Equation 6, term V_{ne} is preceded by a negative sign as this is an attractive interaction and results in negative energy values. Terms for V_{nn} and V_{ee} are preceded by positive signs as these are repulsive interactions and result in positive energy values.

Using the Molecular Hamiltonian as expressed in Equation 6 the Schrödinger Equation cannot be solved exactly. This is because each of the potential energy terms encounters the many-body problem, making the difficulty of solving such a problem scale exponentially with the number of particles.

2.1.1.2 The Born-Oppenheimer and Adiabatic Approximations

The Hamiltonian (see Equation 5) can be re-written as a centre of mass system:

$$\widehat{H}_{tot} = T_n + \widehat{H}_e + \widehat{H}_{mp}.$$

Equation 8: The Mass-centred Hamiltonian.

In Equation 8, T_n represents the kinetic energy of nuclei, $\widehat{H}_e$ is the electronic Hamiltonian and $\widehat{H}_{mp}$ is a mass polarisation term.

In Equation 8, $\widehat{H}_e$ is the electronic Hamiltonian:

$$\hat{\mathrm{H}}_e = T_e + V_{ee} + V_{ne} + V_{nn}.$$

Equation 9: The electronic Hamiltonian after the Born-Oppenheimer approximation.

The Born-Oppenheimer Approximation uses the relative masses of nuclei and electrons (a nucleus is approximately 1800-fold more massive than an electron in the hydrogen atom) to make the assumption that from the perspective of an electron's movement, nuclei remain stationary. Conversely, with respect to the motion of nuclei the motion of electrons is considered instantaneous under the Born-Oppenheimer Approximation. In the Molecular Hamiltonian, the Born-Oppenheimer Approximation allows for the assumption that $T_n = 0$ and allows for the V_{nn} term in Equation 5 to be calculated as a constant (this constant can have no effect upon the electronic wavefunction so is often omitted).

The result of this approximation is that the nuclei of a system are represented as moving upon a potential energy surface (PES), the topology of which is calculated by solving the electronic Hamiltonian:

$$\hat{\mathrm{H}}_e = T_e + V_{ee} + V_{ne}.$$

Equation 10: The electronic Hamiltonian.

After the Born-Oppenheimer approximation, the Hamiltonian Operator can be expressed as a sum of the electronic Hamiltonian (Equation 10) and the mass-polarisation term $\widehat{H}_{mp}$ introduced in Equation 8. Under the adiabatic approximation, the mass-polarisation term (which contains information on nuclear motions) is neglected. For non-adiabatic calculations such as frequency calculations, this approximation is inappropriate.

This simplified Hamiltonian can be expressed as an eigenvalue equation:

$$\hat{\mathrm{H}}_e \Psi_e = E_e \Psi_e.$$

Equation 11: The eigenvalue equation for the electronic Hamiltonian.

In Equation 11, the subscript e denotes the electronic Hamiltonian, wavefunction and energy. The Born-Oppenheimer Approximation removes one many-body term (V_{nn}) from the Molecular Hamiltonian, however one many-body term (V_{ee}) remains. As such, further approximations are required to accurately approximate the Schrödinger Equation.

2.1.1.3 The orbital approximation and spin integration

The concept of molecular orbitals is used to circumvent the many-body property of the V_{ee}term in Equation 10. Each electron in a molecule is assigned its own wavefunction (otherwise known as a molecular orbital). The full electronic wavefunction (Ψ_e in Equation 11) is constructed as a combination of molecular orbitals (described later in Equation 19). Each molecular orbital is constructed as a linear combination of other functions:

$$\psi_i = \sum_j c_{ij} \phi_{j.}$$

Equation 12: Construction of a molecular orbital.

Where ψ_i represents the molecular orbital for electron i, constructed as a linear combination of functions ϕ_j and coefficients c_{ij}. So far, orbitals have been described as functions of three spatial components (x, y, and z). In order to account for the spin component of electrons, the concept of spin orbitals is introduced:

$$\chi(x_1) = \psi(\mathrm{r}_1)\sigma(\omega_1)$$

Equation 13: The formula for a spin orbital.

In Equation 13, the spin orbital is constructed via the multiplication of a spatial orbital $\psi(\mathrm{r}_1)$ with a spin function $\sigma(\omega_1)$. Since electrons are fermions, the ω_1 value exists in either spin-up (α) or spin-down (β) states.

2.1.1.4 Hartree Products, the Antisymmetry Principle and Slater Determinants

If, as a thought experiment, it is assumed that the repulsive term in Equation 7 V_{ee} = 0 it is possible to express the electronic Hamiltonian as a sum of one-electron integrals:

$$\hat{\mathrm{H}} = \sum_{i=1}^{N} \hat{h}(i).$$

Equation 14: The electronic Hamiltonian expressed as a sum of one-electron integrals.

$$\hat{h}(i) = \frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{ia}}.$$

Equation 15: the form of a one electron integral.

The energy of an individual spin-orbital can thus be calculated as an eigenvalue in the equation:

$$\hat{h}(i)\chi_i(x_i) = \epsilon\chi_i(x_i).$$

Equation 16: The energy of a spin-orbital expressed as an eigenvalue.

The wavefunction can therefore be written as a Hartree product of each of the spin-orbitals:

$$\Psi = \prod_{i=1}^{N} \chi(x_i) = \chi_1(x_1)\chi_2(x_2) \dots \chi_N(x_N).$$

Equation 17: The wavefunction of a system expressed as a Hartree product of its spin-orbitals.

Hartree products treat electrons as distinguishable (i.e., we know which electron occupies which orbital) and do not account for electron spin. According to the Antisymmetry Principle, the sign of the wavefunction must change (from positive to negative or vice-versa) when the occupancies of two electrons are switched. This behaviour results from the electron being a fermion. Hartree Products do not satisfy this principle, instead an antisymmetric linear combination of the spin orbitals is used. For a two-electron system this is written as:

$$\Psi(1,2) = \frac{1}{\sqrt{2}}[\Psi_1(1)\Psi_2(2) - \Psi_1(2)\Psi_2(1)].$$

Equation 18: The wavefunction for a two-electron system, written to satisfy the Antisymmetry Principle.

As the number of electrons in a system increases, Equation 18 becomes increasingly tedious to write out (as the number of terms rises steeply with the number of electrons). As such, Slater Determinants are used to express a wavefunction as an antisymmetric linear combination of N spin

orbitals:

$$\Psi(1,2,\dots,\mathrm{N}) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix}.$$

Equation 19: The Slater determinant used to express the combination of antisymmetric spin-orbitals.

Slater Determinants satisfy the Antisymmetry principle for any number of electrons. By using Slater Determinants, the inter-electron interaction term is calculated as a sum of each individual electron's interactions with the electron density from the remaining spin-orbitals occupied by all the other electrons. This is an approximation (often called the Mean-Field Approximation) and will never give the exact solution to the Schrödinger Equation for systems with more than one electron.

2.1.1.5 Many-electron integrals

In Equation 14, the core Hamiltonian is expressed. This can be used to calculate the energy of a system where the inter-electron interactions are neglected. In order to calculate the energy of a 'real' system with interacting electrons, the V_{ee} term in Equation 5 must be evaluated. This interaction is evaluated as the interaction of one molecular orbital belonging to an individual electron with the molecular orbitals of all other electrons in the system. These interactions are calculated one at a time, as such, V_{ee} is considered a two-electron term. The energy of this two-electron interaction E_2 is calculated as:

$$E_2 = \frac{1}{2}\sum_{ij}^{Nelec}(J_{ij} - K_{ij})$$

Equation 20: The energy of two-electron interactions in Hartree-Fock Theory.

In Equation 20, the inter-electron energy is calculated as a sum of the difference between Coulomb integrals (J_{ij}) and Exchange integrals (K_{ij}) for all electron pairs. The $\frac{1}{2}$ prefactor in Equation 20 accounts for the double-interaction which is a result of the sum running over all electrons. The Coulomb integral is analogous to the inter-electron repulsion term described in Equation 6. J_{ij} is expressed as:

$$J_{ij} = \iint \chi_i(1)\,\chi_i(1)^* \frac{1}{r_{ij}} \chi_j(2)\chi_j(2)^*\, dr_1 dr_2$$

Equation 21: The Coulomb integral.

In Equation 21, J_{ij} is calculated as the energy value produced by the interaction of electrons 1 and 2 in their respective spin orbitals χ_i and χ_j. The reciprocal distance operator $\frac{1}{r_{ij}}$ is analogous to the classical V_{ee} operator in Equation 6.

The Exchange integral K_{ij} has no classical analogue and is expressed as:

$$K_{ij} = \iint \chi_i(1)\, \chi_i(2)^* \frac{1}{r_{ij}} \chi_j(2) \chi_j(1)^*\, dr_1 dr_2$$

Equation 22: The Exchange integral.

Equation 22 is the same as the Coulomb integral (Equation 21) but with the occupancies of electrons 1 and 2 exchanged.

At this point, we can express the total energy of a system as:

$$E_{HF} = \sum_{i}^{N} h_i + \frac{1}{2} \left(J_{ij} - K_{ij}\right) + V_{nn}$$

Equation 23: The Hartree-Fock energy.

In Equation 23, the total energy is expressed as a combination of a sum of all one electron integrals h_i (see Equation 15), Coulombic (see Equation 21) and Exchange integrals (see Equation 22) and the inter-nuclear interaction energy (which is a constant). The ½ prefactor to the $J_{ij} - K_{ij}$ terms is used to remove double-interactions that arise from summing over all electrons interacting with all other electrons in the system.

2.1.1.6 The Variational Principle and the Self-Consistent-Field Process

To find the ground state energy of a system, the wavefunction (which is constructed through a linear combination of basis sets) for that system must be optimised through an iterative process called the self-consistent-field (SCF) process. Small modifications to the coefficients that describe the contribution of each basis function are made, which results in a small change in energy. The Variational Principle states that the 'true' wavefunction for a system, when operated upon by the Hamiltonian will produce the lowest possible energy as its expectation value. This can be expressed as:

$$\frac{\langle \Psi_{trial} | \widehat{H} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} = E_{trial} \geq E_0.$$

Equation 24: The Variational principle.

Thus, any trial wavefunction must have a higher or equal associated energy value than the 'true' wavefunction. The SCF process therefore becomes a series of iterative changes to the wavefunction to minimise the system's calculated energy.

2.1.2 Density Functional Theory

Density Functional Theory (DFT) provides an alternative method to Hartree-Fock methods (discussed in the previous section). As the name suggests, DFT relies on the concept of electron density ρ to calculate properties of a chemical system. The electron density is defined as the square of the absolute value of the wavefunction:

$$\rho(r) = |\Psi|^2.$$

Equation 25: Electron density in relation to the wavefunction.

Electron density is an expression of probability of finding an electron at an infinitesimally small volume of space. As the exact location of an electron cannot be identified exactly, electron density describes the location of electrons as a 'smear' of locations that they are likely to be found in.

2.1.2.1 The Hohenberg-Kohn Theorems

The foundations of DFT are the Hohenberg-Kohn (HK) theorems[50]. The first HK theorem is the Existence Theorem. This states that there is a unique one-to-one mapping between the ground state electron density of a system and its energy. The HK Existence Theorem was proven using a *reductio ad absurdum* method which for brevity will not be recounted in this thesis. The upshot of this theorem is that a 'perfect' functional (a function of functions) of electron density was proposed that would return the exact energy of the system in question. This would, in effect, solve the Schrodinger Equation without having to resort to Hartree-Fock (HF) theory (discussed in the previous section). This is an attractive proposal as it would greatly reduce the dimensionality of the problem as electron density is dependent on 3 spatial dimensions, while HF methods are dependent on 4N dimensions (three spatial and one spin component per electron). This would greatly reduce the computational cost, although as we see in later sections, the re-introduction of terms from HF theory into DFT methods limits these savings.

The second of the HK theorems is the Variational Theorem. This states that a functional of electron density will produce its lowest energy value only when operating upon the exact ground-state density of the system in question. This is the same as saying that the Variational Principal discussed in section 2.1.1.6 can be applied to DFT. In practice this allows for the iterative optimisation of the electron density to find the exact (or close to exact) ground state density.

2.1.2.2 Kohn-Sham Theory

An initial formulation of the 'perfect' HK functional was proposed by Thomas, Fermi, and Dirac[51–53]:

$$E_{TFD}[\rho] = T_{TF}[\rho] + J[\rho] + K_D[\rho] + E_{ne}[\rho].$$

Equation 26: The Thomas-Fermi-Dirac density functional.

In Equation 26, the density functional is separated into the following functionals: electronic kinetic energy ($T_{TF}[\rho]$), electronic Coulombic repulsion ($J[\rho]$), electronic Exchange energy ($K_D[\rho]$), and nuclear-electron attraction ($E_{ne}[\rho]$). This approach treats electron density ρ as a single quantity, providing the attractive reduction in dimensionality discussed previously. Unfortunately, the kinetic energy functional $T_{TF}[\rho]$ in Equation 26 tends to create large inaccuracies in predicted energies to the extent that bonding between atoms is not predicted using these equations. Clearly this is wholly inadequate to computational chemists.

An alternative formulation of the HK functional was proposed by Kohn and Sham[54]. Using Kohn-Sham (KS) theory the concept of molecular orbitals is re-introduced into DFT, with each molecular orbital describing the electron density of one electron in the system:

$$\rho = \sum_{i=1}^{\infty} n_i |\varphi_i|^2.$$

Equation 27: Construction of electron density from KS molecular orbitals.

In Equation 27, the full electron density of a system is constructed via the linear combination of the squares of the absolute values of molecular orbitals, multiplied by occupancy cofactors. This equation is summed up to infinity, which is required to the exact description of the total electron density. In practice, it is sufficient to sum up to the number of electrons in the system.

The concept of adiabatic connection is central to KS theory. A hypothetical system with non-interacting electrons is constructed (i.e., $V_{ee} = 0$). The inter-electron interaction is then gradually 'switched-on' until a fully interacting system is reached. Mathematically, this is expressed using the letter λ, where $\lambda = 0$ in the non-interacting system and increases to $\lambda = 1$ for the 'real' fully interacting system. Values of λ between 0 and 1 therefore represent imaginary systems where the inter-electron interaction is partially 'switched on'.

In KS theory, the density functional is reformulated as:

$$E_{KS} = T_{KS} + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho].$$

Equation 28: The Kohn-Sham density functional.

In Equation 28, the energy in the KS functional E_{KS} is calculated as a sum of electronic kinetic energy T_{KS} and functionals for nuclear-electronic interactions $E_{ne}[\rho]$ and inter-electronic Coulombic interactions $J[\rho]$ as well as an additional functional for exchange-correlational energies $E_{xc}[\rho]$. Note that the electronic kinetic energy term in Equation 28 is no longer a functional of electron density, instead it is calculated simply using a Slater Determinant:

$$T_{SD} = \sum_{i=1}^{N_{elec}} \left\langle \varphi_i \middle| -\frac{1}{2}\nabla^2 \middle| \varphi_i \right\rangle.$$

Equation 29: The electronic kinetic energy calculated using a Slater Determinant in KS theory.

In Equation 29, the electronic kinetic energy (T_e) is calculated as a sum of expectation values of the Laplace operator (see Equation 7) operating upon each KS molecular orbital. This formulation of T_e does not take into account effects from inter-electronic interactions (in this case $\lambda = 0$). In KS theory, the inter-electronic contribution to T_e is instead incorporated into the $E_{xc}[\rho]$ in Equation 28. The functionals $E_{ne}[\rho]$ and $J[\rho]$ are formulated using equations borrowed from HF theory with the nuclear-electronic interaction calculated as:

$$E_{ne}[\rho] = -\sum_{i}^{N} \int \sum_{A}^{M} \frac{Z_A}{r_{iA}} |\varphi_i|^{2.}$$

Equation 30: The nuclear-electronic attraction functional in KS theory.

In Equation 30, the nuclear-electronic functional is constructed as a sum of each KS molecular orbital interacting with each nuclear charge. The variables N and i refer to number and index of electrons, while the variables M, A, and Z_A refer to the number, index and charge of nuclei. The variable r_{iA} refers to the distance between electron i and nucleus A. The Coulombic Operator is constructed using the Coulomb Integral (see Equation 21), operating upon KS molecular orbitals:

$$J[\rho] = \sum_{ij}^{N} \iint |\varphi_i(r_1)|^2 \frac{1}{r_{ij}} \left|\varphi_j(r_2)\right|^2 dr_1 r_2.$$

Equation 31: The Coulombic Operator

At this point, each of the terms used in HF theory to calculate the energy of a system (see Equation 23) has an analogue in KS theory, except the Exchange integral K_{ij}. In KS theory, the energy that would arise from K_{ij} is included in the $E_{xc}[\rho]$ functional in Equation 28. This exchange-correlational functional contains all energy contributions that arise from inter-electron interactions, except from Coulombic interactions which are contained in $J[\rho]$. As a result, $E_{xc}[\rho]$ is by far the most

computationally expensive term in KS theory to calculate. In the next few sections, a series of methods to approximate $E_{xc}[\rho]$ at greater levels of accuracy (and complexity) are detailed.

2.1.2.3 The Local Density Approximation

The simplest method for approximating the $E_{xc}[\rho]$ functional in Equation 28 is the Local Density Approximation (LDA). LDA methods rely upon the concept of a homogeneous electron gas (HEG). This is an imaginary material where electrons are uniformly distributed about a background of uniformly distributed positive charge. HEG can be treated as having any electron density (it is, after all, imaginary). The exchange-correlational energy of HEG can be calculated exactly using the Quantum Monte-Carlo method. In LDA methods, the exchange-correlational energy of real molecules is calculated by assuming that the electron density locally can be treated as HEG:

$$E_{xc}^{LDA}[\rho] = \int \rho(r)\varepsilon_{xc}\big(\rho(r)\big).$$

Equation 32: The exchange-correlational functional under the Local Density Approximation.

In Equation 32, the $\varepsilon_{xc}\big(\rho(r)\big)$ term refers to the exchange-correlational energy of HEG at the electron density of a single point in space in the real molecule of interest. The LDA approximation provides accurate energies for systems that resemble the imaginary HEG (such as continuous metal surfaces). Unfortunately for the majority of molecular systems, LDA methods underestimate the exchange-correlational energy.

2.1.2.4 Gradient Correction Schemes

To improve the accuracy of LDA methods, a series of corrections can be applied. Methods that do this are grouped together as generalised gradient approximation (GGA) methods. These gradient corrections act upon the exchange-correlational energy produced by LDA methods and can be either additive or multiplicative. It is convenient at this point to separate E_{xc} into exchange E_x and correlational E_c energies:

$$E_{xc} = E_x + E_{c.}$$

Equation 33: Exchange-correlational energy spit into its component parts.

The popular functional PBE[55] (a variant of which is used in this work) applies the following multiplicative correction to calculate exchange energy:

$$E_x^{PBE} = E_x^{LDA}\left[1 + \alpha - \frac{\alpha}{1 - bx^2}\right]$$

$$x = \frac{|\nabla\rho|}{\rho^{4/3}}.$$

Equation 34: The exchange functional in PBE methods.

In Equation 34, the variables α and b are constants which can be used to change the extent of the correction. The variable x is dependant on the spatial components of the electron density. In PBE methods, the correlational energy is computed using an additive correction to the correlational energy calculated using the LDA method:

$$E_c^{PBE} = E_c^{LDA} + F_c^{PBE}(t)$$
$$F_c^{PBE}(t) \propto \ln\left[1 + dt^2\left(\frac{1 + At^2}{a + At^2 + A^2t^4}\right)\right].$$

Equation 35: The correlation functional in PBE methods.

In Equation 35, the variables a and A are constants that can be used to modify the extent of the correction. The variable t represents a dimensionless density gradient (not time, as per general convention) and is proportional to x in Equation 34.

2.1.2.5 Hybrid and Double Hybrid Functionals

An inherent problem with LDA and GGA methods is that their E_{xc} functional is evaluated approximately. In HF theory however, the exchange energy can be evaluated exactly using the exchange integral K_{ij} (see Equation 22). In so-called hybrid methods, a portion of the exchange energy is calculated using HF methods (using KS orbitals in place of wavefunctions), while the remainder is calculated using LDA methods. The hybrid functional PBE0[56] is a modification of the previously discussed PBE GGA method. In PBE0 the exchange energy is evaluated as:

$$E_{xc}^{PBE0} = 0.25E_x^{HF} + 0.75E_x^{PBE} + E_c^{PBE.}$$

Equation 36: The exchange-correlational functional in the hybrid functional PBE0.

A further step up "Jacob's Ladder" of density functionals is marked by double-hybrid density functionals. In the same manner as hybrid functionals, double-hybrid functionals calculate exchange energy using a combination of HF and GGA methods (which in turn contain LDA methods). Additionally, double-hybrid functionals use perturbation theory (an extension of HF theory not discussed in this thesis for brevity) in combination with GGA methods to calculate the correlation energy. The double-hybrid functional DSD-PBEP86[57] (used in this work for single-point energy calculations) calculates exchange-correlational energy by:

$$E_{xc}^{PBEP86} = 0.32E_x^{PBE} + 0.68E_x^{HF} + 0.45E_c^{PBE} + 0.51E_o^{MP2} + 0.23E_s^{MP2}.$$

Equation 37: The exchange-correlational functional in the double-hybrid functional DSD-PEB86.

In Equation 37, the terms $0.51E_o^{MP2}$ and $0.23E_s^{MP2}$ refer to the opposite-spin and same-spin correlational energies calculated using the Perturbation theory method MP2. Opposite-spin and same-spin interactions have differing ranges over which they are important, so it is convenient to calculate them separately.

2.1.2.6 Dispersion Corrections

GGA, hybrid, and double-hybrids tend to make systematic errors for interactions between large aliphatic groups and between aromatic rings. These 'soft' interactions are grouped together into the umbrella term dispersion. To account for these systematic errors, dispersion corrections can be applied to any DFT method. These are applied to the overall energy once it has been calculated by the DFT method in question:

$$E_{Tot} = E_{DFT} + E_{disp.}$$

Equation 38: Scheme for applying dispersion corrections to a generic DFT method.

The D3(BJ) dispersion correction[58] (applied to the revPBE and DSD-PBE86 functionals in this work) uses exponents of inter-nuclear distances of r^6 and r^8 to calculate the dispersion energy:

$$E_{disp}^{D3(BJ)} = -\frac{1}{2}\sum_{A\neq B}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n + \left[f_{damp}^{D3(BJ)}\left(R_{BJ}^{AB}\right)\right]^n}.$$

Equation 39: The D3(BJ) dispersion correction.

In Equation 39, the dispersion correction is calculated via pairwise interactions between atoms A and B, with interatomic distances of R_{AB}. The term C_n^{AB} is the isotropic dispersion coefficient and is calculated for all element pairs in the system for $n = 6$ and $n = 8$. The Becke-Johnson dampening function $f_{damp}^{D3(BJ)}\left(R_{BJ}^{AB}\right)$ is applied to modify the output of the overall dispersion correction:

$$f_{damp}^{D3(BJ)}\left(R_{BJ}^{AB}\right) = a_1\sqrt{\frac{C_8^{AB}}{C_6^{AB}}} + a_2.$$

Equation 40: The Becke-Johnson Dampening Function.

In Equation 40, the same isotropic dispersion coefficients used in Equation 39 are used. The variables a_1 and a_2 are constants and may be tuned to affect the extent of the dampening function.

2.1.3 Basis Sets

Both Hartree-Fock and Density Functional Theory require a description of a set of one-electron functions. As previously discussed, the ideal function to describe these one-electron functions would

be a Slater function centred upon each nucleus in the system. Each Slater function can be used to describe the shape of one atomic orbital, this is called a Slater-Type Orbital (STO):

$$\phi_{abc}^{STO}(x,y,z) = Nx^a y^b z^c e^{-\zeta r}.$$

Equation 41: The mathematical description of a Slater-Type Orbital.

The Slater orbital ϕ_{abc}^{STO} is a function of the Cartesian coordinates x, y and z and uses the normalisation constant N. The exponents a, b and c are angular momentum terms such that the angular momentum L=a+b+c. The ζ term describes the diffuseness of the orbital, with smaller ζ values resulting in more diffuse orbitals. Slater-Type Orbitals are costly to compute and do not describe nodal behaviour in atomic orbitals. Gaussian-Type Orbitals (GTO) have become universally used instead of STOs:

$$\phi_{abc}^{GTO}(x,y,z) = Nx^a y^b z^c e^{-\zeta r^2}.$$

Equation 42: The mathematical description of a Gaussian-Type Orbital.

GTOs are convenient for Quantum Chemistry as they follow the Gaussian Product Theorem that states that the product of two Gaussian functions is another Gaussian Function. This makes GTOs significantly faster to calculate than STOs. The GTO on its own provides a less accurate description of an atomic orbital than a STO. For this reason a linear combination of GTOs is used to approximate a single STO:

$$\phi_{abc}^{CGTO}(x,y,z) = N\sum_{i=1}^{n} c_i x^a y^b z^c e^{-\zeta r^2}$$

Equation 43: An atomic orbital described as a linear combination of Gaussian-Type Orbitals.

In Equation 43, a contracted gaussian-type orbital ϕ_{abc}^{CGTO} is constructed from a series of GTOs. The contraction coefficient c_i is used to weigh each 'primitive' Gaussian function. In a contracted basis set, core atomic orbitals have the contraction coefficient frozen at a pre-set value throughout the SCF process as these orbitals are not significantly affected by the surrounding orbitals. In a minimal basis set, one basis function is used per atomic orbital present in the system while in double-zeta basis sets, two are used and in triple-basis sets three are used and so on. In general, the more basis functions per atomic orbital used, the more accurate the calculation at the cost of increased calculation time.

In order to better describe polarisation of molecular orbitals (such as during bonding) polarisation functions are often used. These consist of using orbitals with higher angular momentum. To polarise an orbital with an angular momentum of L, it is common practice to use a polarisation function of

with an angular momentum of $L + 1$ (s-orbitals are polarised by p-orbitals, p-orbitals with d-orbitals and so on).

2.1.4 Transition State Search Using Nudged-Elastic Band Methods

The nudged-elastic band technique is a method widely used to identify saddle-points between local minima[59]. Nudged-elastic band methods generate 'images' that lie in-between two input local minima via interpolation. This produces a 'band' of states that convert one input geometry to the other. The geometry of each 'image' is then optimised separately using a chosen method (e.g., choice of DFT functional). In addition to the energies generated from the geometry optimisation method, an additional nudged-elastic band force is applied to each 'image' as it is being optimised:

$$F_i^{NEB} = F_i^{\perp} + F_i^{S||}.$$

Equation 44: The nudged-elastic band force

In Equation 44, $F_i^{\perp}$ represents the force perpendicular to the band and $F_i^{S||}$ represents a spring force applied along the direction of the band. The , $F_i^{\perp}$ term is constructed as:

$$F_i^{\perp} = -\nabla(\boldsymbol{R_i}) + \nabla(\boldsymbol{R_i}) \times \hat{\boldsymbol{\tau}}_{\boldsymbol{i}}\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}.$$

Equation 45: The perpendicular force of the nudged-elastic band

In Equation 45, $\nabla(\boldsymbol{R_i})$ represents the gradient of the potential energy with respect to the coordinates of image i. In Equation 45, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ represents the tangent vector along the band at image i. The spring force in Equation 44 is constructed as:

$$F_i^{S||} = k(|\boldsymbol{R_{i+1}} - \boldsymbol{R_i}| - |\boldsymbol{R_i} - \boldsymbol{R_{i-1}}|)\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}.$$

Equation 46: The spring constant of the nudged-elastic band method

In Equation 46, k is a unitless force constant that can be used to tune the magnitude of the spring force. The terms $\boldsymbol{R_i}$, $\boldsymbol{R_{i+1}}$, and $\boldsymbol{R_{i+1}}$ represent the coordinates of images at state i, $i - 1$ and i+1 respectively. By combining perpendicular and spring forces, the nudged-elastic band method ensures that adjacent 'images' remain geometrically connected throughout the geometry optimisation procedure. Once the set of 'images' have been optimised a minimum energy pathway between the two input states is produced. In order to find the saddle point of this minimum energy pathway an additional procedure is used: The climbing image method[60] reoptimizes the nudged-elastic band 'image' with the highest energy. In this optimisation, the 'image' does not feel the usual spring force. Instead it feels an inverted form of the perpendicular force described in Equation 45. The force applied during the climbing image optimisation is expressed as:

$$F_{imax}^{CI} = -\nabla(\boldsymbol{R}_{\boldsymbol{i}_{max}}) + 2\nabla(\boldsymbol{R}_{\boldsymbol{i}_{max}}) \times \hat{\boldsymbol{\tau}}_{\boldsymbol{i}_{max}} \hat{\boldsymbol{\tau}}_{\boldsymbol{i}_{max}}.$$

Equation 47: The force applied during a climbing-image step

In Equation 47 $\nabla(\boldsymbol{R}_{\boldsymbol{i}_{max}})$ represents the gradient of the potential energy with respect to the coordinates of $\boldsymbol{i}_{\boldsymbol{max}}$, the 'image' with the highest potential energy. $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}_{max}}$represents the tangent to band at image $\boldsymbol{i}_{\boldsymbol{max}}$. The result of applying Equation 47 to the highest energy 'image' is that the image moves up the potential energy surface along the band, whilst moving down the potential energy surface perpendicular to the band. A fully optimised geometry generated using this method will therefore fit the definition of saddle-point.

2.2 Calculating Energies through Molecular Mechanics

2.2.1 Use of Empirical Forcefields

An alternative to *ab initio* quantum mechanical methods is molecular mechanics (MM). In MM methods, the energy of a system is calculated using an empirical, atomistic, and algebraic function that treats particles using Newtonian physics. These methods have the advantage of being significantly faster than QM methods and are therefore practical for systems with up to a million particles. For this reason, MM methods have become popular for the study of biological macromolecules. The drawbacks of MM methods are that their accuracy does not approach even the simplest QM methods and that their empirical nature creates a reliance upon a large set of input parameters. The energy of a system is calculated as:

$$E_{MM} = E_{bonded} + E_{non-bonded}.$$

Equation 48: The separation of total energy into bonded and non-bonded terms in molecular mechanics.

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions.}$$

Equation 49: The separation of the bonded term in Equation 48 into bonded, angular and torsional energies.

$$E_{non-bonded} = E_{electrostatic} + E_{vdw}.$$

Equation 50: The separation of non-bonded term in Equation 48 into electrostatic and Van der Waals energies.

These set of equations are to say that the full MM energy is calculated as a sum of bonded and non-bonded energies (see Equation 48). The bonded energy term is further calculated as a sum of covalent, angular, and torsional energies (see Equation 49). The non-bonded term in Equation 48 is calculated as a sum of energies of electrostatic and Van der Waals interactions (see Equation 50).

2.2.1.1 Semi-Empirical Treatment of Bonded Terms

The bonded terms in Equation 48 can be separated into 1-2 interactions (bonds), 1-3 interactions (angles) and 1-4 interactions (torsions or dihedral angles). Interactions between particles that are separated by more than three chemical bonds are treated using the non-bonded terms.

The first term in Equation 49, the chemical bonding energy is calculated as a sum of a set of simple spring constant equations according to Hooke's Law:

$$E_{bonds} = \sum_{bonds} K_b(r - r_{eq})^2.$$

Equation 51: Spring constants used to model bonded energy terms in molecular mechanics.

In Equation 51, K_b represents the bond spring constant, r_{eq} represents the equilibrium bond distance and r represents the actual bond distance.

The second term in Equation 49, the energy of bond angles is calculated using another spring constant equation:

$$E_{angles} = \sum_{angles} K_a(\theta - \theta_{eq})^2.$$

Equation 52: Spring constants used to model angular energy terms in molecular mechanics.

In Equation 52, K_a represents the angle spring constant, θ_{eq} represents the equilibrium bond angle and θ represents the actual bond angle. The Hooke's Law equations produce a harmonic potential which differs from the real potential when atoms are very close or very far apart. This creates a limitation for MM methods making them suitable for systems that have bond lengths and bond angles that are close to equilibrium.

The third term in Equation 49 is the torsional energy. This is calculated using a linear combination of cosine functions:

$$E_{torsions} = \sum_{torsions} \sum_{n=1}^{6} \frac{1}{2} V_n[1 + \cos(n\phi - \gamma_n)].$$

Equation 53: Linear combination of cosine functions used to model torsional energy terms in molecular mechanics.

In Equation 53, V_n represents the energy barrier for an n-fold rotation, ϕ represents the actual torsion angle and γ_n represents a phase-offset term.

Out-of-plane terms are often included in the torsional energy term of Equation 49. These terms describe the angle between a pair of planes. Each of these planes is constructed using one central atom bonded to three other atoms. The energy of this term is calculated by:

$$E_{oop} = \sum_{oop} \frac{1}{2} V \left[1 + \cos\left(2\phi - 180^{\circ}\right)\right.$$

Equation 54: Linear combination of cosine functions to model out-of-plane terms in molecular mechanics.

In Equation 54, V represents the rotational energy barrier associated with the out-of-plane angle and ϕ represents the actual out of plane angle. In practice it is unnecessary to include out-of-plane terms for each trivalent or above atom in a system. These terms are mostly included to enforce planarity functional groups such as carbonyls or aromatic groups. Semi-Empirical Treatment of Non-Bonded Terms

The non-bonded term in Equation 48 is separated into Van der Waals interactions and electrostatic interactions. Both interactions are calculated for all pairs of atoms that do not have bonded interactions and are within an arbitrary cut-off distance.

The Van der Waals interactions are calculated as:

$$E_{vdw} = \sum_{i=1}^{N} \sum_{i<j}^{N} \varepsilon_{ij} \left[\left(\frac{r_0}{r_{ij}} \right)^{12} - \left(\frac{r_0}{r_{ij}} \right)^{6} \right].$$

Equation 55: Lennard-Jones style potential used to model Van der Waals terms in molecular mechanics.

In Equation 55, ε_{ij} represents the interaction strength between each atom pair, r_0 represents the Van der Waals radius and r_{ij} represents the actual distance between atoms i and j. Equation 55 has a positive term that scales with r^{-12}; this represents the repulsive Van der Waals interaction between atoms that are close to each other. As functions raised to a higher negative power decay steeply, the repulsive term is very short ranged. Equation 55 contains a negative term which scales with r^{-6}; this represents the attractive forces between atoms due to induced-dipole induced-dipole interactions (dispersion) and is a longer ranged function than the repulsive term.

The electrostatic (or Coulombic) term in Equation 50 is calculated as:

$$E_{elec} = \sum_{i=1}^{N} \sum_{i<j}^{N} \frac{q_i q_j}{r_{ij}}.$$

Equation 56: The modelling of Coulombic interactions in molecular mechanics.

In Equation 56, q_i and q_j represent the charges of atoms i and j and r_{ij} represents the distance between the two atoms. If the two atoms in this term are of opposing charge this term is negative (representing a Coulombic attraction) while atoms of the same charge produce a positive term (representing a Coulombic repulsion). The electrostatic term scales with r^{-1}, so is considered a long-range interaction.

2.2.1.2 Use of Parameters in Molecular Mechanical Methods

In the previous sections, the equations for the energy terms used in MM methods have been described, each of which requires a set of constants to be evaluated. To calculate the energy of a system using MM methods the following parameters are required for each appropriate set of atoms: k_b and r_{eq} for 1-2 interactions, k_a and $\theta_{eq}$for 1-3 interactions, V_n and $\gamma_n$for 1-4 interactions ε_{ij}, r_0 and each atom's charge q for all non-bonding interactions. Parameter sets are used to provide values for these input constants. Values for each of the above parameters have been calculated at high accuracy using QM methods for a wide array of atom groups. These parameters are then collected into a large dataset and used as an input for all MM calculations.

In most parameter sets the concept of atom types is introduced. This makes the approximation that atoms within a functional group behave the same as atoms within the same functional group, but in a different molecule. This allows each atom in a system to be assigned an atom type. Once atom types have been assigned, the number of required parameters can be reduced significantly.

2.2.1.3 Molecular Dynamics

Molecular Dynamics uses a Molecular Mechanical forcefield to simulate the evolution of a system through time. The MM forcefield can be considered a time-independent potential energy function of the system's Cartesian coordinates. This is used to calculate the change in each particle's momentum through time via:

$$\frac{\partial p_{x_i}}{\partial t} = -\frac{\partial V}{\partial x_i}.$$

Equation 57: The scheme used to update a particle's momentum in molecular dynamics simulations.

In Equation 57, $\frac{\partial p_{x_i}}{\partial t}$ represents the partial derivative of momentum in each one Cartesian direction over time and $\frac{\partial V}{\partial x_i}$ represents the partial derivative of the system's potential energy. Equation 57 is

evaluated three times for each atom in the system, once for each Cartesian direction of each atom. This is achieved by assigning initial Cartesian coordinates and velocities for each atom in the system, then updating the positions and velocities of each atom at a series of discrete timesteps:

$$x_{i+1} = x_i + v_i \Delta t.$$

Equation 58: The scheme used to update a particle's Cartesian coordinates in molecular dynamic simulations.

$$v_{i+1} = v_i - \Delta \mathrm{V}(x_i) \Delta t.$$

Equation 59: The scheme used to update a particle's velocity in molecular dynamics simulations.

In Equation 58, x_{i+1} represents the position of a particle after timestep Δt, x_i represents the position of a particle before this timestep, v_i represents the velocity of a particle before this timestep and m represents a particle's mass. In Equation 59, v_{i+1} represents the velocity of a particle after a timestep Δt. v_i and x_i represent the velocity and position of a particle before this timestep and $\Delta \mathrm{V}$ represents the change in potential energy. For each timestep in Equation 58 and Equation 59 are solved for each particle in the system, creating a time-evolved simulation of motion.

2.2.1.4 Ensembles

When we use MD methods to simulate a system, we must select a manner in which the system interacts with the surrounding 'environment' (i.e., anything outside of the system in question). Under the so-called canonical ensemble (otherwise known as the NVT ensemble), the volume of the system is set to be constant. The boundaries of the system are made impenetrable to particles which keeps the number of particles in the system constant. To keep the temperature of the system constant, it is immersed in an imaginary 'heat bath' of constant (and adjustable) temperature of such a size where temperature exchange with the system has a negligible effect upon the temperature of the 'heat bath'. Under the canonical ensemble, the energy and pressure of the system are allowed to fluctuate.

In the same manner as the canonical ensemble, the isothermal-isobaric ensemble (otherwise known as the NpT) ensemble uses the same 'heat bath' and impenetrable boundary methods to keep the number of particles and temperature constant. Instead of holding volume constant, the NpT ensemble holds pressure constant and allows energy and volume of the system to fluctuate throughout the simulation. Since the majority of chemical reactions are performed under an environment with constant pressure (for example a round-bottomed flask at ambient pressure), the NpT ensemble is most commonly used to model chemical and biological systems.

2.3 Hybrid Methods to Calculate Energies of Macromolecules

2.3.1 The Concept of Hybrid Quantum-Mechanical/ Molecular Mechanical Methods

The study of macromolecules such as proteins presents us with a problem: the computational cost of quantum mechanics (QM) methods increases rapidly as its system's size increases, making a full QM treatment unfeasible for macromolecules. Whilst MM methods can be used to simulate large systems, they are incapable of modelling bond-making and bond-breaking events as they depend on a set of known parameters that govern bonding and non-bonding interactions (often called forcefields). This problem is solved by introducing hybrid quantum mechanical/ molecular mechanical (QM/MM) methods[61–65]: the region of the system in which the interesting chemistry is occurring is treated using the more accurate QM method, while the remainder of the system is treated with the computationally inexpensive MM method. The success of this approach was recognized through the Nobel Prize in Chemistry being awarded to Warshel, Levitt, and Karplus "for development of multiscale models for complex chemical systems"[61].

Using the QM/MM methodology, the energy of a system is calculated as:

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM}.$$

Equation 60: The separation of energy in QM/MM calculations into QM, MM and QM/MM interaction energies.

In Equation 60, E_{QM} is the energy of the QM region, which is calculated by solving the Schrödinger equation either using *ab-initio* methods or Density Functional Theory. E_{MM} is the energy of the MM region, calculated using forcefields. $E_{QM/MM}$ represents the interaction energy between the QM and MM regions. This term presents the greatest challenge to evaluate and will be discussed in detail in the proceeding sections.

2.3.2 Treatment of the QM/MM Boundary

2.3.2.1 Polarisation Methods

This section discusses the evaluation of the $E_{QM/MM}$ term from Equation 60 in a system where the QM and MM region share no covalent bonds. The simplest treatment of the QM/MM boundary is to use Unpolarised Boundary Conditions. In this method the energies of the QM and MM regions are computed then the electrostatic and Coulombic interactions between the QM and MM regions are calculated and added as a correction:

$$E_{QM/MM} = \sum_{i}^{QM}\sum_{j}^{MM}\frac{\alpha q_i q_j}{r_{ij}} + 4\varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right).$$

Equation 61: The scheme for QM/MM interaction energy under Unpolarised Boundary Conditions.

In Equation 61, the first term represents the electrostatic interaction, and the second term is a Lennard-Jones potential. The term α is a polarisation term that accounts for the overall charge of the QM region. Terms q_i and q_j represent the point charges of atoms in the QM and MM regions respectively. It is worth noting that point charges for the QM region must be generated from the electron density. The term r_{ij} represents the distance between particles i and j. The terms ε_{ij} represents the interaction strength of the non-bonding interaction and σ_{ij} represents the distance between the atoms i and j where the energy of interaction is minimal.

A more sophisticated (and accurate) method to calculate the interaction energy between the QM and MM regions is the polarised/unpolarised method. In this method the QM region is allowed to be polarised by the surrounding MM region while the MM region remains unpolarised by the QM region. This is achieved by separating Equation 60 into electrons and nuclei in the QM region interacting with point charges in the MM region:

$$E_{QM/MM} = \sum_{i}^{\substack{QM\\electrons}}\sum_{m}^{\substack{MM\\atoms}}\frac{q_m}{r_{im}} + \sum_{n}^{\substack{QM\\nuclei}}\sum_{m}^{\substack{MM\\atoms}}\frac{Z_m q_n}{r_{nm}} + 4\varepsilon_{ij}\left(\frac{\sigma_{nm}^{12}}{r_{nm}^{12}} - \frac{\sigma_{nm}^{6}}{r_{nm}^{6}}\right).$$

Equation 62: The scheme for QM/MM interaction energy under Polarised/Unpolarised Boundary Conditions.

To use a fully polarised scheme where the QM and MM regions are polarised by each other, a polarisable forcefield must be used. In these polarisable forcefields an additional polarizability tensor term α is required to induce a dipole in the MM region. The most popular forcefields do not include polarizability tensors, making the polarised/unpolarised method the most popular.

2.3.2.2 Link Atoms

For systems where the QM region is covalently bound to the MM region it is necessary to draw the QM/MM boundary across either covalent bonds or through the centre of atoms, the prior being the simplest and most practical. The bonds on the QM/MM boundary are cut creating hypovalent atoms on either side. To satisfy these valency requirements of boundary atoms, several boundary schemes have been developed.

The simplest and most popular boundary scheme is the link atom method: a hydrogen atom (other monovalent atoms may be used, but hydrogen atoms are most common) is added to each atom

along the QM/MM boundary. As these dummy hydrogen atoms do not exist in the real system, their energy contribution must be removed from the overall system energy. The introduction of dummy protons has the potential to create undesired over-polarisation of the QM region by the MM region. To alleviate this effect, charge-shift schemes are often used to redistribute the charges of MM atoms close to the boundary. This can reduce spurious over-polarisation effects. Another potential downside of introducing link atoms is that the new QM atom – link atom bond may not describe the cut bond in the real system. To minimise this effect, it is best practice to place the QM/MM boundary across non-polar carbon-carbon bonds as the resulting dummy carbon-hydrogen bonds will have similar polarities.

An alternative boundary scheme is the boundary atom scheme[65]. Instead of placing the boundary across covalent bonds, the boundary is placed across the centre of atoms. These atoms become 'boundary atoms' which are treated as being in both the QM and MM regions. In the QM region, boundary atoms mimic the cut bond between them and other atoms in the QM region. In the MM region, boundary atoms behave as normal MM atoms.

Another treatment of the QM/MM boundary is the use of frozen localised orbitals[66]. Methods that take this approach place the QM/MM boundary across bonds. In the place of cut bonds, frozen hybrid orbitals are added to the QM region. These orbitals are 'frozen', as such they are not allowed to mix with other orbitals in the QM region and are not included in the SCF procedure. There are several methods that apply variations to this scheme such as using highly parameterised frozen orbitals[65].

2.3.2.3 Embedding Schemes

In systems where the QM/MM boundary is drawn across covalent bonds Equation 62 is insufficient to calculate the total energy due to the introduction of link atoms, thus other approaches are needed. The simplest method is called Mechanical Embedding where the energy of the system is calculated using the subtractive scheme:

$$E_{QM/MM}(all) = E_{MM}(all) + E_{QM}(inner + link) - E_{MM}(inner + link).$$

Equation 63: The Mechanical Embedding scheme.

In Equation 63, the total energy of the system $E_{QM/MM}(all)$ is evaluated by calculating the energy of the entire system using a MM method ($E_{MM}(all)$). The energy of the inner region (QM-region) and link atoms is calculated at both the QM and MM levels, with the MM energy subtracted from the total energy. This removes the effect of the link atoms on the final energy. This method does not

include any polarisation effects and relies on forcefield parameters for the QM region which may not be readily available. The mechanical embedding scheme is applied in the popular ONIOM method[67].

An alternative formulation of the subtractive scheme Equation 63 is the so-called additive scheme:

$$E_{QM/MM}(all) = E_{MM}(outer) + E_{QM}(inner + link) - E_{QM-MM}(coupling).$$

Equation 64: The subtractive embedding scheme.

In Equation 64, the energy of the outer region (MM-region) ($E_{MM}(outer)$) is calculated using a MM method, while the energy of the inner region (QM-region) and link atoms is calculated at the QM level ($E_{QM}(inner + link)$). The interaction between the QM and MM regions ($E_{QM-MM}(coupling)$) is then calculated separately.

The most popular way of calculating the $E_{QM-MM}(coupling)$ term in Equation 64 is through the use of the electrostatic embedding scheme.

The electrostatic embedding scheme is an extension of the polarised/unpolarised scheme discussed in Equation 62: The scheme for QM/MM interaction energy under Polarised/Unpolarised Boundary Conditions. In this method the energy of the QM system is calculated by introducing an additional term to the one electron operator:

$$\hat{h}_i^{QM/MM} = \hat{h}_i + \sum_{i}^{N} \sum_{m}^{MM\ point\ charges} \frac{q_m}{r_{im}}.$$

Equation 65: One-electron operator used to calculate the QM/MM interaction energy in the Polarised/Unpolarised scheme, under the electrostatic embedding scheme.

In Equation 65, $\hat{h}_i$ is the Hamiltonian operator seen in Equation 15: the form of a one electron integral. The second term describes the electrostatic interaction between the electron i and each atom in the MM region as a point charge in space. The Overall energy is then calculated using Equation 63.

2.4 Molecular Docking

The challenge of generating enzyme-substrate complexes *in-silico* has necessitated the development of molecular docking algorithms. The goal of these algorithms is to generate binding poses. Binding poses are geometries of a ligand bound to the active site of an enzyme. Molecular docking

algorithms tend to generate multiple binding poses per enzyme-substrate pair, then use scoring functions to decide which binding pose is the 'best'.

Rather than using physics-based methods, molecular docking algorithms tend to use a scoring function. This scoring function is empirically fitted to reproduce known poses of enzyme-substrate complexes, such as those obtained through X-ray crystallography. This section details the scoring function used in the molecular docking software AutoDock Vina[68] (used in this work), which was used to generate enzyme-substrate complexes in this project.

2.4.1 Scoring function of AutoDock Vina

The general form of the scoring function used in AutoDock Vina can be expressed as:

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}).$$

Equation 66: The conformation-dependent scoring function in AutoDock Vina.

In Equation 66, c is the conformation-dependent part of the scoring function. Each atom is assigned a type (expressed as t). c is then evaluated as a summation of interactions of atom pairs that are able to move relative to each other. Since the majority of atoms in a docking calculation are not allowed to move, the number of atoms considered in Equation 66 is limited to interactions between atoms within the ligand and interactions between the ligand and proximal protein atoms. In the case of flexible docking procedures, small portions of the protein (typically large, highly mobile residue side chains within the active site) are allowed to move throughout the simulation. In this case, atoms in the side-chains of flexible residues included as well as ligand atoms when it comes to creating atom pairs in Equation 66.

A sum of a set of symmetric interaction functions $f_{t_i t_j}$ and interatomic distances r_{ij} between atoms i and j are used to construct c. Hydrogen atoms are excluded from these calculations and as such, are only described implicitly. The interaction function $f_{t_i t_j}$is defined as relative to the distance between atomic Van der Waals surfaces of the atom pair in question:

$$d_{ij} = r_{ij} - R_{t_i} - R_{t\,j}.$$

Equation 67: Expression for the distance between interatomic surfaces.

In Equation 67, d_{ij} is the distance between atomic surfaces, r_{ij} is the interatomic distance measured centre-centre and R_{t_x} is the Van der Waals radius of atom x with atom type t_x. This d_{ij} term is then used to redefine the scoring function $f_{t_i t_j}$:

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}).$$

Equation 68: The scoring function in AutoDock Vina expressed using interatomic surface distance.

In Equation 68, $h_{t_i t_j}$is a weighted sum of steric, hydrophobic and (if applicable), hydrogen-bonding interactions between atoms i and j:

$$h_{t_i t_j} = steric(d) + w_4 hydrophobic(d) + w_5 hbonding(d).$$

Equation 69: The scoring function un AutoDock vina decomposed into steric, hydrophobic and hydrogen-bonding terms.

In Equation 69, terms w_4 and w_5 are weights. Note at this point the interatomic surface distance d_{ij} is expressed as simply d. The steric portion of Equation 69 is evaluated in the same manner for all pairs of atom types and is split into:

$$steric(d) = w_1 gauss_1(d) + w_2 gauss_2(d) + w_3 repulsion(d).$$

Equation 70: Terms used to calculate the steric interaction of an atom pair.

Equation 70, the terms w_1, w_2, and w_3 are weights. The terms in Equation 70 can be further decomposed to:

$$gauss_1(d) = e^{-\left(\frac{d}{0.5\ \text{Å}}\right)^2}$$

Equation 71: The steric term $gauss_1$ used in AutoDock Vina.

$$gauss_2(d) = e^{-\left(\frac{d-3\ \text{Å}}{2\ \text{Å}}\right)^2}$$

Equation 72: The steric term $gauss_2$ used in AutoDock Vina.

$$repulsion(d) = \begin{cases} d^2, if\ d < 0 \\ 0, if\ d \geq 0 \end{cases}$$

Equation 73: The repulsive term used in AutoDock Vina.

The hydrophobic term in Equation 69 is evaluated as:

$$hydrophobic(d) = \begin{cases} 1, & if\ d < 0.5\ \text{Å} \\ 1 \to 0, & if\ 0.5\ \text{Å} \leq d < 1.5\ \text{Å} \\ 0, & if\ d \geq 1.5\ \text{Å} \end{cases}$$

Equation 74: The hydrophobic term used in AutoDock Vina.

In Equation 74, the value of the hydrophobic term is 1 when d is less than 0.5 Å and 0 when d is equal to or more than 1.5 Å. For values of d in between 0.5 Å and 1.5 Å, the value of the hydrophobic term is linearly interpolated between 1 and 0.

The final term in Equation 69 is associated with hydrogen-bonding. This term is only non-zero when the atom pair in question is capable of forming a hydrogen-bond. If appropriate, the hydrogen-bonding term is evaluated as:

$$hbonding(d) = \begin{cases} 1, & if\ d < -0.7\ \text{Å} \\ 1 \rightarrow 0, & if\ -0.7\ \text{Å} \leq d < 0\ \text{Å} \\ 0, & if\ d \geq 0\ \text{Å} \end{cases}$$

Equation 75: The hydrogen-bonding term used in AutoDock Vina.

In Equation 75, the value of the hydrogen-bonding term is 1 when d is less than -0.7 Å and 0 when d is equal to or more than 0 Å. For values of d in between -0.7 Å and 0 Å, the value of the hydrogen-bonding term is linearly interpolated between 1 and 0.

2.4.2 Optimisation algorithm of AutoDock Vina

AutoDock Vina seeks to minimise the value of Equation 66, producing a binding pose for the desired system with the lowest possible binding energy (which is calculated by a function of the intermolecular portion of c in Equation 66). To obtain this global minimum value of Equation 66, the Iterated Global Search global optimiser is used. This algorithm consists of a sequence of steps in which a small conformational modification is made, followed by a local optimisation, using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The step is then accepted or rejected based upon the Metropolis criterion:

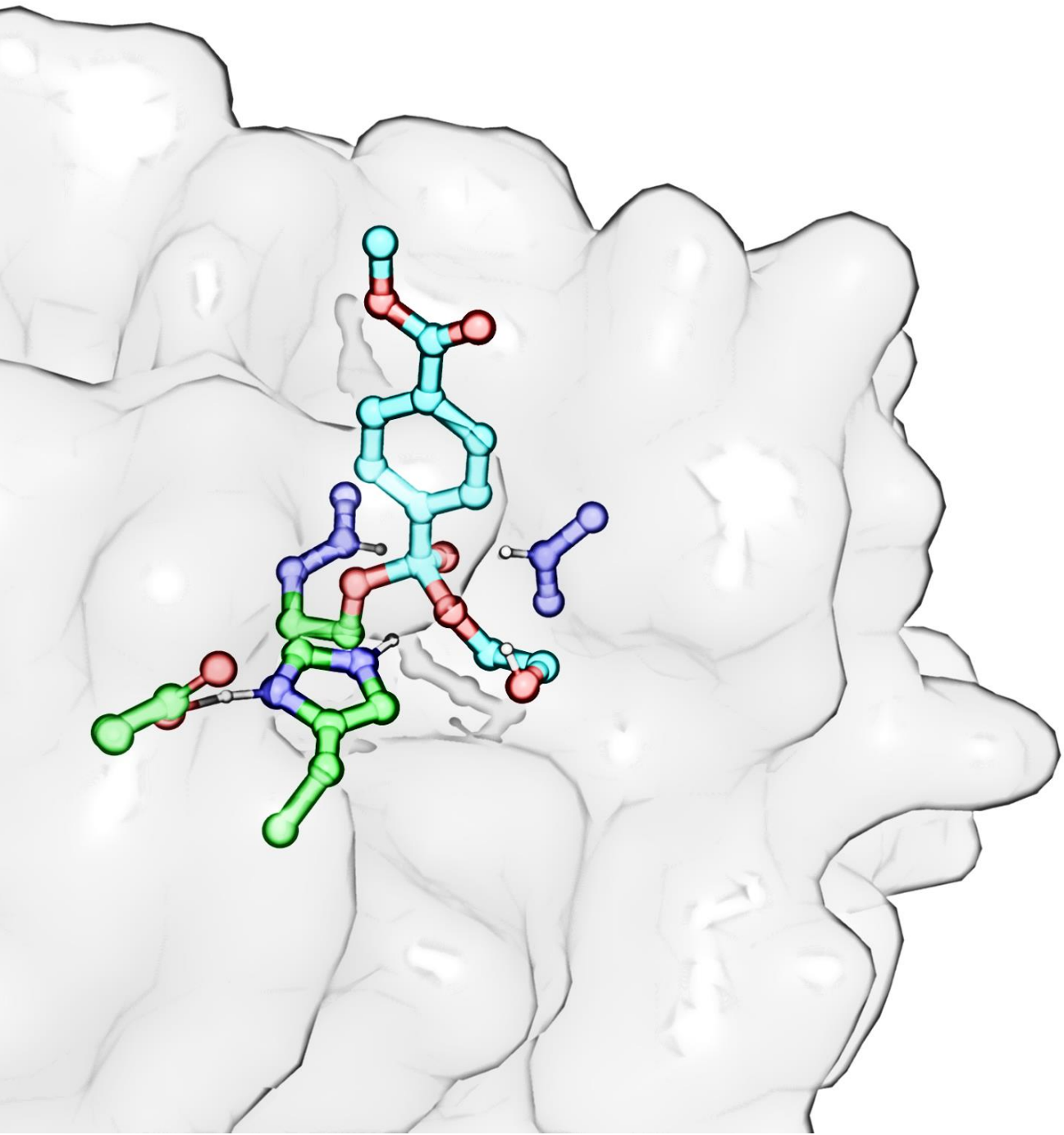
$$A = min\left(\frac{f(x_t)}{f(x_{t-1})}, 1\right)$$

Equation 76: The Metropolis acceptance criterion.

In Equation 76, A represents the probability of accepting the step. The term $f(x_t)$ represents the value of the function in question after a step has been taken, while $f(x_{t-1})$ represents the function's value before the step has been taken. The use of the min function ensures that steps that reduce the value of $f(x)$ always get accepted.

Metropolis methods often have problems of slow convergence for some non-global minima. To circumvent this problem, AutoDock Vina uses multiple starting conformations for its enzyme-substrate complexes and runs several separate optimisations.

Chapter 3: Evaluation of the Mechanism of Is-PETase



3.1 Previous Computational Studies on the Mechanism of Is-PETase

At the outset of this project, several studies had used either molecular docking[18,27,39] or molecular dynamics simulations[21] to explore the binding modes of various substrate mimics to the active site of PETase. No comprehensive studies using QM/MM methods had been performed to explore the catalytic mechanism of *Is*-PETase. At the time of writing (four years later) five such studies have been published and four studies have been published that have explored the catalytic mechanism of *Is*-PETase's sister enzyme, MHETase using QM/MM methods. This section will discuss important features of the methodologies applied in each study as well as their results.

Through a combination of X-ray crystallography and molecular docking studies, it has been theorised that *Is*-PETase functions via a four-step catalytic cycle (see Figure 17). Unless otherwise stated each of the QM/MM studies have created reaction profiles for this reaction mechanism.

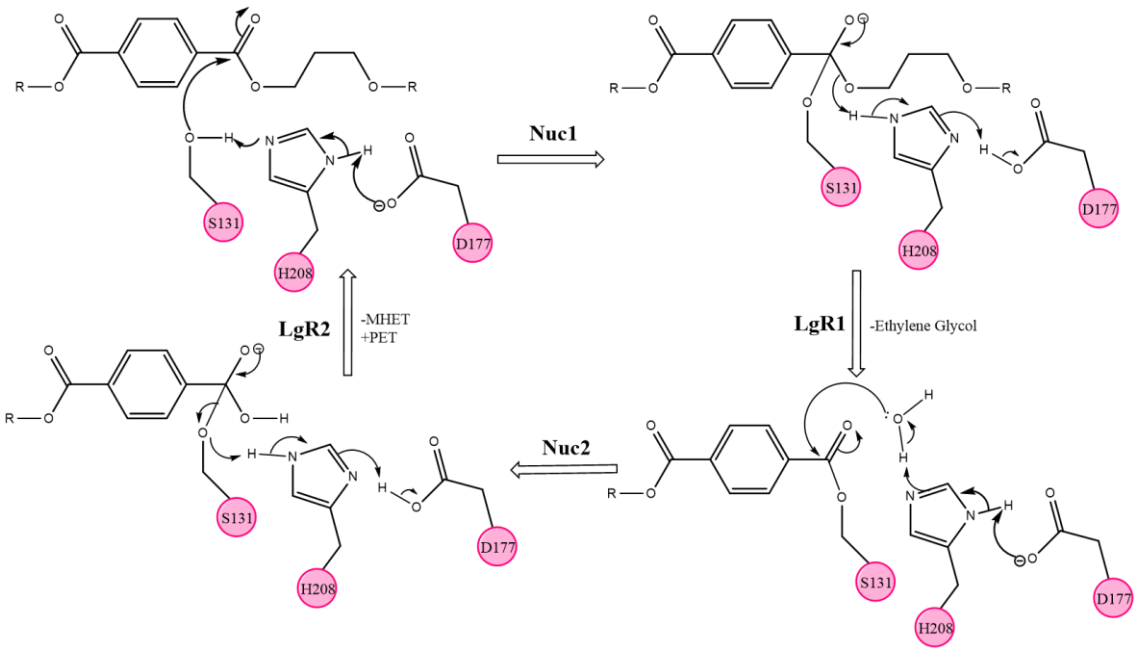


Figure 17: The four-step catalytic cycle for Is-PETase.

3.1.1.1 Semi-Empirical methods used to explore the mechanism of MHETase

The first QM/MM study related to Is-PETase was by Knott et al.[69]. This study used the umbrella sampling method to explore the free energy surface of the degradation of MHET via the enzyme MHETase. MHETase acts synergistically with Is-PETase, degrading the major product of Is-PETase's reaction, mono hydroxyethyl terephthalate. Whilst this does not provide information on the catalytic mechanism of Is-PETase, the results from this study can be used for comparison. For example, it is

known that MHETase has a higher turnover rate than Is-PETase. This is either due to lower energy barriers within the catalytic mechanism or to faster rates of substrate binding or product release.

The semi-empirical SCC-DFTB method was used to describe the QM region. It is unclear what forcefield was used to describe the MM region. The proposed mechanism for the degradation of MHET via MHETase followed the same scheme as that of *Is*-PETase (see Figure 18).

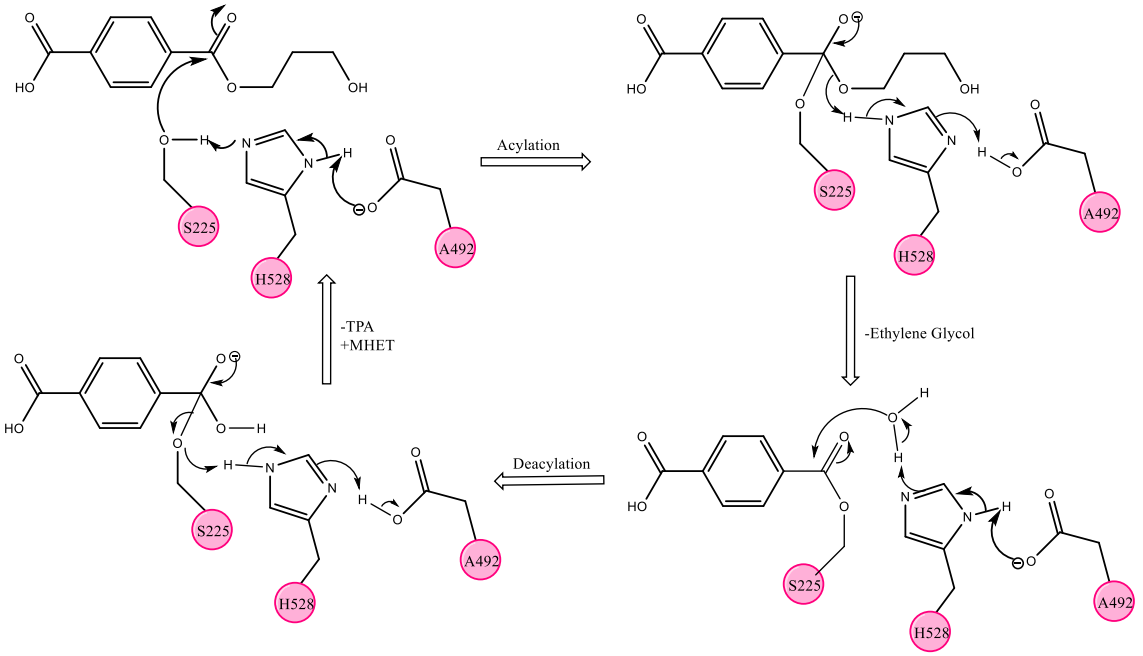


Figure 18: The four-step mechanism for MHETase.

This method resulted in a two-step reaction mechanism for MHETase, as opposed to the canonical four-step mechanism. The enzyme-bound tetrahedral intermediates (see Figure 18, top right and bottom left) in the proposed mechanism were found to be unstable. This led to a truncated two-step mechanism being proposed, comprising an initial acylation/ester cleavage step followed by a hydrolysis/deacylation step (see Figure 19).

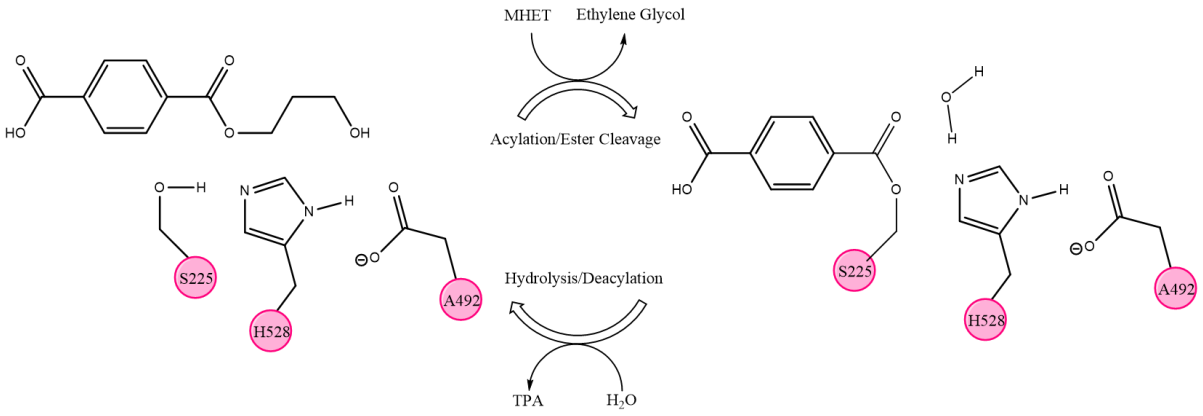


Figure 19: The truncated two-step mechanism for MHETase proposed by Knott et al.[69].

The free energy barriers for these two steps (see Figure 19) were calculated to be 58.2 kJ mol^{-1} and 82.8 kJ mol^{-1} for the acylation/ester cleavage and hydrolysis/deacylation steps respectively, making the latter step rate-limiting.

The activation energy for the catalytic mechanism of MHETase may not be fully transferable to the mechanism of its sister enzyme *Is*-PETase. It is known however that MHETase has a faster turnover rate than *Is*-PETase under the same conditions. Comparison of the calculated energy barriers for MHETase and *Is*-PETase can provide insights into the causes of this discrepancy in reaction rates. If the energy barrier for the reaction catalysed by *Is*-PETase is lower or comparable to that of MHETase, it can be inferred that MHETase's superior rate is due to more rapid substrate binding and/or product dissociation events. If the activation barrier for *Is*-PETase is found to be significantly larger than that of MHETase, it could be inferred that the activation energy itself is the cause of the discrepancy in rates.

3.1.1.2 Semi-Empirical Methods used to Explore the Mechanism of Is-PETase

The first QM/MM study on the catalytic mechanism of Is-PETase was by Boneta et al.[29] In this paper, umbrella sampling was used to sample the FREE ENERGY SURFACE of Is-PETase degrading the substrate mimic $MHET_2$ (a dimer of PET). A 2-dimensional potential energy surface for each reaction step (see Figure 20) was first generated using the umbrella sampling method with the semi-empirical AM1 method[70] used to describe a small QM region (consisting of only the ligand and the side-chain atoms of Ser131, Asp177 and His208), with the AMBER ff03 forecefield used to describe the remainder of the system. The full free energy surface was then generated using an interpolation correction with the hybrid meta exchange-correlation functional M06-2X with the 6-31+G(d,p) basis set. The results of this work concluded that Is-PETase follows the classical four-step hydrolysis mechanism, conserved across the α/β-hydrolase family of enzymes (of which Is-PETase is a member).

It was concluded that the highest energy barrier in the reaction profile was associated with the reaction step LgR1: the cleavage of the substrate's ester bond. The activation free energy for this step was calculated to be 98.7 kJ mol^{-1}. This paper makes comparisons to ΔG values obtained from previous experimental studies, concluding that their calculated values are in agreement with those obtained experimentally. Comprehensive kinetic studies of the reaction catalysed by *Is*-PETase would be required to accurately obtain these ΔG values. At the time of writing, these studies are yet to be performed. It is therefore doubtful that these comparisons are particularly valuable.

The small QM region (see Figure 20) used in this paper's calculations may have had an adverse effect on their results. A key feature of the canonical α/β-hydrolase mechanism is the stabilisation of

negative charges that develop on the carbonyl oxygen of the substrate during the acylation and hydrolysis steps (reaction steps 1 and 3) through hydrogen bonding interactions with backbone nitrogen atoms. The backbone nitrogen atoms of Tyr58 and Met132, which make up the oxyanion hole of *Is*-PETase, have not been included in the QM region in this paper's calculations. Since the substrate has been included in the QM region, this important interaction has been evaluated as an interaction between the QM and MM regions. It is difficult to quantify the extent of the accuracy loss that would result from this, as this would depend on the interaction between the specific QM and MM methods used.

The geometries that make up the free energy surface of each reaction step were generated using the semi-empirical method AM1. Whilst this method is reasonably inexpensive, it is known to inadequately describe hydrogen bonding interactions, which are crucial to the reactivity of *Is*-PETase. Perhaps the replacement of AM1 with a more modern semi-empirical method such as xTB[71] would have yielded more accurate geometries.

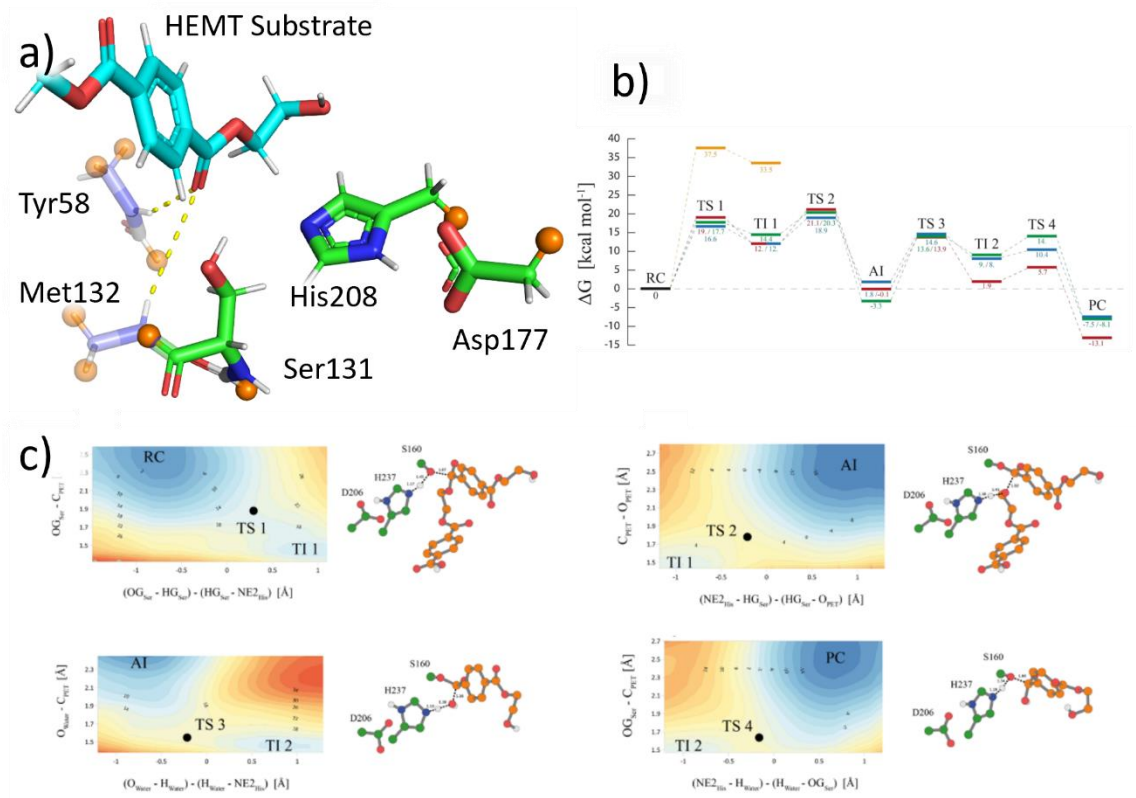


Figure 20: The QM region selected by Boneta et al. (a) shown in solid sticks. The oxyanion hole (shown in transparent sticks) was not included in this QM region. Relative energies of intermediates and transition states calculated by Boneta et al. (b) for enzyme-substrate pairs Is-PETase/MHET$_2$ (green), Is-PETase/MHET$_3$ (blue), LCC-ICCG/MHET$_2$ in a flipped orientation (yellow) and LCC-ICCG/MHET$_2$ in its 'normal' configuration. Plots showing two-dimensional potential energy surfaces

for the degradation of PET via Is-PETase, generated though umbrella sampling (c). Images (b) and (c) were taken from Boneta et al.[29].

3.1.1.3 DFT Method used to Explore the Mechanism of Is-PETase

Another approach to examine the catalytic mechanism of Is-PETase was taken by Feng et al.[31]. In this work, the degradation of the substrate mimic 2-hydroxyethyl-(mono-hydroxyethyl terephthalate)$_2$ (2-HE(MHET)$_2$) via Is-PETase was investigated using QM/MM methods using a DFT method to describe a large QM region. The CHARMM27 forcefield was used to describe the MM region.

The reaction profile for each step in the canonical α/β-hydrolase reaction mechanism (see Figure 17) was calculated using a microiterative transition state optimisation method. The geometries for intermediates and transition states were optimised at the M06-2X/6-31G(d):CHARMM27 level, while their energies were evaluated at the higher M06-2X/6-311G(d,p):CHARMM27 level. The QM region used for these calculations was larger than the one used by Boneta *et al.* (as described in Chapter 3.1.1.3) with atoms from the ligand as well as all atoms within residues in the catalytic triad (Ser131, His208 and Asp177) as well as all atoms in within residues that make up the oxyanion hole (Tyr58 and Met132). This larger QM region is an improvement over that of Boneta's, as the hydrogen bonding between OC_{LIG} and the oxyanion hole is now being described at the DFT level. The choice of QM region does however place the QM/MM boundary across several carbon-nitrogen bonds; this can result in spurious charges developing in the QM region. Better practice would be to include the backbone carbonyl carbon and oxygen of residues at the N-terminus of each residue of interest in the QM region. This ensures that the QM/MM boundary is placed across the bond between a carbonyl carbon and a α-carbon.

Twenty initial structures were used for each reaction step, resulting in twenty independent reaction profiles for each step. The overall energy barrier for each reaction step was then generated by Boltzmann-weighted averaging over the corresponding reaction profiles. The rate-determining step was determined by comparing the arithmetic average activation energy of each reaction step rather than the Boltzmann-weighted average. The initial Nuc1 and the final LgR2 product-release step had energy barriers of 85.8 kJ mol^{-1} and 87.0 kJ mol^{-1-} using arithmetic averaging. These energies are well within the standard deviation their respective set of energies calculated, making them functionally equivalent. Interestingly, these same steps are more distinguishable when using Boltzmann-weighted averaging, with the initial acylation step having a barrier of 58.2 kJ mol^{-1} and the fourth product release step having an energy barrier of 49.8 kJ mol^{-1}. This small change in methodology not only has a significant impact on the magnitude of the reported energy barriers but also firmly shows that the acylation step is rate-limiting. As discussed later in section 3.3.4, Boltzmann-weighted averaging is biased towards low-lying reaction profiles and thus mimics natural processes better

than arithmetic averaging. As such, the values obtained via Boltzmann-weighted averaging should be considered to be more reliable.

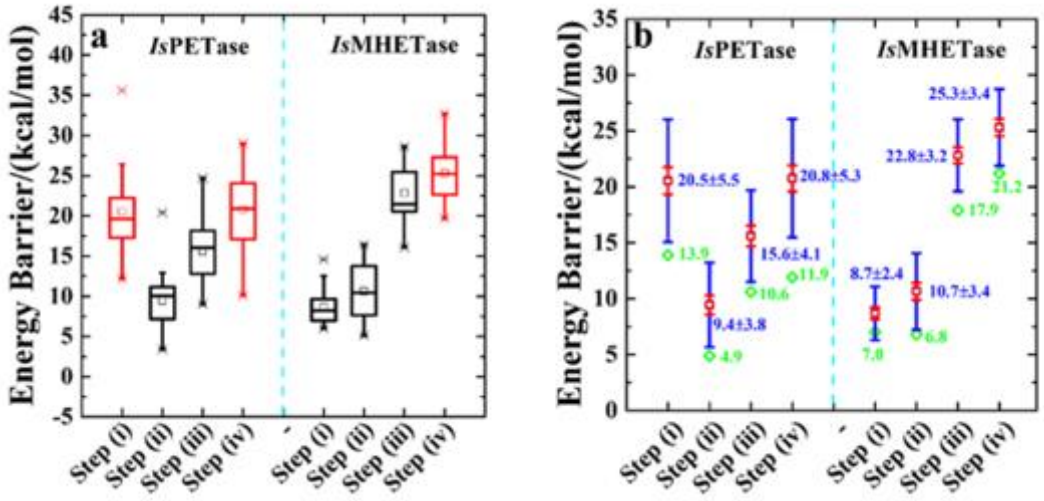


Figure 21: (a): Energy ranges for reaction steps 1 to 4 shown as box and whisker plots for the degradation of MHET by Is-PETase and MHETase. (b): The arithmetic average ± standard deviation in energy for each step is shown in blue and the arithmetic average ± standard error is shown in red. The Boltzmann-weighted average energy barrier for each step is shown in green. This figure was taken from Feng et al.[31]. Steps i), ii), iii) and iv) are referred to as Nuc1, LgR1, Nuc2 and LgR2 in this work.

Identical methods were applied to the mechanism of MHETase. These methods found that MHETase does indeed follow the canonical four-step α/β-hydrolase mechanism, with stable enzyme-bond tetrahedral intermediates found after the initial acylation step and after the nucleophilic attack of the solvent water molecule. These results directly contradict the conclusions made by the earlier study by Knott *et al.* Since the computational methods used in this study are at a significantly higher level (energies calculated at the M06-2X/6-311G(d,p)//CHARMM27 vs. the semi-empirical SCC-DFTB method), it is reasonable to assume that MHETase is more likely to adopt the four-step mechanism. The highest energy barrier for the reaction mechanism of MHETase was calculated to be the final (fourth) product release step, with a Boltzmann-weighted average activation energy of 88.7 kJ mol^{-1}. This value is close to the highest energy barrier calculated by Knott *et al.* of 82.8 kJ mol^{-1}.

3.1.1.4 DFT/MM Molecular Dynamics simulations used to Explore the Mechanism of Is-PETase

Another study by Jerves *et al.* [REF] used the umbrella sampling method to explore the catalytic mechanism of *Is*-PETase. The density functional PBE with the DZVP-GTH basis set was used to describe a large QM region in these calculations, while the Amber ff14SB forcefield was used to describe the remaining MM region. To achieve the most accurate results, input geometries for

umbrella sampling calculations were generated using the highly demanding Born-Oppenheimer molecular dynamics (BOMD) method, using the same functional and forcefield.

50 ns MD simulations of *Is*-PETase with a bound dimer of PET (2PET) were performed using BOMD, analysis of the resulting trajectories revealed that the residues Thr59, Ala60, Trp130, Trp156, Ile179, Ser207 and Ser209 are involved in the binding of the 2PET, mostly through van der Waals interactions with the aromatic rings of 2PET. Whilst this has been observed in previous studies, this BOMD simulation has the highest accuracy of any MD simulation run on this enzyme to this date. It is worth noting the GGA method PBE has been shown to poorly describe Van der Waals interactions[72]. To rectify this shortcoming, perhaps a dispersion correction should have been introduced.

In contrast to the work performed by Boneta *et al.* and Feng *et al.*, this study treated the catalytic mechanism of *Is*-PETase as a two-step process (rather than the canonical four-step mechanism common to α/β-hydrolases) consisting of an initial acylation step followed by a deacylation step. In the acylation step, the nucleophilic attack of Ser131 and the cleavage of the ligand's ester bond occur in concert. Similarly, in the deacylation step the nucleophilic attack of an activated water molecule and the cleavage of the Ser131-ligand ester bond occur in concert. The energy barriers calculated for the acylation and deacylation steps were calculated to be 83.7 kJ mol^{-1} and 63.2 kJ mol^{-1} respectively. These results lead to the conclusion that the acylation step is rate-determining.

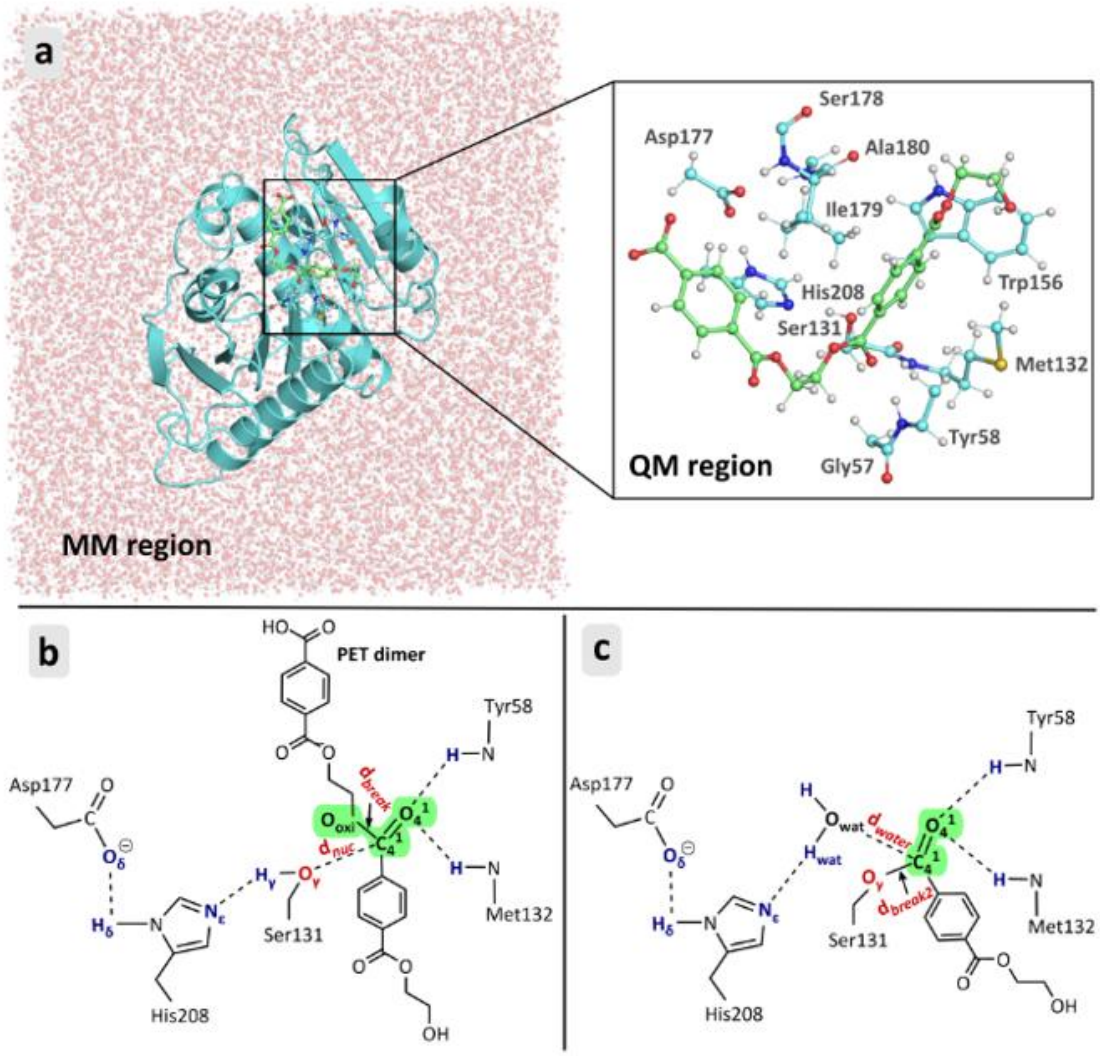


Figure 22: a): the large QM region used by Jerves et al. The transition states for the two-step mechanism proposed by Jerves et al are shown in (b) and (c). Figure taken from Jerves et al.[30]. The protonation state of the PET substrate in (b) and (c) are for illustrative purposes only. Calculations were performed on deprotonated PET, as shown in (a).

Before this work, the diffusion of the product of the cleavage of the ligand's ester bond (in this case MHET) had not been considered. A significant energy barrier could be associated with the removal of this product and its replacement by a water molecule, which is required for the following deacylation step. If this dissociation barrier exists, or if the relative energies before and after this process are significantly different, then the direct connection between the acylation and deacylation steps is not appropriate. In this study, no stable minimum was found for the end point of the acylation step; this was thought to be due to spontaneous dissociation of the MHET product. To confirm this, BOMD simulation was performed using this end-point as an input structure. It was found that the MHET product dissociated from the enzyme in less than 1 ns, leading to the conclusion that the dissociation of the products of the acylation step from *Is*-PETase is a spontaneous, entropically-driven process.

3.1.1.5 DFT Method used to compare activity of Is-PETase and TfH

An additional study applying QM/MM methods to the activity of Is-PETase was performed by Aboelnga et al.[32]. Reaction profiles for the degradation of a 2PET ligand (see Figure 23) were generated for the enzymes Is-PETase and TfH (discussed in Chapter 1.1.2.1). Through molecular dynamics simulations, it was revealed that 2PET binds to TfH in two modes; 'bent' and 'linear', while it binds to Is-PETase in the 'bent' mode only (see Figure 23). For this reason, one reaction profile was generated for Is-PETase with 2PET bound in the 'bent' mode. Two reaction profiles were generated for TfH, one for each binding mode observed. Large QM regions were selected for the QM/MM calculations containing the catalytic triad (Ser131-His208-Asp177 for Is-PETase and Ser130-His208-Asp176 for TfH), oxyanion hole (Tyr58-Met132 for Is-PETase and Tyr58-Met131 for TfH) and aromatic residues thought to contribute to substrate binding (Trp130-Trp156 for Is-PETase and His129-Trp155 for TfH). Intermediate and transition state geometries were generated using the DFT functional B3LYP with the dispersion model D3BJ and the 6-31G(d,p) basis set. Frequency calculations were performed upon intermediates and transition states to verify that they were energy minima and saddle-points respectively. Single-point energy calculations were performed using the same DFT method, but with the larger 6-311-G(3df,3pd) basis set. A visual inspection was performed to ensure that transition states lay on the same reaction coordinate as adjacent intermediates. The reaction mechanism used in this work follows the four-step reaction scheme (see Figure 17), although the second intermediate in the acylation reaction (in this work I1, IC1 in Figure 23) was not found to be stable for the degradation of 2PET by TfH in either binding pose.

Using the methods described above, it was identified that the second step of the deacylation process (in this work LgR2) was rate-limiting for *Is*-PETase, with an energy barrier of 75.3 kJ mol^{-1}. For TfH it was found that the rate-limiting step for the 'bent' binding pose was also LgR2, with an energy barrier of 87.0 kJ mol^{-1}. For the 'linear' binding pose of 2PET to TfH the first step of the deacylation process (in this work Nuc2) was found to be rate-limiting, with an energy barrier of 85.7 kJ mol^{-1}. From these calculations it was concluded that *Is*-PETase has a faster rate of reaction than TfH due to its lower overall energy barriers. Additionally, it was concluded that TfH's degradation reaction proceeds via the 'linear' binding pose rather than the 'bent' binding pose due to its preference for this pose in a long 800 ns MD simulation.

The role of the novel disulfide bond in *Is*-PETase (discussed previously in section 1.1.2.3 and Figure 7) was further investigated. An 800 ns MD simulation was performed on wild-type *Is*-PETase and a mutant with the disulfide bond removed (through knock-out C175A-C211A mutations). It was found that hydrogen bonding distances in the catalytic triad were significantly shorter (and therefore stronger) in wild-type *Is*-PETase than in the knock-out mutant.

It is worth noting that this study did not attempt to sample the conformational space of the reactions studied. Only one reaction profile was generated for each of the three reactions. In earlier studies[29–31] (discussed in Chapters 3.1.1.2, 3.1.1.3, and 3.1.1.4), conformational sampling has been achieved through generating multiple reaction profiles[31], umbrella sampling[29] and Born-Oppenheimer Molecular Dynamics methods[30]. It is quite possible that through generation of several additional reaction profiles using the techniques described above, reaction pathways with lower overall energy barriers would have been found. As described in Chapter 3.1.1.3, the selection of QM atoms has not been performed using best-practice methods. The QM-region has been selected using whole residues. This means that the QM/MM interface has been placed through multiple amide bonds which has been shown to have unpredictable effects on the charge distribution in the QM region.

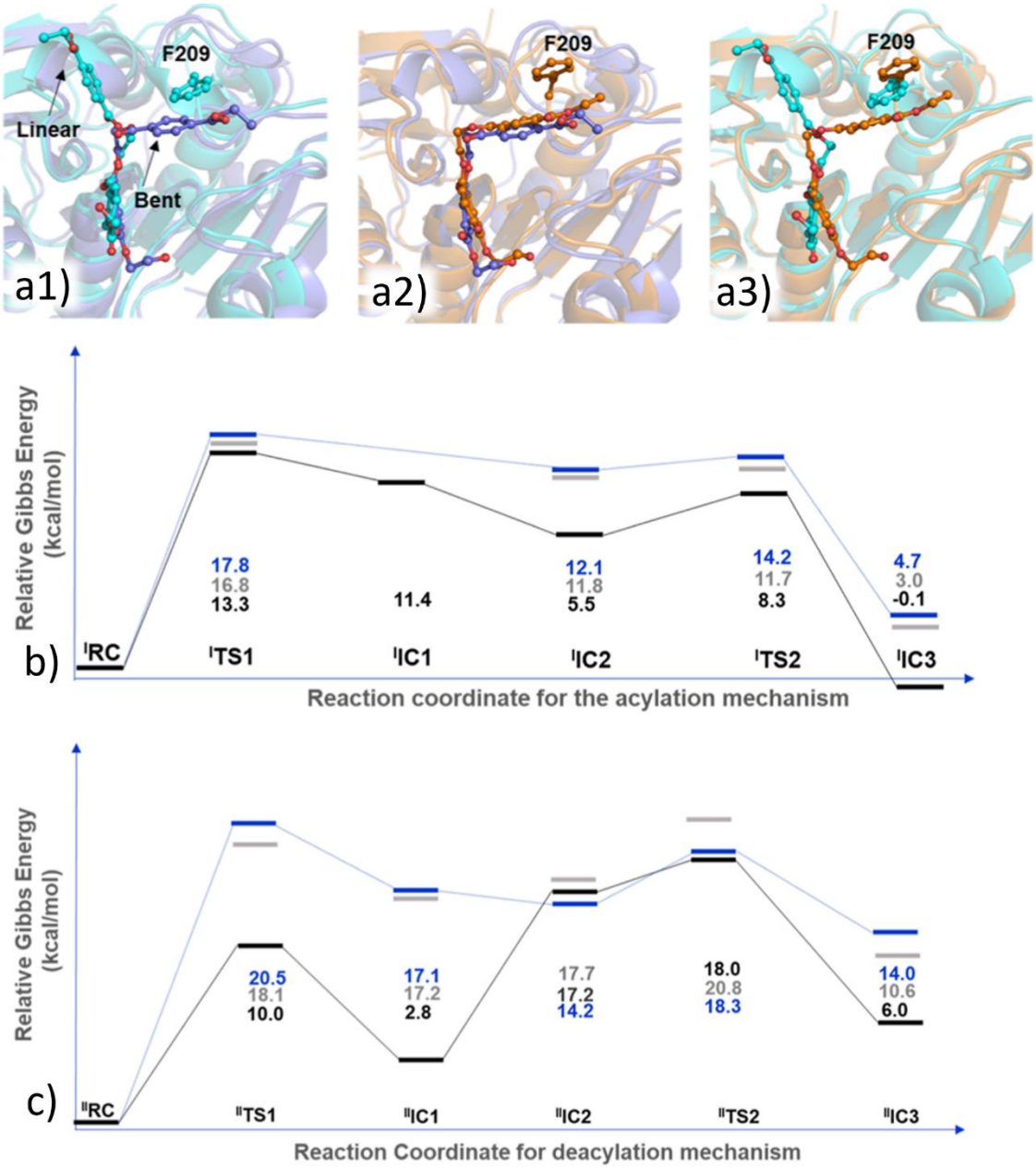


Figure 23: a) Overlays of conformations of 2PET the active site of Is-PETase and TfH enzymes. The 'bent' conformation of 2PET in Is-PETase (purple) is shown in a1 and a2. The 'bent' conformation of 2PET in TfH (orange) is shown in a2 and a3. The 'linear' conformation of 2PET in TfH (cyan) is shown in a1 and a3. b) the energies of intermediates and transition states for the acylation reaction for Is-PETase (black), TfH with 'bent' 2PET (blue) and 'linear' 2PET (grey). Images taken from Aboelnga et al.[32].

3.1.1.6 DFT methods used to explore the role of Asp177 stabilising residues in Is-PETase's mechanism

An exploration of the mechanism of *Is*-PETase using DFT QM/MM methods has been performed by Fernandes *et al.* with the goal of yielding additional insight into the roles of non-catalytic residues in the enzyme's active site[33]. This knowledge could prove useful in attempts to engineer *Is*-PETase to have favourable characteristics such as enhanced catalytic rate or substrate affinity.

Following the four-step reaction mechanism (see Figure 17) for *Is*-PETase, intermediates along acylation and deacylation reactions (NS1 and NS2 in this work) were generated using QM/MM geometry optimisations with the B3LYP density functional with the 6-31G(d,p) basis set. The AMBER18 forcefield was used to treat the MM region for all calculations in this work. Transition state geometries were generated using internal reaction coordinate calculations; their identity as saddle-points were confirmed using frequency calculations. Single-point energy calculations were performed upon the geometries of intermediates and transition states using an extrapolation method between the original B3LYP/6-31(d,p) method and the DLPNO-CCSD(T) with cc-pVDZ and cc-pVTZ basis sets and their corresponding correlation fitting basis sets.

Novel to this study, the residues between Ser178 and Ala180 were included in the QM region, alongside a solvent molecule proximal to the side chain of Asp177 and the backbone nitrogen of Ala180. It was found that a hydrogen bonding network formed between the backbone nitrogen of Ala180, the water molecule and the carboxyl or Asp177. Throughout the acylation and deacylation steps, Boneta *et al.* found that the proton originally bound to the ND1 atom in His208 was transferred to Asp177. This proton transfer was observed in this work when the Ser178-Ala180 (and accompanying water molecule) region was not included in the QM region. With the addition of this region to the QM region, no proton transfer was observed. It was concluded that this water-mediated hydrogen bonding interaction is key to the stabilisation of the catalytic Asp177 residue.

The role of Ile179 was investigated using a couple of MD simulations, one with wild-type PET and one with the I179A mutation applied to *Is*-PETase. Analysis of the hydrogen-bonding distance between the catalytic residues Asp177 and His208 revealed significant destabilisation for the I179A mutant. It was concluded that the steric bulk of Ile179 is key to maintaining the catalytic orientation of Asp177. The smaller 1179A residue allowed Asp177 to rotate to become solvent exposed, replacing the hydrogen bonding interaction with His208 with hydrogen-bonding interactions with the solvent water molecules.

A single reaction profile was generated for the acylation and deacylation reactions. The rate-limiting step was found to be the first step of the deacylation reaction (in this work Nuc2) with an activation barrier of 52.3 kJ mol^{-1}. This is the lowest energy barrier found for the degradation of PET via *Is*-

PETase. As with the study performed by Aboelnga *et al.*[32], no attempt to explore the conformational space of either reaction was made in this work. Generation of additional reaction profiles is likely to have yielded lower overall energy barriers.

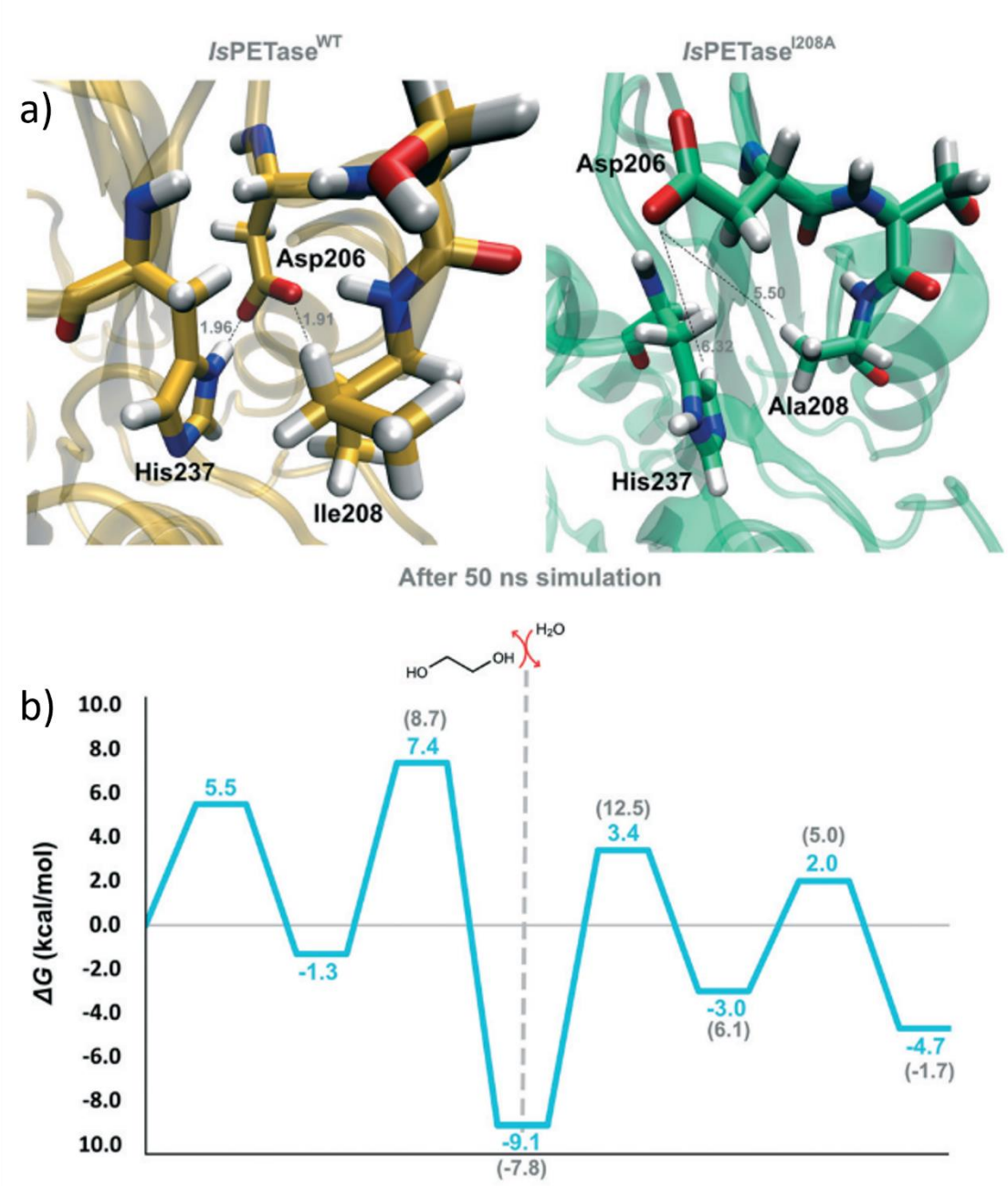


Figure 24: a): snapshots taken at 50 ns during MD simulations or wild-type Is-PETase (yellow) and the I179A mutant (note that the residue numbers displayed are 29 higher). b): The energy profile for the degradation of PET via Is-PETase. The reaction profiles for the acylation and deacylation reactions have been joined at the dotted line. Images taken from Fernandes et al.[33].

3.1.1.7 Summary of QM/MM approaches to Explore the Catalytic Mechanism of Is-PETase

In five of the studies discussed in the previous sections, different computational methods have been applied to the problem of the catalytic mechanism of *Is*-PETase. Each study is in agreement that *Is*-PETase adopts the same chemical mechanism (see Figure 18). Despite this, each approach resulted in slight discrepancies in each study's conclusions. Table 2 summarises the methodological differences between these studies, as well as their key results.

Table 2: Comparison of results and methods of QM/MM studies performed on Is-PETase.

Main Author	Geometry Optimisation Method	Single-Point Method	Conformational Sampling	Transition state search	Rate-Determining Step	Activation Energy kJ mol^{-1}
Boneta	AM1 /M06-2X /6-31+G(d,p) /AMBER ff03	None	Umbrella Sampling	Umbrella Sampling	**LgR1**	98.7
Feng	M06-2X /6-31G(d) /CHARMM27	M06-2X /6-311G(d,p) /CHARMM27	20 energy profiles per reaction step	Microiterative Optimisation	**Nuc1 and LgR2**	86.6 and 87.0
Jerves	PBE /DZVP-GTH /AMBER ff14SB	None	Umbrella Sampling	Umbrella Sampling	**Acylation**	83.7
Aboelnga	B3LYP-D3(BJ) /6-31-G(d,p)	B3LYP-D3(BJ) /6-311-G(3df,3pd)	None	Frequency Calculations	**LgR2**	75.3
Fernandes	B3LYP /6-31(d,p) /AMBER18	Extrapolation to DLPNO-CCSD(T) /cc-pVTZ	None	Internal Reaction Coordinate Scan	**Nuc2**	52.3

The most obvious point of difference is that each paper concluded that a different reaction step (or pair of steps) was rate-determining. In the work by Boneta *et al.* the rate-determining step was found to be the LgR1 step (ligand ester cleavage) with an energy barrier of 98.7 kJ mol^{-1}. In the work by Feng *et al.* the rate-determining step was assigned using an arithmetic averaging of several reaction profiles. This resulted in Nuc1 (nucleophilic attack of Ser131) and LgR2 (product release) having activation energies within their respective standard deviations; as such both steps were concluded to be rate-determining. In the work by Jerves *et al.* stable tetrahedral intermediates in between steps Nuc1 and LgR1, and in between steps Nuc2 and LgR2 were not found. As a result, the mechanism of *Is*-PETase was said to follow a two-step process – combining steps Nuc1 and LgR1 and steps Nuc2 and LgR2. The first acylation/ester cleavage step was found to be rate-limiting with an activation energy of 83.7 kJ mol^{-1}. In partial agreement with Feng *et al.*, Aboelnga *et al.* found the LgR2 step to be rate limiting, with an activation energy of 75.3 kJ mol^{-1}. Finally, Fernandes *et al.* concluded that the Nuc2 (nucleophilic attack of water molecule) reaction step was rate-limiting, with an activation energy of 52.3 kJ mol^{-1}. From these five studies, each of the four possible reaction steps has been separately identified as the rate-determining step. Clearly, more work in this area is required to determine the 'true' rate-determining step.

The differences in the methodologies in each paper can be separated into two categories: protocol used to generate the reaction profile and level of theory used. In general, it is assumed that

computational calculations performed at higher levels of theory have more accurate results. Using this simple heuristic, the work by Feng *et al.* would be considered the most accurate study. Feng *et al.* used the hybrid meta-GGA density functional M06-2X to evaluate a large QM region, the basis set 6-31+G(d) was used to generate geometries, the energies were then re-evaluated using the larger 6-311+G(d,p) basis set. Jerves *et al.* used the less accurate GGA density functional PBE with the DZVP basis set and GTH pseudopotentials to calculate the geometries along the reaction profile. Crucially, the energies of energy minima and transition states were not recalculated using a higher level of theory. Boneta *et al.* used the semi-empirical AM1 method with an interpolation correction using the hybrid meta-GGA M06-2X density functional with the 6-31+G(d,p) basis set. No further single-point energy calculations were performed. The studies by Fernandes[33] and Aboelnga[32] used the hybrid functional B3LYP with the 6-31(d,p) basis set to generate geometries. Whilst B3LYP once enjoyed considerable popularity, it has been shown to perform worse than modern GGA functionals[73].

The protocol component of the methodological differences in these studies is far harder to quantify. Boneta[29] and Jerves[30] used umbrella sampling techniques to construct geometries along the reaction profile of *Is*-PETase. By contrast, Feng[31] used a microiterative transition state optimising protocol to find transition states between independently generated intermediates. Umbrella sampling has the benefit of generating a large enough set of geometries to calculate the free-energy surface of the reaction profile, while the approach by Feng et al. generates the potential energy surface instead. The downside of umbrella sampling techniques is that they may not be able to find transition state geometries accurately. This could result in inaccuracies in the calculated energies of transition states. As noted in Chapters 3.1.1.5 and 3.1.1.6, Aboelnga *et al.*[32] and Fernandes *et al.*[33] did not use any methods to explore the conformational space of the reactions catalysed by *Is*-PETase. The result of this is that the energy values quoted by these papers are from a single reaction profile. If multiple reaction profiles were generated, the energy values would likely be lower.

Overall, there has been no consensus as to the rate-limiting step of the reaction catalysed by *Is*-PETase. Several differing computational approaches have been applied to this question, with superficially similar energy profiles for the reaction but important differences in their conclusions. This makes additional study in this area valuable.

3.2 Methodology

3.2.1 Preparation of an Enzyme-Substrate Complex

3.2.1.1 Crystal Structure of Is-PETase

The geometry for the enzyme *Is*-PETase was taken from the Protein Data Bank (PDB id.: 5xh3)[25]. All molecules other than the enzyme including the ligand, crystallised water molecules, ions and crystallisation additives were removed using the molecular visualisation software Pymol[14]. To co-crystalise *Is*-PETase and an inhibitor, the following single-point mutations had been introduced: S131A R013G. Using Pymol, these mutations were reversed to generate the structure of the wild-type enzyme. The PropKa tool[74] was used to assign the protonation states of the titratable residues His, Asp and Glu assuming a pH of 9 as this has been found to be the optimal pH for the function of *Is*-PETase[18,23]. In all cases, histidine residues were assigned the HSD protonation state, where the δ-nitrogen atom is protonated and the ε-nitrogen atom is deprotonated. In all cases, all aspartate and glutamate residues were assigned to their deprotonated states.

3.2.1.2 Molecular Docking Simulations

Molecular docking simulations were performed using the program Autodock Vina[68]. Structures for ligands (HEMT, 2PET and 3PET) used were created in Gaussview[75] as a pdb file. Enzyme and ligand pdb files were then converted to three pdbqt (a format used by Autodock tools) files: receptor, flexible residues and ligand using a custom python script calling in-built functions in Autodock Tools[76]. A cubic grid box with dimensions of 40 Å was placed over the active site of the enzyme by centring it over the hydroxyl oxygen of Ser131. The ligand is then placed within the box with a random position and orientation. The built-in searching and scoring functions of Autodock Vina then generated up to 12 poses of the ligands bound to the active site.

Docking poses were selected based on the orientation of the ligand's ester carbonyl in relation to the nucleophilic γ-oxygen of Ser131 and the backbone nitrogen of Tyr58 and Met130 which make up the oxyanion hole of *Is*-PETase.

3.2.2 Molecular Dynamics Simulations

3.2.2.1 Parameterisation

All molecular dynamics simulations were performed using parameters from the CHARMM36 forcefield to represent the protein and ions[77]. Water molecules were parameterised using the TIP3P model[78]. Ligand parameters were generated using the webservice SwissParam[79]. These ligand parameters are made by analogy and without any optimisation. This would significantly reduce the accuracy of any MD simulations involving a ligand. Fortunately, MD was only used to generate input structures for subsequent QM/MM geometry optimisations and therefore these parameters were suitable for this purpose. For the QM/MM calculations, the ligand parameters were used only for their non-bonded interaction terms. The use of the other terms in the ligand parameters was avoided by placing the entire ligand in the QM region for all QM/MM calculations.

3.2.2.2 Solvation and Charge Neutralisation

The enzyme-ligand complexes were placed in a cubic box with dimensions of approximately 75 Å containing approximately 40,000 TIP3P water molecules. The dimensions of the box were defined by a minimum 10 Å distance between the face of each side of the box and the enzyme-substrate complex. Although there is a degree of randomness in the orientation of the water molecules placed in this step, they are created using a grid resulting in an unphysical pattern to the water's positions. To enforce an overall neutral charge on the system our protocol randomly replaces solvent molecules with chloride or sodium counterions. In the case of *Is*-PETase, which has an overall charge of +7, seven chloride ions were introduced.

3.2.2.3 Equilibration and Ensembles

Following the protocol detailed by Justin Lemkul[80], three short steps were used to equilibrate the water molecules placed in the previous steps: an energy minimisation step, a NVT ensemble simulation, and NpT ensemble simulation. This was followed by a longer production molecular dynamics simulation using the NpT ensemble.

In each of these steps, periodic boundary conditions in the x, y and z dimensions were applied to the edges of the solvent box, allowing solvent molecules and ions to travel out of one face of the simulation box and to reappear at the opposite face. The Verlet[81] cut-off scheme was used to generate the nearest neighbours list of atoms during these three calculations. Long range electrostatic interactions were calculated using the Particle-Mesh Ewald method[82]. The LINCS/SHAKE constraints were applied to hydrogen atoms in each of these simulations. For the NVT, NPT and production MD simulations, three temperature coupling groups were used: one containing the protein, another with containing the solvating water molecules and counterions and a final group

containing the ligand. A temperature of 300 K was set for each temperature coupling group in all simulations.

The energy minimisation step was performed using Gromacs' in-built steepest decent integrator with an energy tolerance of 1000 kJ mol^{-1} and an energy step size of 0.01 kJ mol^{-1}. NVT and NPT simulations were both performed for 100 ps and were performed with position restraints placed upon all heavy atoms. The NPT ensemble simulation used the Berendsen method[83] for pressure coupling, while the longer MD simulation used the Parrinello-Rahman method[84]. The production MD simulation was run for 1 ns without position restraints on any atoms.

Geometries were extracted from snapshots from the production MD simulation, the first at 100 ps, the next at 150 ps then at 150 ps intervals (100ps, 150 ps, 300 ps, 450 ps, 600 ps, 750 ps and 900 ps). Geometries were not taken from snapshots before 100 ps as a large reduction in the overall system energy was observed in the first 100 ps associated with a relaxation of the orientation of the solvating water molecules. A further steepest decent energy minimisation step was then applied to these snapshot geometries; this was performed as a pre-optimisation to speed up subsequent QM/MM geometry optimisations.

3.2.3 Hybrid Quantum Mechanical/ Molecular Mechanical Calculations

3.2.3.1 Choice of Density Functionals and Basis Sets

For QM/MM geometry optimisations and QM/MM nudged-elastic band calculations, the revPBE density functional[72] was used using the def2-SVP basis set[85], selected for their performance against the GMTKN55 database[73]and compatibility with our Chemshell software[86]. Initially the semi-empirical method AM1[87] was used for geometry optimisations. Whilst AM1 calculations were significantly faster than revPBE/def2-SVP calculations, it was found that AM1 is inadequate for describing hydrogen bonding. This is a known problem and is due to its neglect of both polarisation effects and electron correlation[88].

For QM/MM single-point calculations the DSD-PBEP86 density functional[57] with the ma-def2-TZVP basis set[85] was used, chosen for its performance against the GMTKN55 database[73].

3.2.3.2 Choice of QM region and Link atom Scheme

The QM region (see Figure 25) for all calculations was selected using the following principle: "Use the minimal amount of QM atoms to reduce calculation time whilst having enough such that energies and geometries can be calculated accurately"[66].

The ester functional group in each ligand is where the majority of bond-breaking and bond-making events occur throughout the proposed mechanism. These cannot be described using molecular

mechanics, necessitating the inclusion of this functional group in the QM region. The MD parameters for the ligands were produced using the webservice SwissParam[79] and are not suitable for detailed MD calculations. For this reason, the entirety of each ligand was included in the QM region. This causes the calculation to ignore the MD parameters in lieu of the more accurate QM description.

CHARMM parameters for amino acid residues split the residue into groups. Typically these are: The backbone carbonyl (carbonyl carbon and carbonyl oxygen), the backbone (α-carbon, its proton, amide nitrogen and its proton) and the side-chain. The net charge of any group must always be an integer; thus, the QM/MM boundary must be made in-between these groups.

The side chains of the residues of the catalytic triad (Ser131, Asp177 and His208) were included in the QM region. Throughout the reaction pathway of both reactions NS1 and NS2 (see Figure 27), proton transfers occur between these residues, leaving groups on the ligand and the nucleophilic water in NS2. These proton transfers cannot be described by molecular mechanics necessitating these residues' inclusion in the QM region.

The backbone atoms of residues Tyr58 and Met132 were included in the QM region. Whilst these atoms do not undergo any bond-making or bond-breaking events during the two reactions, they are thought to provide significant charge-stabilisation to the various negatively charged tetrahedral intermediates along the reaction pathways through hydrogen bonding interactions.

To ensure that the QM/MM boundary has a minimal effect on the charges of QM atoms, only non-polar carbon-carbon bonds are 'cut' to create this boundary. To satisfy this requirement, the backbone carbonyls of Gly57, Trp130 and Ser131 and the backbone of Ser131 were included in the QM region.

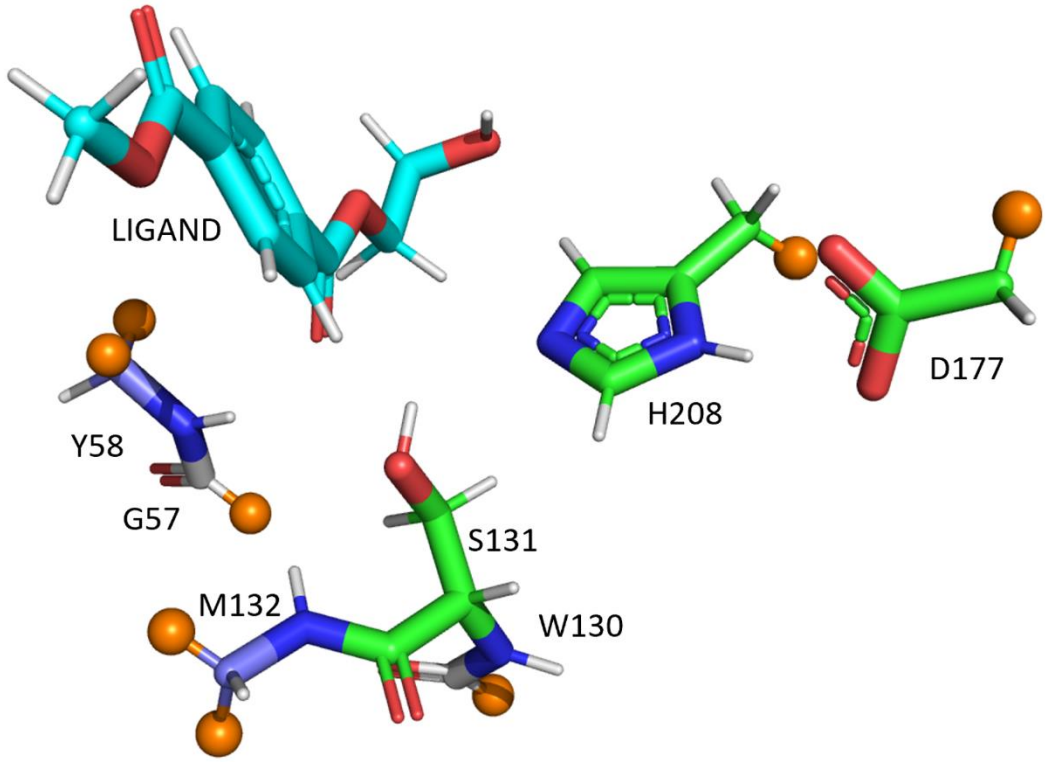


Figure 25: The QM region used in our QM/MM calculation. In this image, the ligand HEMT is shown, which was used in all calculations for the reaction NS1. For the reaction NS2, the ligand HMT was used. In all cases the entirety of each ligand was included in the QM region.

The link atom scheme[86] was used to treat the QM/MM boundary with the charge-shift scheme used to re-distribute charges in MM atoms adjacent to the QM/MM boundary.

3.2.3.3 Generation of Intermediate Geometries through Restrained Optimisations

To create geometries that represent local minima on the potential energy surface (PES) the DL-FIND optimiser[89] was used with the L-BFGS minimisation algorithm. For unrestrained QM/MM geometry optimisations, the density functional revPBE-D3(BJ)[72] with the def2-SVP basis set[85] was used to evaluate the energy of the QM region. This functional/basis set pair was chosen for a combination of speed and its high performance against the GMTKN55 benchmark database[73]. The CHARMM36 forcefield was used to evaluate the energy of the MM region. An active region was selected in the initial QM/MM calculation in the generation of each reaction profile; this list of atoms was then maintained throughout all subsequent QM/MM calculations. All atoms in the enzyme and ligand were included in the active region. Solvent atoms within 10 Å of CO_{LIG} (see Figure 26) were included in the active region.

The geometries of intermediates were produced using geometries of previous intermediates at starting points. A combination of harmonic potentials (hereafter called restraints) were placed

between specified atoms in the QM region during QM/MM geometry optimisations. These harmonic potentials caused the desired bonds to form or be broken in order to create a geometry for each desired intermediate.

It was found that in some cases the restraints caused unphysical behaviours in the TIP3P water molecules proximal to the ligand. This resulted in large solvent cavities above the ligand (i.e., no solvent molecules were directly interacting with a portion of the ligand). Initially a further QM/MM geometry optimisation was performed upon the output geometry of each restrained geometry optimisation; this was insufficient to remove these unphysical solvent cavities. The initial restraints used to produce intermediates RS and PS were found to produce the largest solvent cavities. To remove the solvent cavities from RS and PS, a further series of MD simulations (see section **3.2.2**) were used to completely re-solvate the enzyme-ligand complex. These MD simulations were performed with position restraints placed on all atoms in the protein and ligand to preserve the intermediate's geometry. The resultant geometry from this 'frozen' MD simulation was then re-optimised using an unrestrained QM/MM geometry optimisation as described above. For the production of all other intermediates, this treatment was inappropriate as the positions of the solvent molecules had to be consistent throughout the reaction profile in order to calculate energy barriers. Thus, the positions of water molecules were re-set to their positions prior to each restrained QM/MM geometry optimisation. This was followed by an unrestrained QM/MM geometry optimisation to produce the final geometry for each intermediate.

3.2.3.4 Generation of Transition State Geometries through the Nudged-Elastic-Band Method

The nudged-elastic-band method[59] (NEB) was used to generate geometries of the transition states. This method requires two consecutive intermediate geometries as inputs. It then generates a set of structures by interpolating in-between these two structures. A spring interaction between adjacent structures is applied, then the structures are optimised to generate the geometries along the minimum energy pathway (MEP) connecting the two input intermediates. The geometries of structures that are close to the saddle-point on the MEP are then optimised using the climbing-image protocol[60]. All NEB calculations were performed at the same level as our QM/MM geometry optimisations: The density functional revPBE(D3)-BJ[72] was used to describe the QM region using the def2-SVP[85] basis set. The MM region was described using the CHARMM36 forcefield.

3.2.3.5 Single-Point Energy Calculations

Using the geometries of the reaction intermediates and transition states (produced by unrestrained QM/MM geometry optimisations and NEB calculations respectively at the rev2PBE(def2-SVP)/CHARMM36 level) energies were re-calculated at a higher level of theory. These single-point

calculations were performed using the double-hybrid DSD-PBEP86 density functional with the polarised triple-zeta ma-def2-TZVP basis set. The introduction of exchange-correlation terms is expected to increase the accuracy of the calculated energies and thus a more accurate calculation of the energy barriers. As with all our QM/MM calculations, the CHARMM36 forcefield[90] was used to describe the MM region.

3.3 Results and Discussion

In September 2022, we published our first paper in *Chemistry, A European Journal*[48]. This section will discuss the results of this study in greater detail.

3.3.1 Atom naming Conventions

In the following sections, key atoms in the QM region will be referenced repeatedly. For the sake of conciseness, the following naming convention will be used to reference these key atoms.

It is worth noting that the atoms O_{SOLV} and OE_{LIG} are the same atoms in the reaction NS2. When this oxygen is bound to the HMT ligand it will be referred to as OE_{LIG}. While this oxygen is bound to two protons (i.e., it is part of a water molecule) it will be referred to as O_{SOLV}.

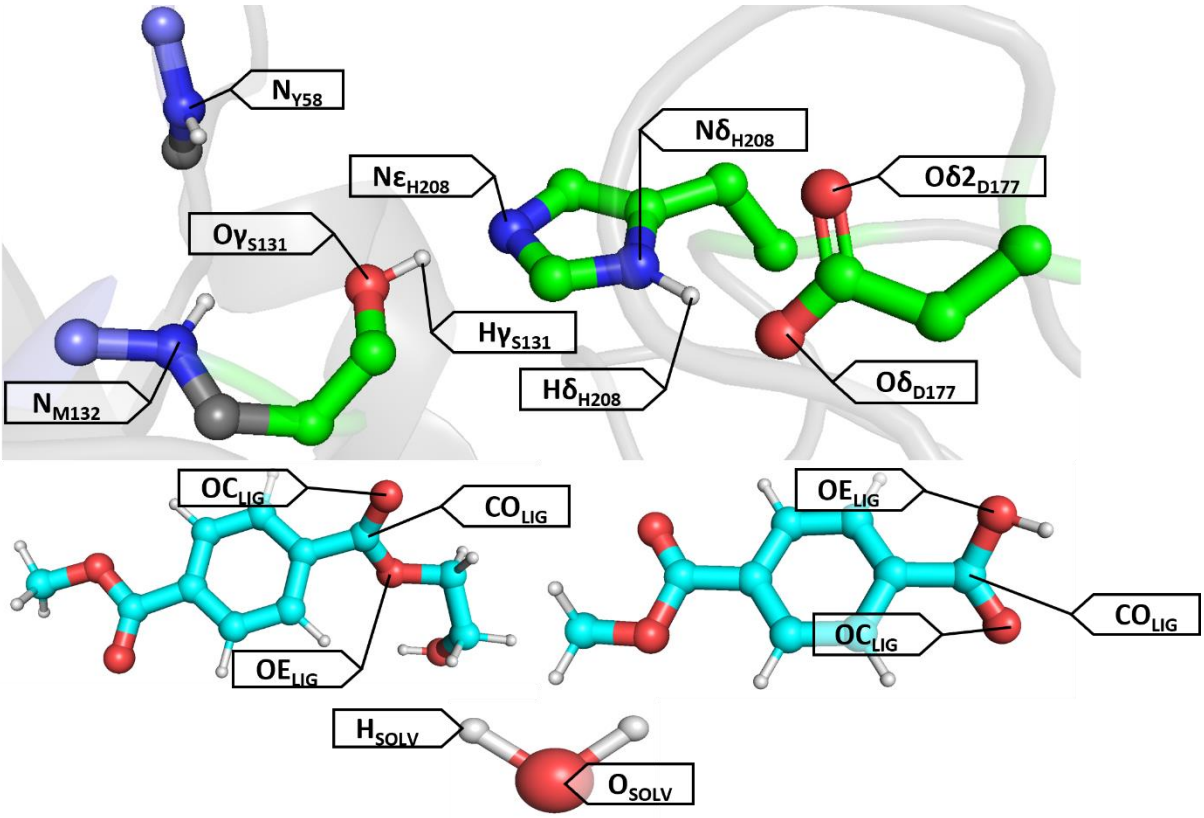


Figure 26: Naming conventions for key atoms throughout the NS1 and NS2 reactions. Protons will be referred to by their bonded heteroatom. For reactions NS1 and NS2 different ligand molecules are used; HEMT (middle left) and HMT (middle right). As in the combined reaction profile HEMT is converted to HMT, the same atom names for key atoms are used for both ligands. The water molecule (bottom) refers only to the water molecule that undergoes a nucleophilic attack in the Nuc2 reaction step.

3.3.2 Proposed Mechanism of Is-PETase

The mechanism of *Is*-PETase has been assumed to follow the same steps as all α/β hydrolases. This reaction comprises of two consecutive nucleophilic substitutions at an ester carbonyl of PET (see Figure 27). The initial substitution cleaves the ester bond and results in an enzyme-bound intermediate. The second substitution hydrolytically cleaves the enzyme-substrate bond, releasing the product and regenerating the enzyme to its original catalytic state. The mechanism is described as a series of four consecutive steps:

The negatively charged side chain of Asp177 deprotonates the $N\varepsilon_{H208}$ transferring $H\varepsilon_{H20}$ to $O\delta_{177;}$ this triggers a rearrangement of the imidazole ring of His208 causing $O\gamma_{S131}$ to be deprotonated by the $N\varepsilon_{H208}$. The deprotonated $O\gamma_{S131}$ attacks the ester carbonyl of the PET substrate creating a tetrahedral intermediate, forming a new bond with CO_{LIG}. The negative charge that develops on OC_{HEMT} is stabilised through hydrogen bonding with the backbone nitrogen atoms N_{Y58} and N_{M132}. The driving force for this nucleophilic attack is the relocation of a negative charge from the side chain of Asp177 to the more δ-positive oxyanion hole where it receives greater stabilisation.

The negatively charged tetrahedral intermediate collapses, regenerating the double bond between CO_{LIG} and OC_{LIG} and breaking the bond between CO_{LIG} and OE_{LIG}. This formally places the negative charge on OE_{LIG} of the substrate which deprotonates $N\varepsilon_{H208}$; this converts the functional group of OE_{LIG} from an ester oxygen to a hydroxyl group. A rearrangement in the imidazole ring of His208 triggers the deprotonation of $O\delta_{D177}$ by $N\varepsilon_{H20}$, regenerating the original protonation states of the catalytic residues Asp177 and His208. This results in an ester bond between the enzyme and substate via $O\gamma_{S131}$ and CO_{LIG} and the elimination of ethylene glycol from the substrate as a leaving group.

The negatively charged side chain of Asp177 deprotonates the $N\varepsilon_{H20}$ via $O\delta_{D177}$; this triggers a rearrangement of the imidazole ring of His208. $N\varepsilon_{H208}$ deprotonates a solvent water molecule. The activated water molecule nucleophilically attacks the enzyme-bound CO_{LIG} resulting in a negatively charged tetrahedral intermediate. The negative charge that develops on OC_{LIG} in this intermediate is stabilised through hydrogen bonding with the backbone nitrogen atoms N_{Y58} and N_{M132}.

The negatively charged tetrahedral intermediate collapses, regenerating the double bond between CO_{LIG} and OC_{LIG}, and breaking the ester bond between CO_{LIG} and $O\gamma_{S131}$ and converting the substrate's ester group into a carboxylic acid group. The negative charge is transferred to the $O\gamma_{S131}$ which deprotonates the $N\varepsilon_{H208}$. A rearrangement in the imidazole ring of His208 triggers the deprotonation of $O\delta_{D177}$ by the $N\varepsilon_{H20}$, regenerating the original protonation states of the catalytic residues Ser131

Asp177 and His208 readying the enzyme for another catalytic cycle.

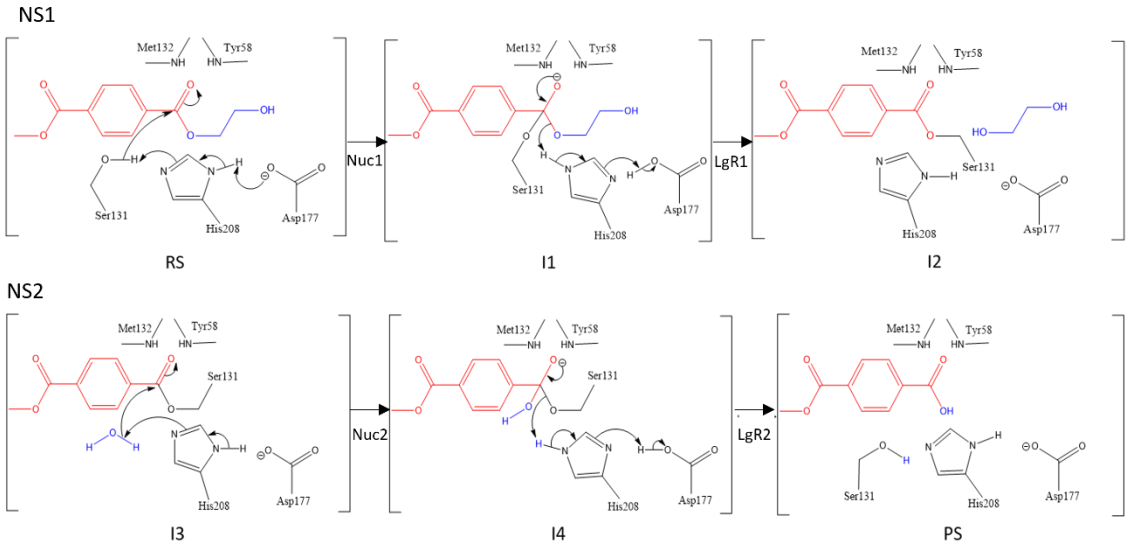


Figure 27: Proposed mechanism for the degradation of HEMT via Is-PETase. The overall mechanism has been split into two nucleophilic substitution reactions: NS1 and NS2.

As of the outset of this project, the importance of the catalytic residues Ser131, Asp177 and His208 had been confirmed by mutagenesis experiments where the replacement of any of these residues with an inert alanine residue abolished the enzymes' activity. The veracity of this exact mechanism had not been examined in detail through either chemical techniques such as deuteration experiments and radioisotope labelling or through computational modelling using QM/MM methods.

For the purposes of this research, the overall reaction catalysed by *Is*-PETase has been separated into two independent nucleophilic substitution reactions. This was necessary as generation of the geometry of intermediate I3 from I2 involved the replacement of the EG leaving group with a water molecule. Initially this was performed (by replacing the majority of the EG molecule with a proton) and produced a water molecule that was ideally orientated to initiate the second nucleophilic attack step. Unfortunately, the replacement of the EG with a water molecule created a solvent void where the removed atoms from larger EG molecule were located. This solvent void was not 'filled' using geometry optimisations at the QM/MM level. Filling this solvent void would be possible using a subsequent MD simulation with the protein and ligand atoms frozen in place. this method was not used for the following reasons: MD simulations involving an enzyme-bound ligand (as exists in I2 and I3) require a parameter set for the ligand, which is now chemically bonded to Ser131; CHARMM will therefore interpret this as a new amino acid. The parameterisation of novel amino acids is an extensive and difficult task. The second reason MD simulations could not be used to connect the

geometries I2 and I3 is that this would result in a change of orientation and position of all the water molecules in the solvent box. This would create a large energy difference between geometries I2 and I3 and as such would not achieve a continuous energy profile across the two nucleophilic substitutions. An alternative method would have been to use the geometry of I2 as the geometry of I3; this would necessitate the calculation of the reaction pathway of NS2 in the presence of the EG leaving group. This method was unfeasible as a significant rearrangement of solvating water molecules as well as the leaving group EG is required such that a water molecule can be deprotonated by His208, initiating the Nuc2 step. Ideally, this ligand replacement could have been achieved through a series of constrained MD simulations using a technique such as umbrella sampling[91]. Unfortunately, this approach was not available to us due to lack of forcefield parameters for the enzyme-bound ligand. The dissociation of EG and its subsequent replacement by a nucleophilic water molecule was studied by a different group[30]. It was found that this process is not accompanied by a significant energy barrier. We have discussed this study in greater detail in section 3.1.1.4.

The separation of nucleophilic substitutions NS1 and NS2 reveals that the catalytic cycle of *Is*-PETase is best viewed as five consecutive steps:

1. Substrate association
2. The chemical steps in reaction NS1
3. Dissociation of products of NS1 and association of a nucleophilic water molecule
4. The chemical steps in reaction NS2
5. Dissociation of the products of NS2

The energy profiles for the chemical steps NS1 and NS2 have been calculated using QM/MM methods in this work. To assess the energy barriers of the three remaining steps, less accurate methods such as molecular docking or molecular dynamics could be used. The results from these methods would however be non-comparable to the results from our QM/MM methods due to the differences in techniques.

3.3.3 Generation of Reaction Intermediates and Transition States

To determine the energy barriers associated with each reaction step of reactions NS1 and NS2, the geometries of the intermediates were first produced through a series of restrained and unrestrained QM/MM optimisations (discussed in section 3.2.3.3). Geometries for transition states that connect these intermediates were produced using NEB calculations (discussed in section 3.2.3.4). In order to sample the conformational space of this reaction, snapshots of an initial MD simulation were taken at regular intervals and used as starting structures. From each of the starting structures, geometries

for intermediates and transition states were generated in parallel, resulting in several different reaction profiles for both NS1 and NS2.

In the following sections, a general method is outlined for the production of geometries for intermediates and transition states along the reaction pathways of NS1 and NS2. The following schemes were applied for each snapshot and required modification in some cases. Important modifications to the scheme will be discussed, but details of the production of each intermediate and transition state across each reaction profile will be omitted for the sake of brevity.

3.3.3.1 Generation of Intermediate and Transition State Geometries for Nucleophilic Substitution 1

The intermediate RS (see Figure 27) represents the enzyme-bound HEMT substrate in an optimal position and orientation for the subsequent nucleophilic attack of Ser131 upon its ester carbonyl. This requires the ester carbonyl oxygen to be forming hydrogen-bonding interactions with the backbone nitrogen of Tyr58 and Met132 in the oxyanion hole. This reduces the electron density situated on the carbonyl carbon making it more electrophilic, activating it toward nucleophilic attack. To generate this geometry, starting with a snapshot structure from the MD simulation a QM/MM geometry optimisation with two restraints of 2.9 Å (the typical length of a hydrogen bond) applied between the ester carbonyl oxygen of HEMT and the backbone nitrogen of Tyr58 and Met132 (see Figure 28). The resulting structure was re-solvated using a series of MD simulations with position restraints placed on all ligand and protein atoms to remove the solvent void created by the restraints. A further unrestrained QM/MM geometry optimisation was performed to produce

the final geometry of RS.

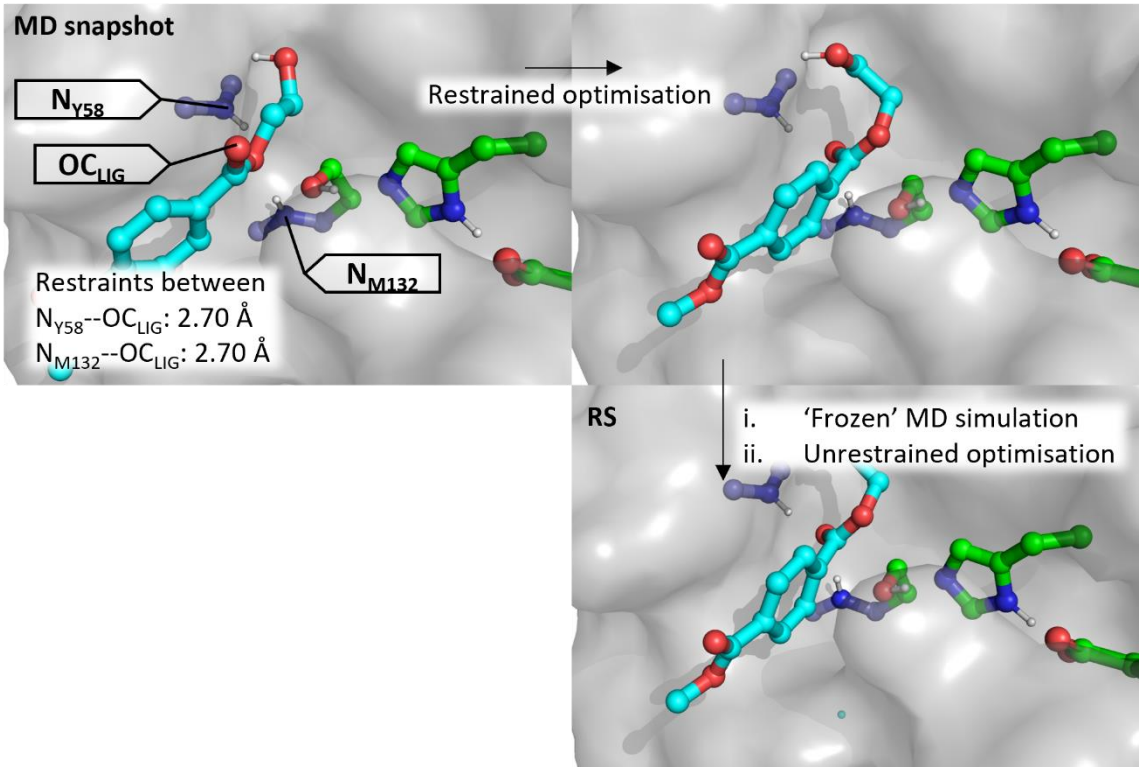


Figure 28: Scheme for production of RS geometries using restrained optimisations followed by frozen MD simulations then unrestrained geometry optimisations.

The intermediate I1 represents the tetrahedral (at the reactive ester carbonyl) intermediate resulting from the nucleophilic attack of the γ-oxygen of Ser131 on the ester carbonyl of HEMT. The geometry for I1 was generated starting from the geometry for RS. A QM/MM geometry optimisation was performed on this structure with a restraint of 1.43 Å (the typical bond length for a C-O single bond in an ester) between the γ-oxygen of Ser131 on the ester carbonyl of HEMT. The coordinates for water molecules were re-set to their positions in RS, then a further unrestrained QM/MM geometry optimisation was performed to generate the final geometry of I1 (see Figure 29).

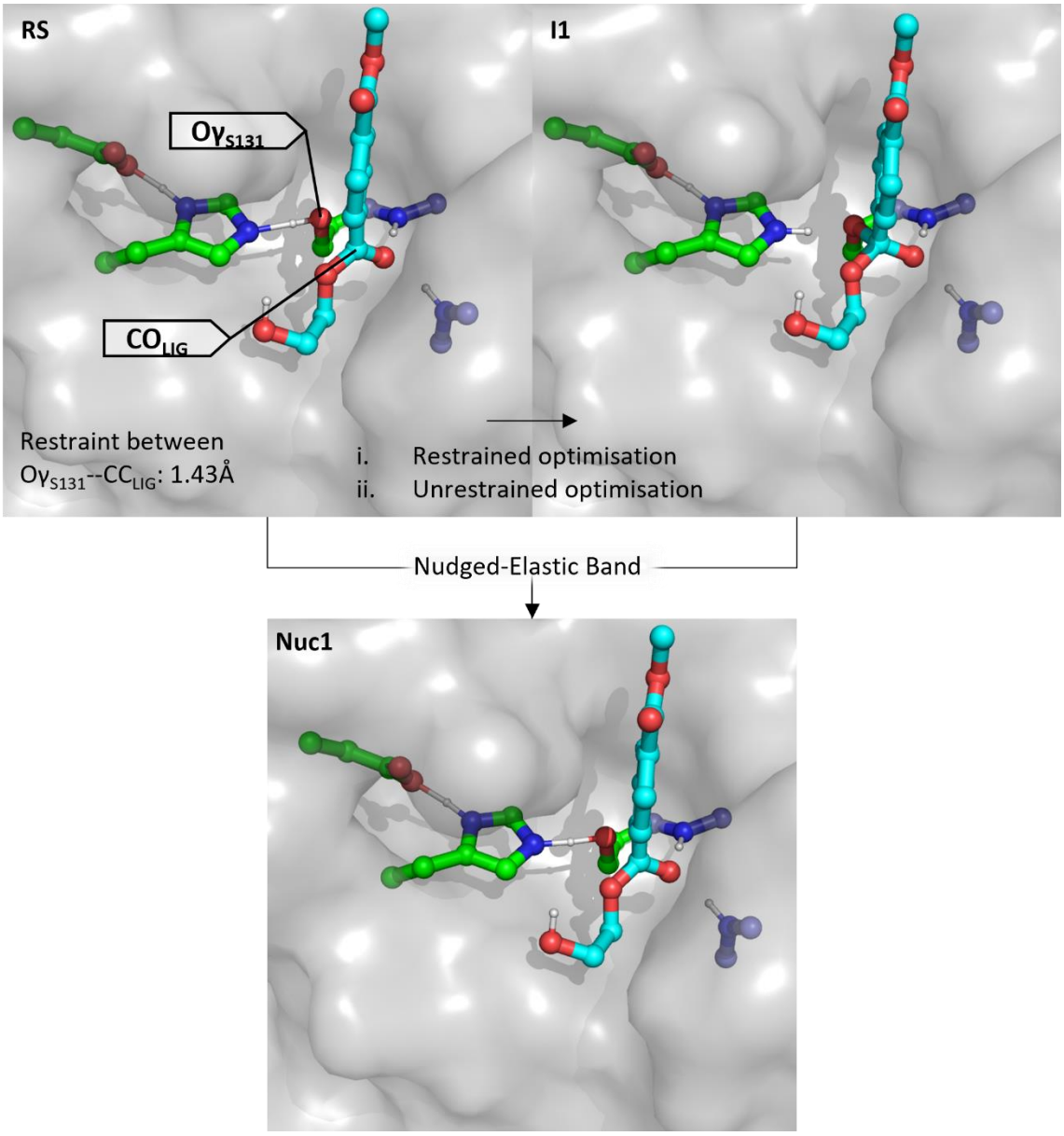


Figure 29: Scheme for the production of I1 and Nuc1 geometries using restrained and unrestrained geometry optimisations, followed by nudged-elastic band calculations.

The intermediate I2 represents the trigonal (at the reactive ester carbonyl) intermediate following the regeneration of the ester carbonyl and the loss of EG as a leaving group. During this loss of EG, the ester oxygen deprotonates the δ-nitrogen of His208. Starting from the structure of I1, a QM/MM geometry optimisation was performed with a restraint of 1 Å (the typical distance for an O-H bond) between the δ-proton of His208 and the ester oxygen of HEMT. During this optimisation, the ester carbonyl spontaneously regenerated and the new leaving group EG moved away from the ester carbonyl. The coordinates for water molecules were re-set to their positions in I1, then a further unrestrained QM/MM geometry optimisation was performed to generate the final geometry of I2 (see Figure 30).

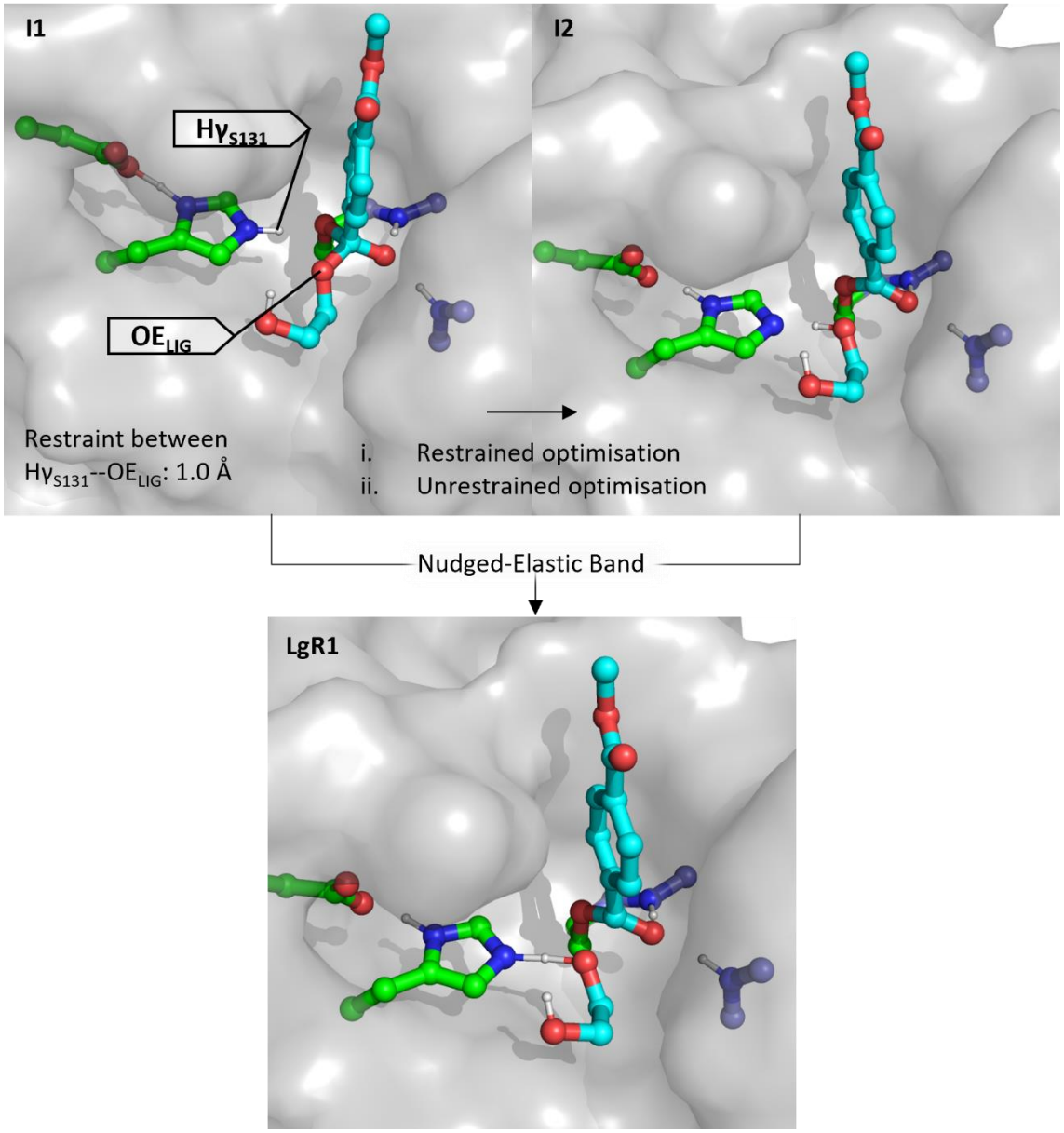


Figure 30: Scheme for the production of I2 and LgR1 geometries using restrained and unrestrained geometry optimisations, followed by nudged-elastic band calculations.

The transition state Nuc1 was produced using a NEB calculation using the geometries of RS and I1 as input structures. The highest energy structure along the MEP was then selected as the final geometry for Nuc1 (see Figure 29).

The transition state LgR1 was produced using a NEB calculation using the geometries of I1 and I2 as input structures. The highest energy structure along the MEP was then selected as the final geometry for LgR1 (see Figure 30).

3.3.3.2 Generation of Intermediate and Transition State Geometries for Nucleophilic Substitution 2

The intermediates and transition states associated with NS2 were generated in reverse order with respect to the reaction profile. The molecule HMT (which is the hydrolysis product of HEMT) was

first placed in the active site with molecular docking. The resulting enzyme-product complex was then solvated and subjected to a MD simulation. Snapshots at 100, 150, 300, 450, 600, 750 and 900 ps were taken from the MD simulation. Each snapshot was used as a starting structure to generate a separate reaction pathway.

The intermediate PS represents the enzyme-bound product, proceeding from the breaking of the bond between the ester carbonyl carbon of HMT and the γ-oxygen of Ser131 during the LgR2 step. The transition state LgR2 is stabilised by hydrogen bonding interactions between the ester carbonyl oxygen of HMT and the backbone nitrogen of Tyr58 and Met132, thus the PS geometry is expected to retain these hydrogen bonding interactions. The structure of PS was generated from the MD snapshot by performing a QM/MM geometry optimisation with a pair of restraints of 2.9 Å (the typical length of a hydrogen bond) between the ester carbonyl oxygen of HMT and the backbone nitrogen of Tyr58 and Met132. The resulting structure was then re-solvated using a further MD simulation with position restraints applied on all protein and ligand atoms. A snapshot from this frozen MD simulation was taken (at a timestep after 100 ps) and then reoptimized using an unrestrained QM/MM geometry optimisation to produce the final geometry of PS (see Figure 31).

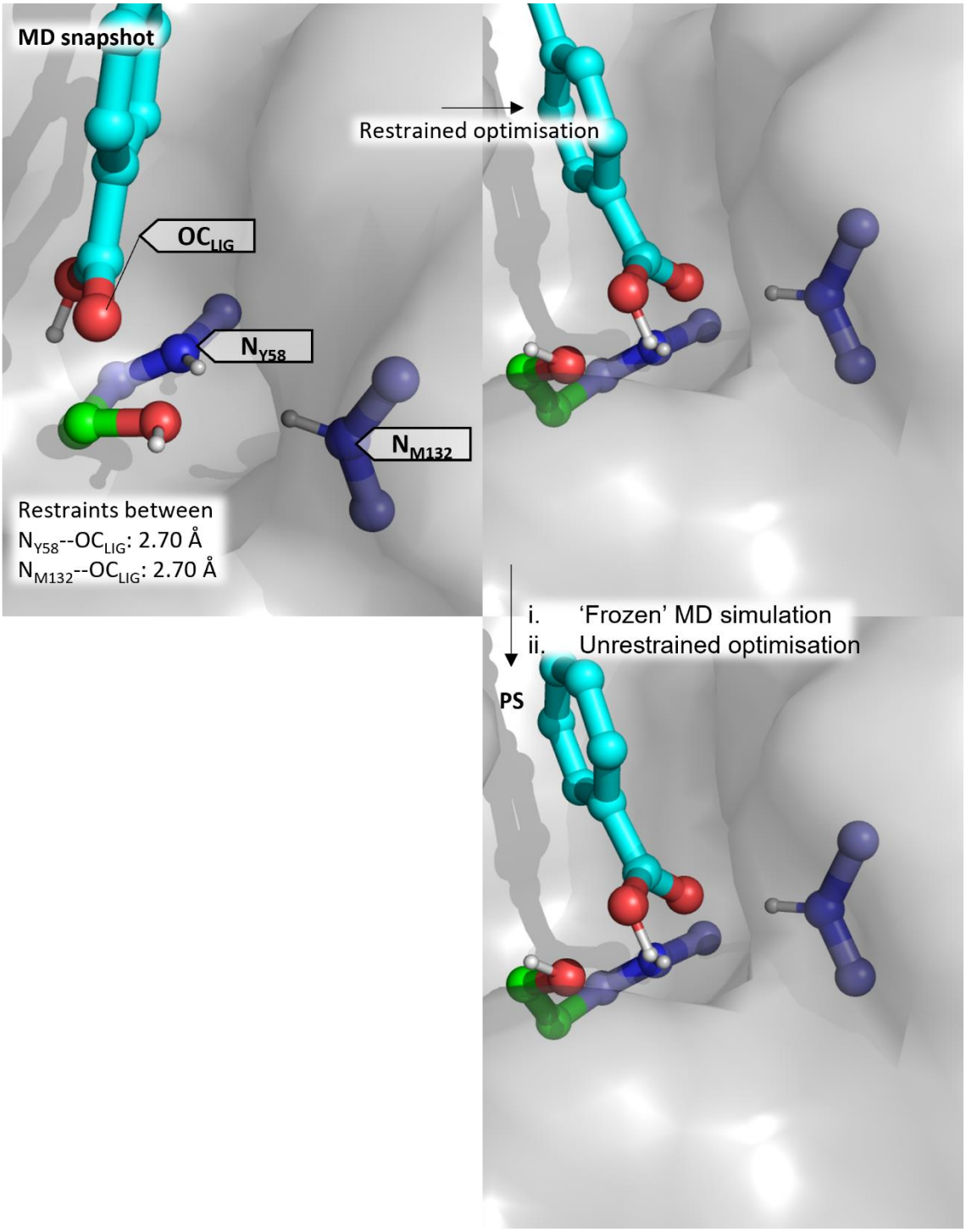


Figure 31: Scheme for production of PS geometries using restrained optimisations followed by frozen MD simulations then unrestrained geometry optimisations.

The intermediate I4 represents the tetrahedral (at the reactive ester carbonyl) intermediate following the nucleophilic attack of an activated water molecule at the ester carbonyl of HMT in the Nuc2 step. From our reverse perspective of the NS2 reaction, I4 is the result of the nucleophilic attack of the activated γ-oxygen at the ester carbonyl, starting from the PS geometry (i.e., the LgR2 step in reverse is a nucleophilic attack). The geometry for I4 was thus generated starting from the

geometry of PS, a QM/MM geometry optimisation was applied with a restraint of 1.43 Å (the typical length of a C-O single bond in an ester bond in our optimised geometries) between the γ-oxygen of Ser131 and the ester carbonyl carbon of HMT. The coordinates for water molecules were re-set to their positions in PS, then a further unrestrained QM/MM geometry optimisation was performed to generate the final geometry of I4 (see Figure 32).

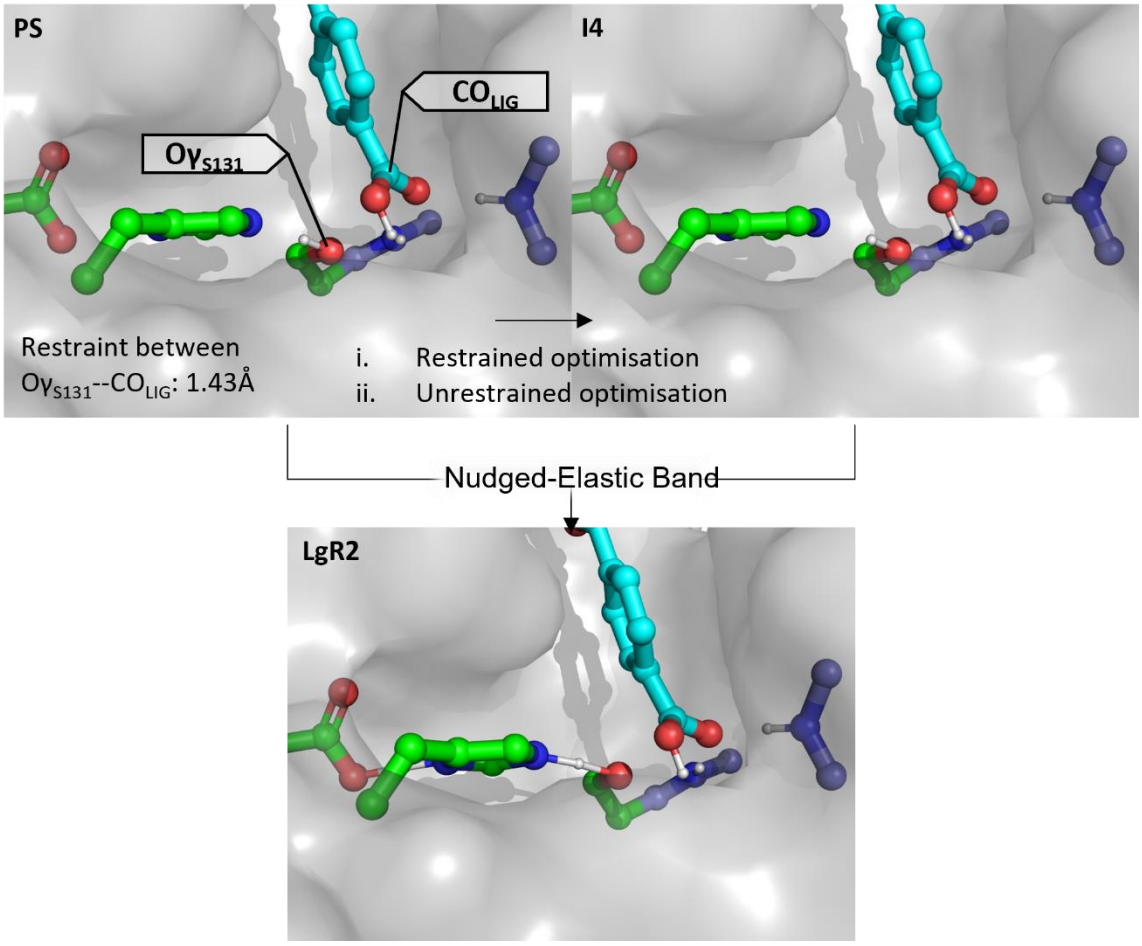


Figure 32: Scheme for the production of I4 and LgR2 geometries using restrained and unrestrained geometry optimisations, followed by nudged-elastic band calculations.

The intermediate I3 represents the trigonal (at the reactive ester carbonyl) intermediate preceding the nucleophilic attack of an activated water molecule at the ester carbonyl of HMT. From our reverse perspective of the NS2 reaction, this can be seen as the reformation of the ester carbonyl in I4 and the loss of water as a leaving group (i.e., the Nuc2 step in reverse is a condensation step). The geometry for I3 was thus generated starting from the geometry of I4, a QM/MM geometry optimisation with a restraint of 1 Å (the typical bond length of an O-H bond in our optimised geometries) between the δ-proton of His208 and the oxygen of the water 'leaving group'. The coordinates for water molecules were re-set to their positions in I4, then a further unrestrained

QM/MM geometry optimisation was performed to generate the final geometry of I3 (see Figure 33).

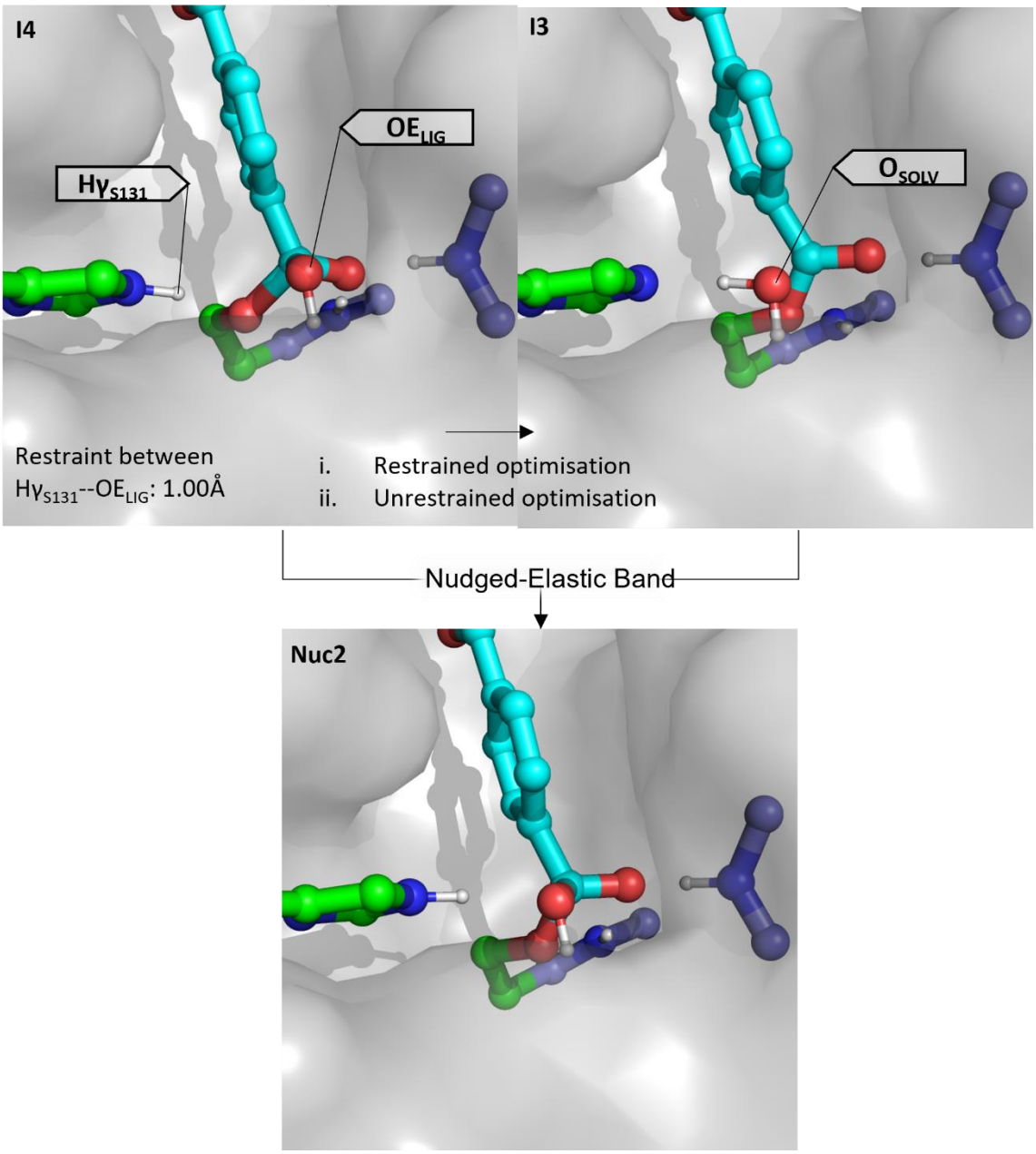


Figure 33: Scheme for the production of I3 and Nuc2 geometries using restrained and unrestrained geometry optimisations, followed by nudged-elastic band calculations.

The transition state Nuc2 was produced using a NEB calculation using the geometries of I3 and I4 as input structures. The highest energy structure along the MEP was then selected as the final geometry for Nuc2 (see Figure 33).

The transition state LgR2 was produced using a NEB calculation using the geometries of I4 and PS as input structures. The highest energy structure along the MEP was then selected as the final geometry for LgR2 (see Figure 32).

3.3.4 The Energy Profile of Is-PETase catalysed reactions

Energy profiles for the reactions NS1 and NS2 were calculated using the protocol described in section 3.1. Seven snapshots from the initial MD directory were used as starting points to generate geometries along the reaction profile of NS1, while six snapshots were used for NS2 (initially seven were used for NS2, but a reaction profile with suitably low activation barriers was not found when starting from the snapshot at 900 ps).

The use of an explicit solvent model presents us with a problem: The absolute energy of each geometry is largely dependent on the orientations of its solvent water molecules, but the orientation of these water molecules has very little effect on the activation barriers of the reaction in question. The result of this is that the absolute energies of the intermediates and transition states calculated from different snapshots are not directly comparable as they have different solvent environments and thus have vastly different energies (up to 3000 kJ mol^{-1} differences were found between equivalent geometries). This large energy difference due to solvation obscures energy differences that may be due to the reaction of interest. Although geometries generated from different snapshots have differing solvent environments, geometries generated from the same snapshot have a largely shared solvation environment as the orientation of solvent water molecules changes only slightly during the sequence of QM/MM optimisations used to generate the intermediates and transition states. By expressing the energy of each geometry relative to the energy of another geometry that shares the same solvation environment, the effect of the solvent environment was removed from the energy of geometries. The result of this is that these relative energies are now comparable across different snapshots. The first intermediate in each reaction pathway (intermediate RS in reaction NS1 and intermediate I3 in reaction NS2) was chosen to be this reference. As such, the energies for intermediates RS and I3 have been set to zero. This makes them appear to be energetically equivalent although it is likely that there are important differences in their geometries that contribute towards an energy difference. As we are primarily interested in the activation barriers of the two reactions, the relative energies of intermediates across different snapshots are unimportant.

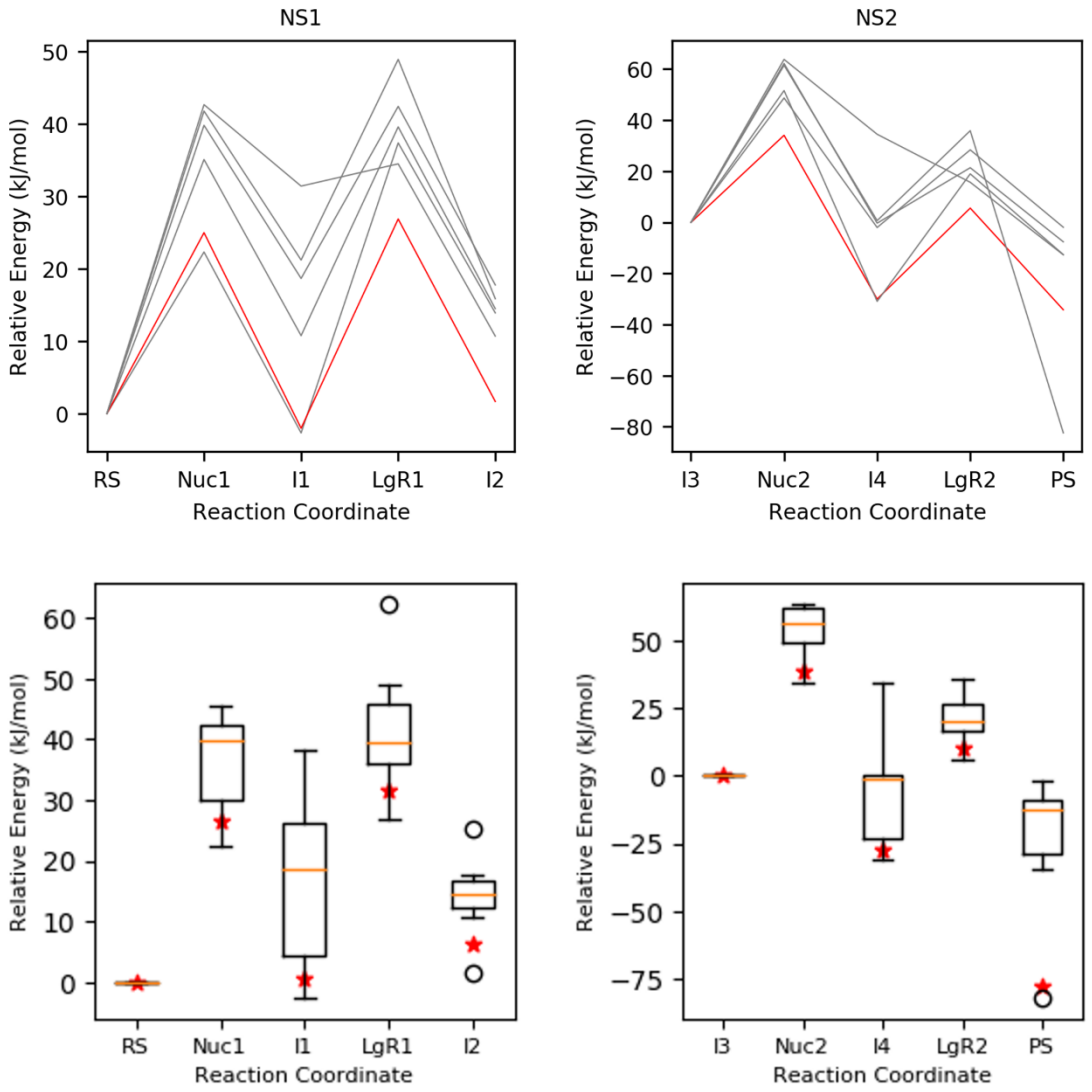


Figure 34: Energies for reaction profiles calculated for NS1 (top left) and NS2 (top right). The lowest energy profiles for NS1 and NS2 have been displayed in red. Box plots for the energies of intermediates and transition states along reaction profiles for NS1 (bottom left) and NS2 (bottom right). The arithmetic average for each geometry is displayed as an orange line, while the Boltzmann-weighted average is displayed as a red star.

Using the reaction profiles we have generated (see Figure 34), several methods can be used to choose our rate-determining step. One method is to select the reaction profiles (one for NS1 and one from NS2) with the lowest overall barrier heights, then by selecting the reaction step with the highest energy activation barrier from each of these lowest energy reaction profiles. The overall rate-determining step is then selected as the one of these two steps with the highest energy barrier. This technique will be referred to as the "lowest energy profile" method. According to transition state theory, if the energy of the preceding intermediate is higher than the intermediate preceding it (i.e., the starting intermediate in the previous reaction step), then the activation energy is calculated as the energy of the transition stated subtracted from energy of the lower energy intermediate.

Another method that could be used to determine the rate-determining step is via arithmetic averaging. In this method, the average energy of each activation barrier is calculated; the rate-determining step is then selected as the step with the highest average energy barrier. A similar method is to select the rate-determining step using a Boltzmann-weighted averaging. This method is preferable to arithmetic averaging as it is more sensitive to low-lying activation barriers than high energy barriers. This better reflects a chemical system, which will preferentially use low energy reaction pathways over high energy ones.

Using the "lowest energy profile" method, the LgR1 step is selected as the rate-determining step for the reaction NS1. In six out of seven reaction pathways (each starting from a different snapshot of the MD trajectory) for the reaction NS1, the rate-determining step was calculated to be LgR1.The range in activation energies for this reaction was calculated to be between 28.91 kJ mol^{-1} and 62.4 kJ mol^{-1}. Rationalisations that relate these energies to their respective geometries will be explored in later in section 3.3.5. In the remaining reaction pathway, the rate-determining step was found to be Nuc1, with an activation energy of 42.69 kJ mol^{-1}.

In five out of six reaction pathways for the reaction NS2, the rate-determining step was calculated to be Nuc2. The range in activation energies for this reaction was calculated to be between 48.65 kJ mol^{-1} and 63.77 kJ mol^{-1}. In the remaining reaction pathway, the rate-determining step was found to be LgR2 with an activation barrier of 35.61 kJ mol^{-1}. It is worth noting that in this case the energy difference between the LgR2 and Nuc2 is very small (1.51 kJ mol^{-1}) and may be within the margins of error for our methods (ca. 4 kJ mol^{-1} is considered to be as accurate as DFT methods get).

The lowest energy barrier found for the NS1 reaction was calculated to be 28.91 kJ mol^{-1}. The lowest energy barrier found for the NS2 reaction was calculated to be 35.61 kJ mol^{-1}. Using the "lowest energy profile" method, this makes the overall rate-determining step for the degradation of PET via *Is*-PETase LgR2 with an activation energy of 35.61 kJ mol^{-1}.

Using the arithmetic averaging, the highest average energy barrier is associated with reaction step Nuc2 with an activation energy of 53.7 kJ mol^{-1}. Using the Boltzmann-weighted averaging technique, the rate-determining step is also selected as Nuc2, with an activation energy of 38.6 kJ mol^{-1}, while the activation energy for LgR2 is 37.8 kJ mol^{-1}.

The use of either the "lowest energy profile" method or Boltzmann-weighted averaging is appropriate for selecting the rate-determining step. These methods select either the Nuc2 or LgR2 steps as the rate-limiting step in our overall reaction. In each case, the difference in activation

energy between the Nuc2 and LgR2 steps is less than 2 kJ mol^{-1}. For these reasons, we believe it is appropriate to find both Nuc2 and LgR2 steps to be rate-limiting.

For one of our reaction profiles for NS2 (see Figure 34), an anomalously low energy was observed for the product state, compared to our other PS energies. Upon further inspection of this geometry, we found that a proton transfer had occurred between OE_{LIG} and $O\gamma_{S131}$. This effect and its implications are discussed more fully in section 3.3.6.

3.3.4.1 A Note on the Kinetics of PETase Enzymes

To verify our calculated energy barrier, ideally a comparison would be made with experimental data. In most enzymatic systems this can be done by using the Arrhenius equation:

$$k = Ae^{\frac{-E_A}{RT}}.$$

Equation 77: The Arrhenius Equation.

In Equation 77, k is the rate constant, A is a pre-exponential factor, E_A is the activation energy of the reaction, R is the universal gas constant and T is the reaction temperature. Unfortunately, the rate constant for the degradation of polymeric PET via *Is*-PETase has not been experimentally determined. The enzymatic degradation of polymeric PET is a heterogenous process whereby insoluble PET fibres are degraded by a solvated enzyme. It is therefore necessary to treat this reaction as being under substrate-limiting conditions. This makes traditional Michaelis-Menten kinetics (which require a large excess substrate) inappropriate for the study of *Is*-PETase. An alternative method for describing PETase kinetics has been proposed to describe the adsorption, hydrolysis and product dissociation processes that occur[92]:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} EP \underset{K_p}{\rightleftharpoons} E + P.$$

Equation 78: Scheme showing a model for describing PETase kinetics.

In Equation 78, the adsorption of the enzyme to an area of PET where a hydrolysable ester bond is available is governed by the rate constant k_1 and the reverse process by the rate constant k_{-1}. The actual hydrolysis process is assumed to be irreversible and is governed by the rate constant k_2. The dissociation of the products is governed by the dissociation constant K_p for all products of the reaction.

In this work, we have calculated the energy barriers for the hydrolysis steps, and as such are interested in the k_2 value for this reaction. In a similar manner to classical Michaelis-Menten kinetics, the initial reaction rate is related to the concentration of the enzyme-substrate complex:

$$v_0 = k_2[ES].$$

Equation 79: Initial rate expressed in terms of enzyme-substrate complex concentration.

In Equation 79, v_0 is the initial rate and $[ES]$ is the concentration of the enzyme-substrate complex. Using a Langmuir-type isotherm model, the concentration of the enzyme-substrate complex can be related to the enzyme concentration:

$$[ES] = A_0 \frac{\mathrm{K}_A[E]}{1 + K_A[E]}.$$

Equation 80: Enzyme-substrate complex concentration expressed in terms of enzyme concentration.

In Equation 80, A_0 is the concentration of the initial surface area of the polymer, K_A is the adsorption constant for the association of the PETase enzyme to the PET polymer and $[E]$ is the enzyme concentration. Equation 79 and Equation 80 can be combined to give:

$$v_0 = k_2 A_0 \frac{\mathrm{K}_A[E]}{1 + K_A[E]}.$$

Equation 81: Initial rate expressed in terms of enzyme concentration.

A further improvement has been made to Equation 4 by including a term that considers the molar density of ester bonds in the substrate (ρ_{EB}):

$$v_0 = k_2 \rho_{EB} A_0 \frac{\mathrm{K}_A[E]}{1 + K_A[E]}.$$

Equation 82: Initial rate expressed in terms of enzyme concentration with an additional term that describes the density of hydrolysable ester bonds.

It is possible to improve Equation 5 further, but the added complexity makes comparison with experimental data more difficult. This approach has been applied to several PETase type enzymes, and have been used as metrics to measure the impact of single-point mutations on the enzyme's activity[93]. Unfortunately, this approach has not yet been applied to *Is*-PETase, so direct comparisons with the data presented in our work cannot be made.

3.3.4.2 Comparison of Our Study with other QM/MM studies of the mechanism of Is-PETase

At a late point in my PhD, five papers that applied QM/MM methods to the study of the catalytic mechanism of *Is*-PETase were published[29–33]. The methodologies and results of these studies have been discussed previously in sections 3.1.1.2 to 3.1.1.6, with a summary in section 3.1.1.7. Across these five studies, a wide array of methods have been applied, resulting in the key findings of each study conflicting with each other. This section will discuss our methods and results in the context of these five studies.

The most striking disagreement in the literature regarding the mechanism of *Is*-PETase is whether the enzyme's catalysis proceeds via a four-step or two-step mechanism. Four of the previous studies[29,31] concluded that *Is*-PETase proceeds via the four-step mechanism, consistent with the canonical mechanism for the α/β-hydrolase family of enzymes, of which *Is*-PETase is a member. The study published by Jerves *et al.*[30] found no stable minima associated with tetrahedral intermediates I1 and I4 (see Figure 27). Our study's results support the theory that *Is*-PETase follows the four-step reaction mechanism. In each of our generated reaction profiles, stable intermediates for intermediates I1 and I4 were found. In our lowest energy profiles for NS1 and NS2 intermediates I1 and I4 were not only stable but had lower associated energies than their preceding RS and I3 geometries respectively. This supports the idea that the driving force for the initial nucleophilic attack of Ser131 and for the following nucleophilic attack of a solvent water molecule is provided by greater stabilisation of the systems negative charge in the oxyanion hole.

Across the five QM/MM studies performed on the catalytic mechanism of *Is*-PETase, none of them are in total agreement as to the rate-determining step. Our results have led to the conclusion that the reaction steps Nuc2 and Lgr2 are dual rate-determining steps, with activation energies of 34.1 kJ mol^{-1} and 35.6 kJ mol^{-1} respectively. Our results do not agree wholly with any of the previous studies, although Fernandes *et al.* concluded that LgR2 was rate-determining (with an activation energy of 53.2 kJ mol^{-1}) and Feng *et al.* concluded that Nuc2 was one of two rate-determining steps (with an activation energy of 86.6 kJ mol^{-1}).

Of all of these QM/MM studies, our calculations have yielded the lowest energy barriers for the degradation of a PET substrate mimic via *Is*-PETase. As will be discussed in section 3.3.5, we found that energy profiles with lower energy values for tetrahedral intermediates I1 and I4 resulted in energy profiles with transition states with lower energy barriers. It is possible that our methods are capable of identifying lower energy tetrahedral intermediates, which has resulted in our lower energy barriers.

Table 3: Comparison of methods and key results of QM/MM studies of the mechanism of Is-PEtase, including our own study.

Main Author	Geometry Optimisation Method	Single-Point Method	Conformational Sampling	Transition state search	Rate-Determining Step	Activation Energy kJ mol^{-1}
Boneta	AM1 /M06-2X /6-31+G(d,p) /AMBER ff03	None	Umbrella Sampling	Umbrella Sampling	**LgR1**	98.7
Feng	M06-2X /6-31G(d) /CHARMM27	M06-2X /6-311G(d,p) /CHARMM27	20 energy profiles per reaction step	Microiterative Optimisation	**Nuc1 and LgR2**	86.6 and 87.0
Jerves	PBE /DZVP-GTH /AMBER ff14SB	None	Umbrella Sampling	Umbrella Sampling	**Acylation**	83.7
Aboelnga	B3LYP-D3(BJ) /6-31-G(d,p)	B3LYP-D3(BJ) /6-311-G(3df,3pd)	None	Frequency Calculations	**LgR2**	75.3
Fernandes	B3LYP /6-31(d,p) /AMBER18	Extrapolation to DLPNO-CCSD(T) /cc-pVTZ	None	Internal Reaction Coordinate Scan	**Nuc2**	52.3
Our Work	revPBE-D3(BJ) /SVP /CHARMM27	DSD-PBEP86-D3(BJ) /TZVP /CHARMM27	6 or 7 energy profiles per reaction step	Nudged-Elastic Band Calculations	**Nuc2 and LgR2**	34.1 and 35.6

Our methods most resemble those of Feng *et al.*, although we have used the nudged-elastic band technique to generate geometries for our transition states instead of microiterative optimisation methods (see Table 3). To sample the conformational space of reactions NS1 and NS2 we have generated multiple reaction profiles, each starting from a different timestep from a MD simulation. Boneta *et al.* and Jerves *et al.* used umbrella sampling techniques to sample the reactions' conformational space. This method should, in theory, provide more complete sampling but relies on appropriate selection of reaction coordinates and on the step-size used in these calculations. The studies by Aboelnga *et al.* and Fernandes *et al.* produced only one reaction profile for each reaction NS1 and NS2, attempting no sampling of the reactions conformational space. The energy profiles generated in our work and in the work by Feng *et al.* had large ranges in their activation energies and their corresponding rate-determining steps. By selecting rate-determining steps using the lowest energy profiles or through the use of Boltzmann-weighted averaging more appropriate and lower lying activation energies have been selected. This shows the great importance of conformational sampling in the application of QM/MM methods to the question of enzyme mechanism.

There is variation in the density functionals used in each study of the mechanism of *Is*-PETase (see Table 3). Using the simple 'Jacobs Ladder' heuristic, we have used the lowest level of theory for our geometry optimisations as revPBE is only a GGA functional. However, our functional has been shown to outperform the hybrid functional B3LYP [73], used by Aboelnga *et al.* and Fernandes *et al.,* against the GMTKN55 database. The functional we have used to calculate our single-point energies (DSD-PBEP86-D3(BJ)) represents the functional highest up on 'Jacobs Ladder' applied to *Is*-PETase so far as

it is the only double-hybrid functional that has been applied to this enzyme. Additionally, DSD-PBEP86-D3(BJ) performed very well against the GMTKN55 database[73]. This lends us great confidence that our energy profiles are accurate.

Another key, but easily overlooked, difference in methodology across these studies is the size and content of the QM region used. Boneta *et al.* used smallest QM region, with only the catalytic triad and the ligand included. This resulted in the exclusion of the oxyanion hole from the QM region, meaning that the crucial stabilisation of negative charge on OC_{LIG} in intermediates I1 and I4 was treated at the QM/MM boundary. To test the effect of this, we recalculated our single-point energies using this smaller QM region. Using both our single-point functional (DSD-PBEP86-D3(BJ)/ma-def2-TZVP) and the functional used by Boneta *et al.* (M06-2X), we found that the use of this smaller QM region resulted in higher energies for intermediates I1 and I4 as well as higher overall energy barriers (see Figure 35). It is of course worth noting that we did not perform any optimisations with this reduced QM region, as this would have been too computationally costly. Neglecting the oxyanion hole in the QM region during optimisation steps is likely to have had a significant effect on the resulting geometries.

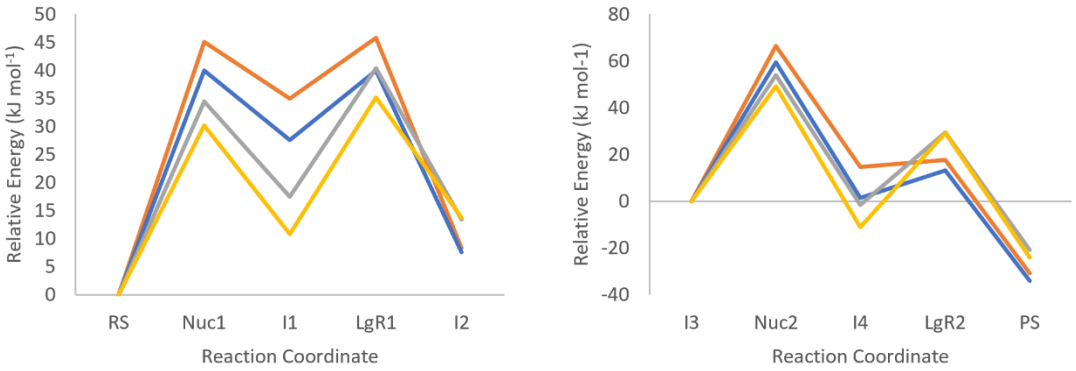


Figure 35: Average single-point energy values calculated using our functional and our QM region (grey), our functional and the QM region from Boneta et al. (orange), the functional used by Boneta et al. and our QM region (yellow) and the functional and QM region used by Boneta et al. (blue).

Other studies included the side chains of aromatic residues Tyr58[31,32] and Trp156[31–33] in their QM regions. These residues are located in the active site and have been shown to interact with the aromatic rings of PET, making them crucial for substrate binding. Whilst we have not tested the effect of inclusion of these aromatic residues into the QM region, it is our opinion that that their inclusion would have a negligible effect, as they do not interact directly with any atoms involved in the enzyme's mechanism. Additionally, their interaction with the aromatic rings of the substrate is likely to remain constant throughout the degradation reaction, therefore contributing little to changes in relative energies. Other residues such as Ser178[30,33], Ile179[30,33] and Ala209[33] as well as a

water molecule that forms a hydrogen bond with the catalytic Asp208 residue[33] have also been included in QM regions. It is thought that these residues have some stabilising effect upon the catalytic residue Asp208. It is unclear what effect the inclusion of these atoms in the QM region had upon the energy profiles produced.

3.3.5 Analysis of Geometries throughout the Is-PETase reaction profile

3.3.5.1.1 Geometry Analysis Nuc1 Step

The first step in the NS1 reaction is the Nuc1 step. In this step $O\gamma_{S131}$ is deprotonated by $N\varepsilon_{H20}$, $H\varepsilon_{H208}$ is deprotonated by $O\delta_{D177}$ and $O\gamma_{S131}$ nucleophilically attacks CO_{LIG} according to the canonical mechanism of a catalytic triad. This results in a transfer of a formal negative charge from $O\delta_{D177}$ to OC_{LIG} which is stabilised through hydrogen bonding interactions with N_{Y58} and N_{M132} which arguably provides the driving force for the reaction step.

Since Nuc1 is a nucleophilic attack step upon a carbonyl, the Bürgi-Dunitz angle of attack[94] (consisting of the angle between CO_{LIG}, CC_{LIG} and $O\gamma_{S131}$) can be analysed (see Figure 36). The ideal angle of attack is 107°, as this allows the lone pair of $O\gamma_{S131}$ to interact with the π* orbital of the carbonyl bond located on CO_{LIG} with maximum overlap. Analysis of the Bürgi-Dunitz angle of attack in each of our optimised RS geometries revealed an average angle of 83°, which is significantly more acute than the ideal geometry. A correlation ($R^2 = 0.63$) was found between the Bürgi-Dunitz angle of attack in each RS geometry and its associated Nuc1 energy barrier. Unexpectedly, the RS geometries with the more acute angles (and therefore further away from the ideal angle of 107°)

resulted in lower Nuc1 energy barriers (see Figure 36).

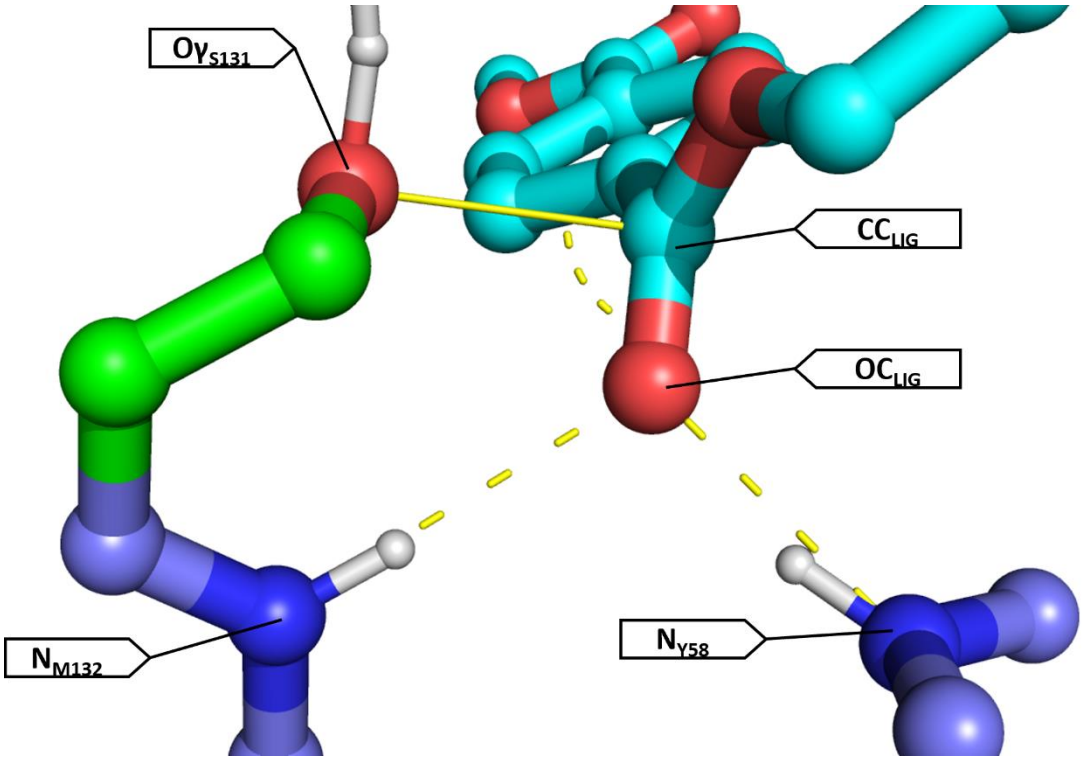


Figure 36: Oxyanion hole distances and Bürgi-Dunitz angle of attack annotated onto an example RS geometry.

Another key factor in Nuc1 is the stabilisation of the developing negative charge on OC_{LIG} by the oxyanion hole (OAH) (consisting of N_{Y58} and N_{M132}, see Figure 36). An ideal geometry would have strong hydrogen bonding interactions between hydrogen bond acceptor OC_{LIG} and the hydrogen bond donors N_{Y58} and N_{M132} (these will be referred to as OAH_{Y58} and OAH_{M132} respectively). The ideal length of a hydrogen bond is between 2.7 Å and 3.3 Å. In the RS geometries the OAH_{Y58} distances are consistently within this range (actual range is between 2.74 Å and 2.85 Å). This shows that N_{Y58} forms a strong hydrogen bonding interaction with OC_{LIG}. By contrast, the OAH_{M132} distance is consistently longer than OAH_{Y58} with only one RS geometry within the ideal range for a hydrogen bond. A plot of OAH_{Y58} and OAH_{M132} distances against their associated Nuc1 energy barrier revealed a weak positive correlation for OAH_{Y58} ($R^2 = 0.16$) and a strong correlation for OAH_{M132} ($R^2 = 0.81$). This can be rationalised as the OAH_{Y58} distances are all suitably short to allow efficient hydrogen bonding so reducing the distance only slightly increases the stabilisation effect. As the majority of OAH_{M132} distances in RS geometries are too long for efficient hydrogen bonding, a significant amount of stabilisation of the negative charge on OC_{LIG} is gained when this distance is reduced by introducing a stronger hydrogen bond. Another way to look at the effect of the oxyanion hole is to measure the CO_{LIG}-OC_{LIG} distance in the RS geometries. A plot of this bond length against OAH_{Y58} and OAH_{M132} revealed a strong negative correlation for OAH_{M132} ($R^2 = 0.83$) showing that as the hydrogen bonding

interactions in the oxyanion hole get stronger (i.e., a shorter OAH_{M132} distance) the CO_{LIG}-OC_{LIG} bond is destabilised more. A plot of the carbonyl bond length against the associated Nuc1 energy barrier revealed a negative correlation (R^2 = 0.68), showing that the destabilisation of the carbonyl makes it less energetically costly to break in the Nuc1 step, lowering the activation barrier for this step.

The energy of I1 geometries (relative to their preceding RS geometries) varies greatly with a range between -2.7 kJ mol^{-1} and +38.5 kJ mol^{-1}. In theory, the I1 geometry should have a lower energy value than the RS geometry as the negative charge is better stabilised when located on OC_{LIG} via hydrogen bonding with the oxyanion hole than it is in the apo-enzyme (where the formal negative charge is located on $O\delta_{D177}$ and only stabilised by one hydrogen bond to $N\varepsilon_{H208}$). Through visual inspection of the energy profiles of each snapshot it is clear that the snapshots with lower I3 energies also have lower activation energies for their Nuc1 steps. A plot of I3 energies against their equivalent Nuc1 energy barriers revealed a strong positive correlation (R^2 = 0.91). Visual inspection of the I1 geometries revealed a rotation in the side chain dihedral of His208 (consisting of C_{H208}, $C\alpha_{H208}$, $C\beta_{H208}$ and $C\gamma_{H208}$). This dihedral however has only a weak correlation with the I1 energy (R^2 = 0.33). A plot of the carbonyl bond length (made up of CO_{LIG} and OC_{LIG}) in each I1 geometry against the geometry's energy revealed a very strong negative correlation (R^2 = 0.95). This means that I1 energy depends strongly on the stabilisation of the negative charge located upon OC_{LIG}, allowing the bond between CO_{LIG} and OC_{LIG} to have a more single-bond-like length and reducing the overall energy of the system.

3.3.5.1.2 Geometry Analysis of LgR1 Step

The second step in NS1 is the LgR1 step; the side chain of His208 rotates such that it can transfer a proton from $N\varepsilon_{H20}$ to OE_{LIG}; at the same time the carbonyl bond between CO_{LIG} and OC_{LIG} reforms. This causes the bond between CO_{LIG} and OE_{LIG} to break, creating an ethylene glycol molecule as a leaving group.

The lowest energy difference between I1 and LgR1 geometries was calculated as 3.1 kJ mol^{-1}. Upon further inspection it was found that the transition state geometry has a far longer CO_{LIG}-OC_{LIG} bond length than the LgR1 transition state geometries from the other snapshots (1.30 Å compared to values below 1.26 Å). This long CO_{LIG}-OC_{LIG} bond length is typical of a single-bonded C-O bond, suggesting that the transition state occurs before the reformation of the ligand's carbonyl has begun. In this anomalous LgR1 geometry $N\varepsilon_{H20}$ is practically equidistant to $O\gamma_{S131}$ and OE_{LIG} (2.79 Å and 2.74 Å respectively); this combined with the long CO_{LIG}-OC_{LIG} bond suggests that in this case the dihedral rotation of His208 is responsible for the step's energy barrier.

For the remainder of the reaction pathways, the energy barrier for LgR1 was calculated to be far higher (between 23.8 kJ mol^{-1} and 40.1 kJ mol^{-1}). In these cases, the CO_{LIG}-OC_{LIG} bond length decreased throughout the LgR1 step with carbonyl bond lengths between 1.30 Å and 1.29 Å for I1 geometries decreasing to between 1.26 Å and 1.25 Å in the transition state LgR1 and further decreasing to bond lengths clustered around 1.24 Å. This observation is consistent with the reformation of the carbonyl throughout the reaction step.

A strong positive correlation was found between the relative energy of intermediate I2 and the activation energy of the preceding LgR1 transition state (R^2 = 0.85). This suggests that stabilisation of the I2 intermediate is important for the reduction energy barrier associated with the LgR1 step. It is unclear however what factors contribute to the variance in energies for our I2 geometries.

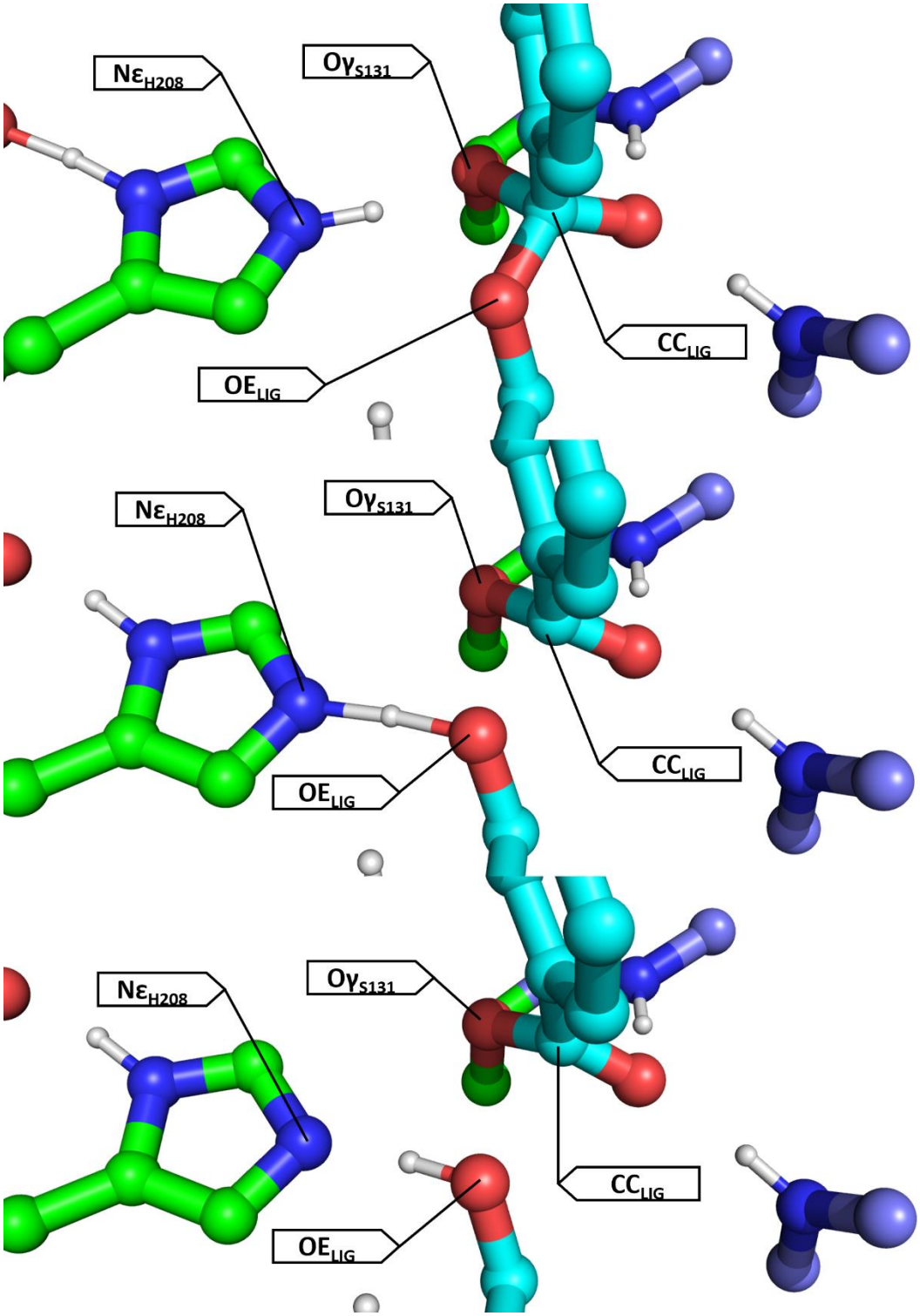


Figure 37: Sample geometries for intermediate I1 (top), transition state LgR1 (middle) and intermediate I2 (bottom), with key atoms annotated.

3.3.5.1.3 Geometry Analysis of Nuc2 Step

In the reaction pathway with the lowest overall energy barrier (calculated as two separate reaction pathways) the Nuc2 step represents has the highest activation energy. This makes Nuc2 the rate-determining-step when considering chemical steps only. In this step, a water molecule is

deprotonated by $N\varepsilon_{H20}$ and undergoes a nucleophilic attack at CO_{LIG}. Concurrently, a rearrangement in the double-bonds in the aromatic ring of His208 occurs accompanied by a transfer of a proton from $N\varepsilon_{H208}$ to $O\delta_{D177}$.

In the Nuc2 reaction step the transfer of a formal negative charge from $O\delta_{D177}$ to OC_{LIG} occurs. This negative charge is stabilised through hydrogen bonding interactions with the nitrogen atoms that make up the oxyanion hole (N_{Y58} and N_{M132}). This charge stabilisation is greater in intermediate I4 than the preceding intermediate I3 and thus provides the driving force for this reaction step. Four out of six I4 geometries were calculated to have lower energies than their associated I3 geometry, representing a greater degree of stabilisation of the negative charge now located upon OC_{LIG}. The importance of the oxyanion hole stabilisation is revealed by plotting the following distances against their respective Nuc2 energy barriers: OC_{LIG}-N_{Y58} and OC_{LIG}-N_{M132}. An interesting observation was made in the plot of these distances for the Nuc2 geometries; a strong positive correlation was found between the OC_{LIG}-N_{M132} distance and the Nuc2 energy barrier ($R^2 = 0.83$) while essentially no correlation was found between the OC_{LIG}-N_{Y58} distance and the Nuc2 energy barrier ($R^2 = 0.04$). In the same manner as discussed previously for the Nuc1 step, the hydrogen bonding distances for OC_{LIG}-N_{Y58} are consistently shorter (with a range between 2.61 Å and 2.68 Å) than the OC_{LIG}-N_{M132} distances (with a range between 3.07 Å and 3.57 Å). The oxyanion hole stabilisation can be observed through the CO_{LIG}-OC_{LIG} bond length, the longer the bond the stronger the stabilisation. A plot of the CO_{LIG}-OC_{LIG} bond length in the Nuc2 geometries was made against the hydrogen bonding distance OC_{LIG}-N_{M132}. It can be observed that for geometries with OC_{LIG}-N_{M132} between 3.23 Å and 3.57 Å the CO_{LIG}-OC_{LIG} bond remains short (all values are 1.25 Å to 2 decimal places). This suggests that in the majority of the reaction pathways calculated, the energy barrier is not associated with the breaking of the CO_{LIG}-OC_{LIG} double bond, but some other factor. One Nuc2 transition state geometry had an OC_{LIG}-N_{M132} distance of 3.07 Å and a CO_{LIG}-OC_{LIG} bond length of 1.31 Å. This transition state corresponds with the lowest energy barrier calculated for the Nuc2 step. In this case, the other factor mentioned previously does not seem to determine the activation energy of this step, rather the breaking of the CO_{LIG}-OC_{LIG} double bond as expected.

It was expected that His208 would maintain a hydrogen-bonding interaction with O_{SOLV} in intermediate I4 and would later rotate to form a new hydrogen bond with $O\gamma_{S131}$ in the proceeding LgR2 step. It was however found that all I4 geometries have shorter distances between $N\varepsilon_{H20}$ and $O\gamma_{S131}$ than between $H\varepsilon_{H20}$ and O_{SOLV}. This shows that the $H\varepsilon_{H20}$-O_{SOLV} hydrogen-bond is broken during the Nuc2 step and may contribute to the activation energy of this step. During each Nuc2 step, the side chain of Ser131 rotates significantly with dihedral angles (consisting of $C\alpha_{S131}$, $C\beta_{S131}$, $O\gamma_{S131}$ and CO_{LIG}) between 118.3° and 136.5° in the set of I3 geometries increasing to between 138.5° and

159.2° in the set of Nuc2 transition states and becoming dramatically more acute in the set of I4 geometries with angles between 97.5° and 135.4°. This rotation most likely serves to orient OC_{LIG} in a better position that facilitates stronger hydrogen-bonding interactions with the oxyanion hole.

The conclusion can therefore be made that the activation energy for the Nuc2 step (and therefore the energy of the rate-determining step for the entire hydrolysis reaction) is largely dependent on OC_{LIG} forming strong hydrogen bonding interactions with the oxyanion hole during the Nuc2 step.

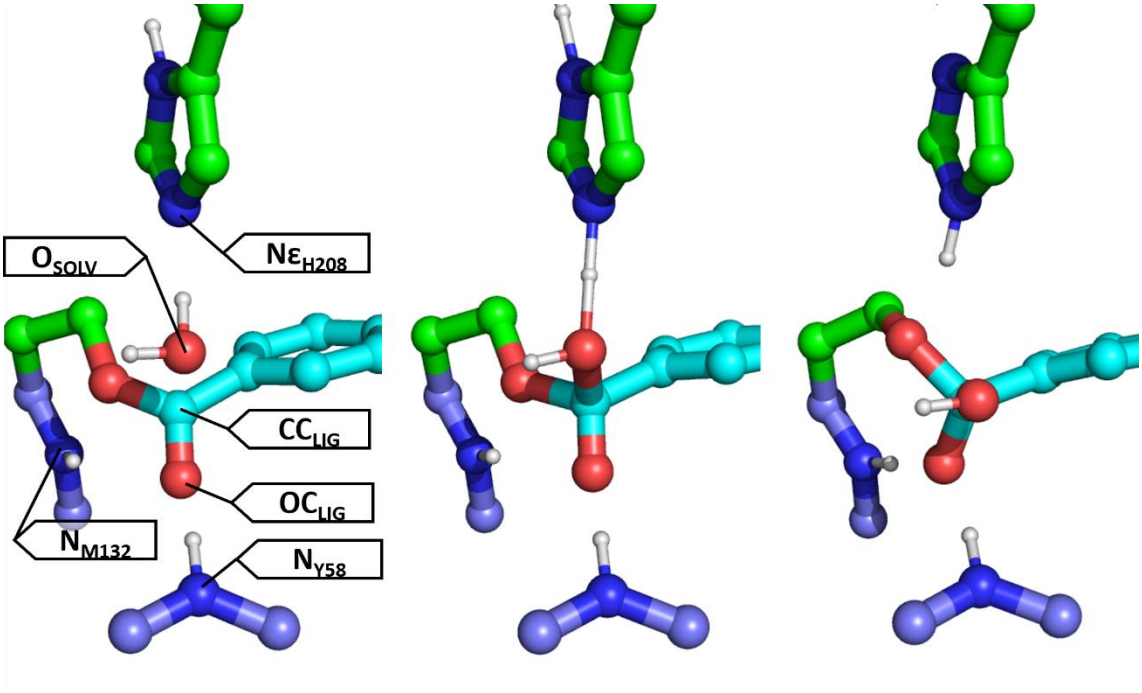


Figure 38: Example geometries for intermediate I3 (left), transition state Nuc2 (middle) and I4 (right), with key atoms annotated.

3.3.5.1.4 Geometry Analysis of LgR2 Step

The last step in the NS2 reaction pathway is the LgR2 step: A proton is transferred from $N\varepsilon_{H20}$ to $O\gamma_{S131}$ facilitating the breaking of the $O\gamma_{S131}$-CO_{LIG} bond and the reformation of the CO_{LIG}-OC_{LIG} bond. The result of the LgR2 step is the regenerated apo-enzyme and the release of the TA portion of the product (in our case this is monomethyl-terephthalic acid).

The LgR2 step starts with intermediate I4 in which His208 forms a hydrogen bond with the newly formed hydroxyl group of this ligand through a $H\varepsilon_{H20}$-O_{SOLV} interaction. For the LgR2 step to proceed this hydrogen bond must be broken, and a new $H\varepsilon_{H20}$-$O\gamma_{S131}$ hydrogen-bond interaction formed. This also requires a rotation in the side-chain of His208 such that $N\varepsilon_{H20}$ is oriented towards $O\gamma_{S131}$ rather than O_{SOLV}.

This reaction step can be seen as analogue of the LgR1 step: His208 starts with a hydrogen-bond to O_{SOLV} and ends up with a hydrogen bond to $O\gamma_{S131}$, resulting in cleavage of the CO_{LIG}-$O\gamma_{S131}$ bond. It was however observed that for all I4 geometries $N\varepsilon_{H208}$ is significantly closer to $O\gamma_{S131}$ than O_{SOLV}. This shows that this rotation of the side-chain of His208 occurs in the previous Nuc2 Step. The major energy component of LgR2 can therefore be concluded to be assocciated with the breaking of the $O\gamma_{S131}$-CO_{LIG} bond, rather than with rotations of the His208 sidechain.

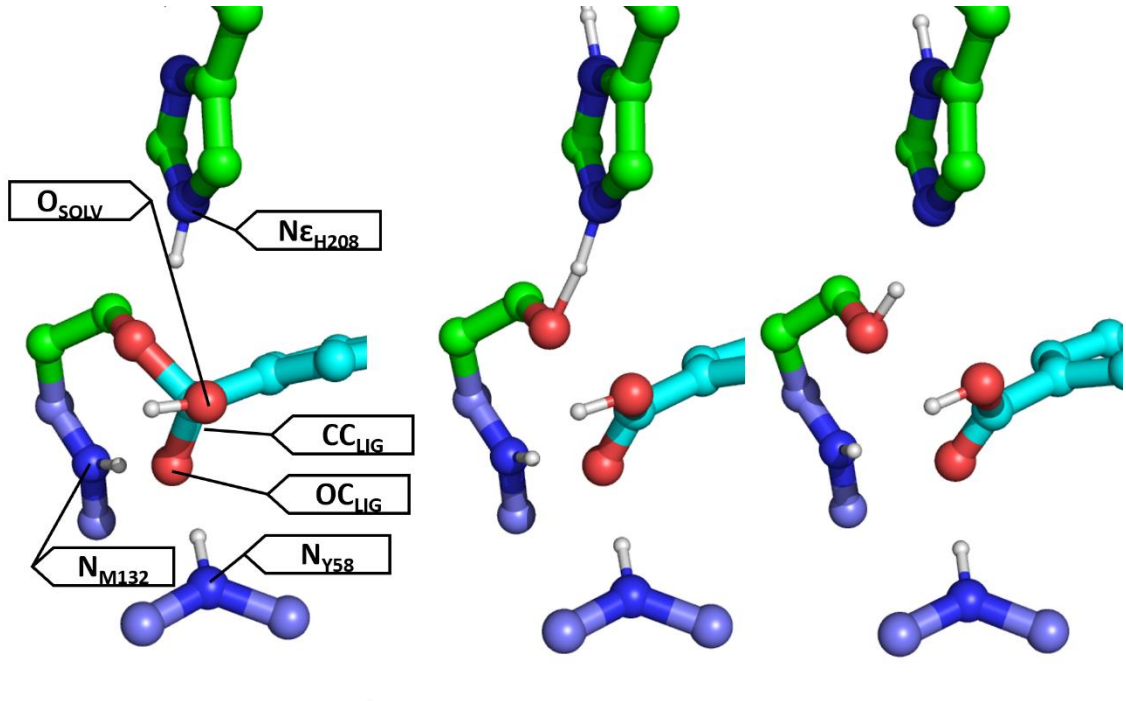


Figure 39: Example geometries for intermediate I4 (left), transition state LgG2 (middle) and intermediate PS (right), with key atoms labelled.

A negative correlation (R^2 = 0.71) was found between the relative energies of I4 geometries and the activation energies of their proceeding LgR2 step (i.e., the higher the relative energy of an I4

geometry, the lower activation energy for the following LgR2 step). It is possible that in high energy geometries of I4, an energy cost has already been paid that does not need to be in the proceeding LgR2 step. Unfortunately, comparisons of the geometries of our I4 intermediates do not reveal any clear factor that could be causing this effect.

3.3.6 Deprotonation of Product as Possible Route to Product Inhibition

In several of the reaction profiles, no stable geometry for the final state PS was found. Instead, QM/MM geometry optimisations resulted in the transfer of a proton from CE_{LIG} to $O\gamma_{S131}$. This resulted in a rearrangement of the aromatic bonds in His208 and proton transfers between $O\gamma_{S131}$ and $N\varepsilon_{H20}$ and between $N\varepsilon_{H208}$ and $O\delta_{D177}$. The negative charge located on the sidechain of D177 in PS had been transferred to the new carboxylic acid group on the ligand (which at this point is the product of the reaction mono methyl terephthalic acid). This new geometry was given the name aPS (After Product State).

For reaction profiles with stable PS intermediates, the aPS geometry was generated starting from the PS intermediate and applying the following optimisations: A QM/MM geometry optimisation with a restraint of 1 Å between the proton bound to OE_{LIG} and $O\gamma_{S131}$, followed by an unrestrained QM/MM geometry optimisation.

For reaction profiles with no stable PS intermediate, the transition state for its LgR2 step was generated using a NEB calculation with I4 and aPS as input geometries. It was found that the transition states generated with this method had similar geometries and energies (relative to their corresponding I3 geometries) to that of LgR2 transition states that had been generated using I4 and PS geometries as inputs for their NEB calculations.

3.3.7 Protonation Transfer Between His208 and Asp177

It was noted in the work by Fernandes et al.[33] (see section 3.1.1.6 for a details of this study) that their QM/MM calculations showed no proton transfer between His208 and Asp177 throughout the reaction mechanism of Is-PETase. It was however reported by Feng et al.[31] (discussed in section 3.1.1.3) that this proton transfer had a small statistical likelihood of occurring, based upon results of

multiple energy profiles generated for each reaction step.

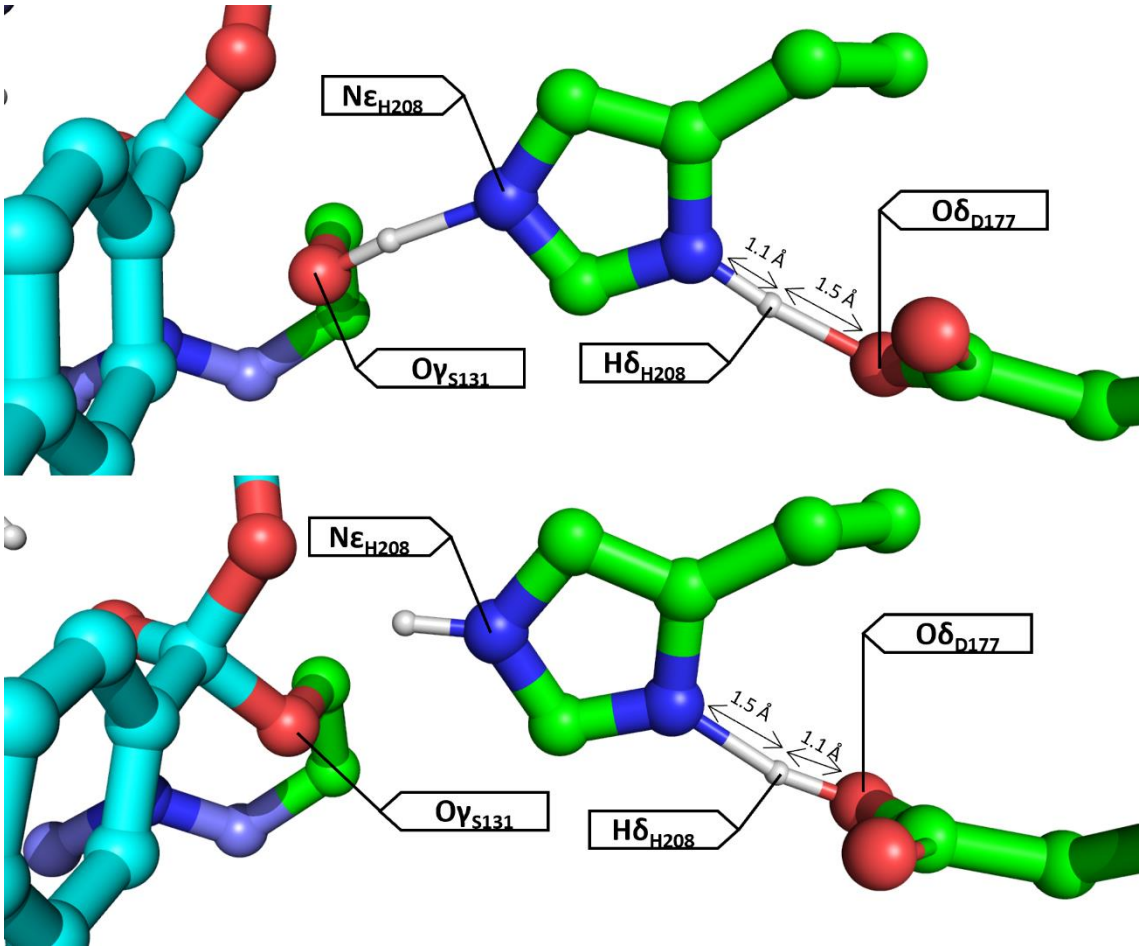


Figure 40: Example geometries where $H\delta_{H208}$ is bound to $N\delta_{H208}$ (top) and where $H\delta_{H208}$ is bound to $O\gamma_{S131}$.

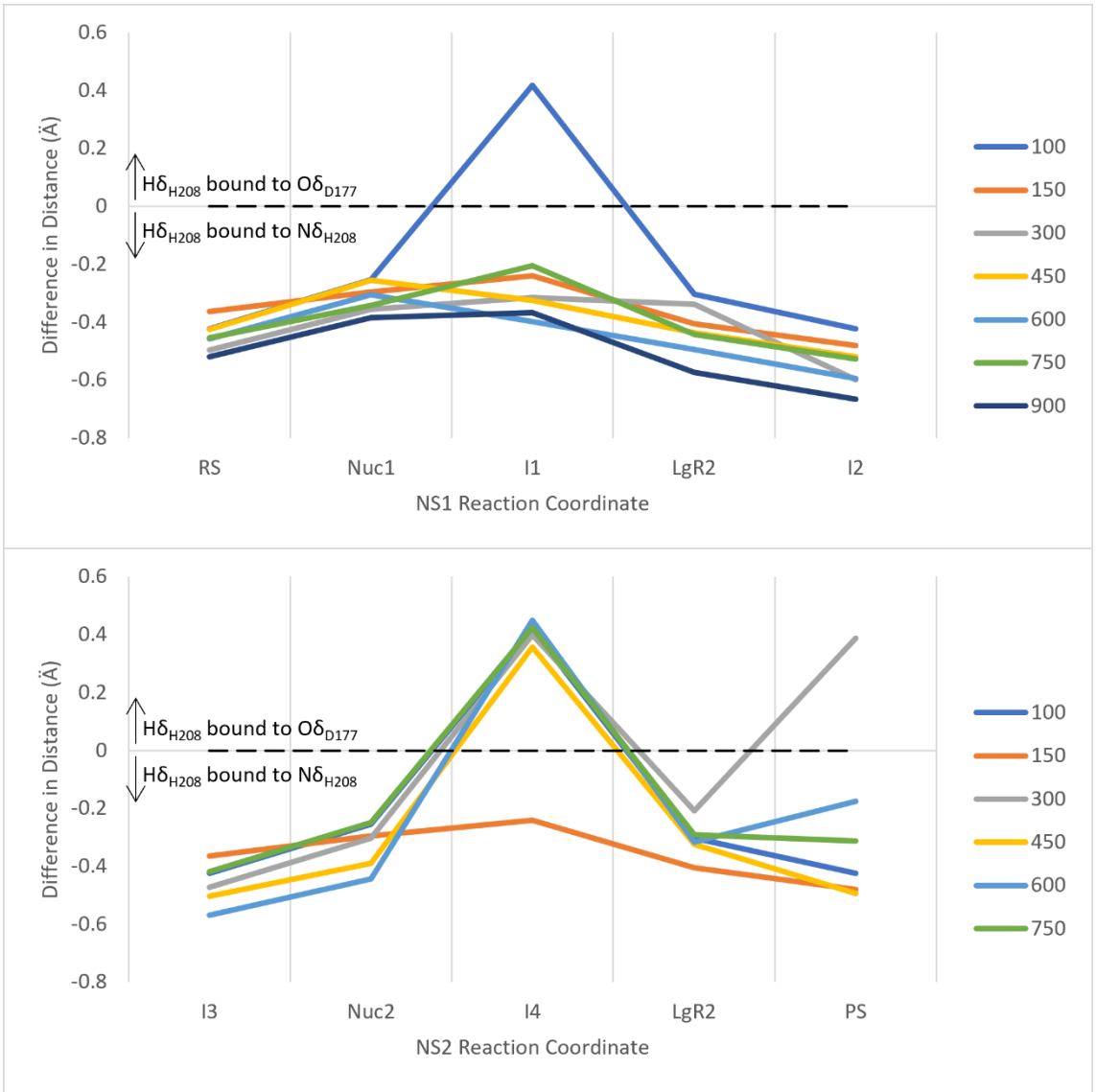


Figure 41: Plot showing the location of Hδ$_{H208}$ throughout reactions NS1 (top) and NS2 (bottom). The y-axis has been calculated as the difference between the distances Hδ$_{H208}$ - Nδ$_{H208}$ and Hδ$_{H208}$-Oγ$_{S131}$.

Analysis of the geometries of intermediates and transition states from our work revealed that for the majority of our reaction profiles for the reaction NS1, no proton transfer was observed (i.e., proton Hδ1$_{H208}$ remained closer to Nδ1$_{H208}$ than to Oδ$_{D177}$). For one of our reaction profiles for NS1, a proton transfer from Nδ1$_{H208}$ to Oδ$_{D177}$ was observed in the I1 geometry only (see Figure 41).

For the reaction NS2, the opposite was found. In five out of six reaction profiles generated, a proton transfer from Nδ1$_{H208}$ to Oδ$_{D177}$ was observed. In four out of the five reaction profiles where this proton transfer was observed, the proton transfer was observed for in intermediate I4 only. In the remaining reaction profile, the proton transfer occurred in the transition state Nuc2. Hδ1$_{H208}$ remained bound to Oδ$_{D177}$ in the proceeding intermediate I4 (see Figure 41).

Our work supports the observation made by Feng et al.[31] We posit that if Fernandes et al.[33] had generated several reaction profiles this behaviour is likely to have been observed in their own results as well.

3.3.8 Charge Analysis

Each of our QM/MM calculations provides the Mulliken gross atomic charge (q_A) for each atom in the QM region. For geometry optimisations, q_A values are provided for QM atoms in the final optimised geometry. As the geometry does not change in our single-point calculations, the q_A value provided corresponds to the QM atoms in the calculation's input geometry.

Mulliken charges are calculated using a Mulliken population Analysis technique. This starts with a set of normalised atomic orbitals (which can be obtained from a DFT calculation), the square of which results in the charge distribution as a probability density. Integration of the charge distribution over all electronic coordinates is performed. This is followed by splitting the components of the integral into a Mulliken population matrix. Each element of this matrix describes the charge distribution in a single molecular orbital. The net population matrix is then formed as a sum of all Mulliken population matrices for occupied molecular orbitals. From the net population matrix, atomic charges for each atom in the system can be extracted[95].

It has been noted that Mulliken charge analysis is highly sensitive to basis set and that calculated Mulliken charge values do not converge with respect to basis set, unlike energy calculations. This makes comparisons between systems calculated using different basis sets meaningless[96]. Whilst the Mulliken charges may not have any direct relation to observable properties in our system, their analysis may reveal addition information on key aspects of our mechanism (i.e., their differences relative to each other may be instructive). It was decided to analyse the q_A values from our set of single-point calculations, as these have all been calculated using the TZVP basis set. The q_A values of key heteroatoms in the QM region were extracted from the output files for each single-point calculation in each reaction profile.

3.3.8.1 Analysis of the Relation Between Mulliken Charge and Activation Energy

From our previous analysis of bond lengths, angles, and dihedrals we concluded that the most important determinant for the activation barrier for reaction steps Nuc1 and Nuc2 was the degree to which the negative charge localised upon OC_{LIG} was stabilised by hydrogen bonding with N_{Y58} and N_{M132} atoms in the oxyanion hole. As such, the q_A values of OC_{LIG}, N_{Y28} and N_{M132} were compared across each reaction profile.

It was found that the q_A values for N_{Y58} and N_{M132} do not vary significantly across any of our NS1 reaction profiles. Similarly for five out of seven of our calculated reaction profiles, the q_A values for OC_{LIG} do not vary significantly across the reaction profile. The remaining two reaction profiles have significant reductions in q_A values from RS geometries to I1 geometries (with q_A values for Nuc1 geometries being in between), followed by an increase in q_A values from I1 geometries to I2 geometries (with q_A values for LgR1 geometries being in between). These two energy profiles with significant changes in q_A values represent the two lowest energy barriers calculated for the Nuc1 step. A plot of the difference between the q_A values centred on OC_{LIG} for geometries of I1 and RS and the corresponding value for the Nuc1 activation energy reveals a strong correlation ($R^2 = 0.79$). If we are to assume that Mulliken charges and activation energies are causally linked, we could conclude that the reduction in q_A value for atom OC_{LIG} during the Nuc1 step (in chemical terms this can be interpreted as an increase in partial charge localised on OC_{LIG}) is important for reduction of the Nuc1 activation energy.

No correlation was found between the difference in Mulliken Charge on OC_{LIG} between our I1 and I2 geometries for the LgR1 reaction step ($R^2 = 0.07$). As the q_A value located on OC_{LIG} did not significantly change for five out of seven reaction profiles, this analysis does not provide conclusions that are particularly instructive.

A similar analysis was performed looking at the difference between q_A values centred upon N_{M132} for Nuc1 (using q_A values from I1 and RS geometries) and LgR1 (using q_A values from I2 and I1 geometries) steps. Plots of these values against their corresponding reaction step's activation energy resulted in a weak correlation in Nuc1 ($R^2 = 0.34$) and no correlation for LgR1 ($R^2 = 0.07$).

In contrast to NS1, the q_A values for OC_{LIG} follow a consistent pattern across the reaction coordinate of NS2: During the reaction step Nuc1, the q_A value on OC_{LIG} drops from intermediate I3 to I4 via transition state Nuc2, then rises again from intermediates I4 to PS via transition state LgR2 (average q_A values of -0.55, -0.76, -0.87, -0.72, -0.59 for geometries I3, Nuc1, I4, Nuc2, and PS respectively). No clear correlation was found between the q_A values centred upon OC_{LIG} throughout the reaction pathways of NS2 and the activation energies of reaction steps Nuc2 and LgR2. As with NS1, the q_A values centred upon N_{Y58} and N_{M132} do not vary significantly throughout the NS2 reaction profile. Again, this analysis has proved to be rather uninstructive.

A strong positive correlation was found between the q_A value centred upon the electrophilic CO_{LIG} atom in RS geometries and the activation energy of the proceeding Nuc1 transition state ($R^2 = 0.84$). This could suggest that a reduction in electron density on CO_{LIG} prior to nucleophilic attack by $O\gamma_{S131}$ results in stabilisation of the proceeding transition state. Unlike NS1, no correlation ($R^2 = 0.06$) was

found between the q_A value centred upon CO_{LIG} in I3 geometries and the activation energy of the proceeding Nuc2 reaction step.

It is unsurprising that strong correlations with energy are not found for q_A values of OC_{LIG}, N_{Y58}, and N_{M132} for our NS2 geometries as they do not vary significantly across each reaction profile (i.e., all reaction profiles have similar q_A values at each intermediate and transition state). It is however surprising that no clear trend was found for the q_A values for CO_{LIG} in I3 geometries with the activation barrier for the proceeding Nuc2 steps, especially given the strong correlation found for the equivalent CO_{LIG} q_A values in RS geometries and Nuc1 activation barriers. The lack of any clear relation between the Mulliken charges on the key heteroatoms in our QM region and the activation energies calculated leads me to conclude that their use may be limited to a qualitative description of the movement of electrons throughout the reaction mechanism.

3.3.8.2 Mulliken Charges Provide a Qualitative Description of Charge Transfer

Analysis of the key Mulliken charges in our QM system across each reaction profile yielded little information as to what factors are key to determining the activation energies of each step in the mechanism of *Is*-PETase. We can, however, use these charges to qualitatively describe the movement of electrons across the QM region throughout the duration of the chemical mechanism.

Figure 42 shows the Mulliken charges for key atoms throughout the lowest energy reaction profile generated for the reaction NS1.

By taking the reaction profiles with the lowest energy barrier for NS1 and NS2, the Mulliken charges for key heteroatoms can be visualised (see Figure 42 and Figure 43). A comparison with the 'curly arrow' mechanism reveals that the redistribution of charge throughout the mechanism is far more complex than the movement of a single negative charge between $O\delta_{D177}$ to OC_{LIG}.

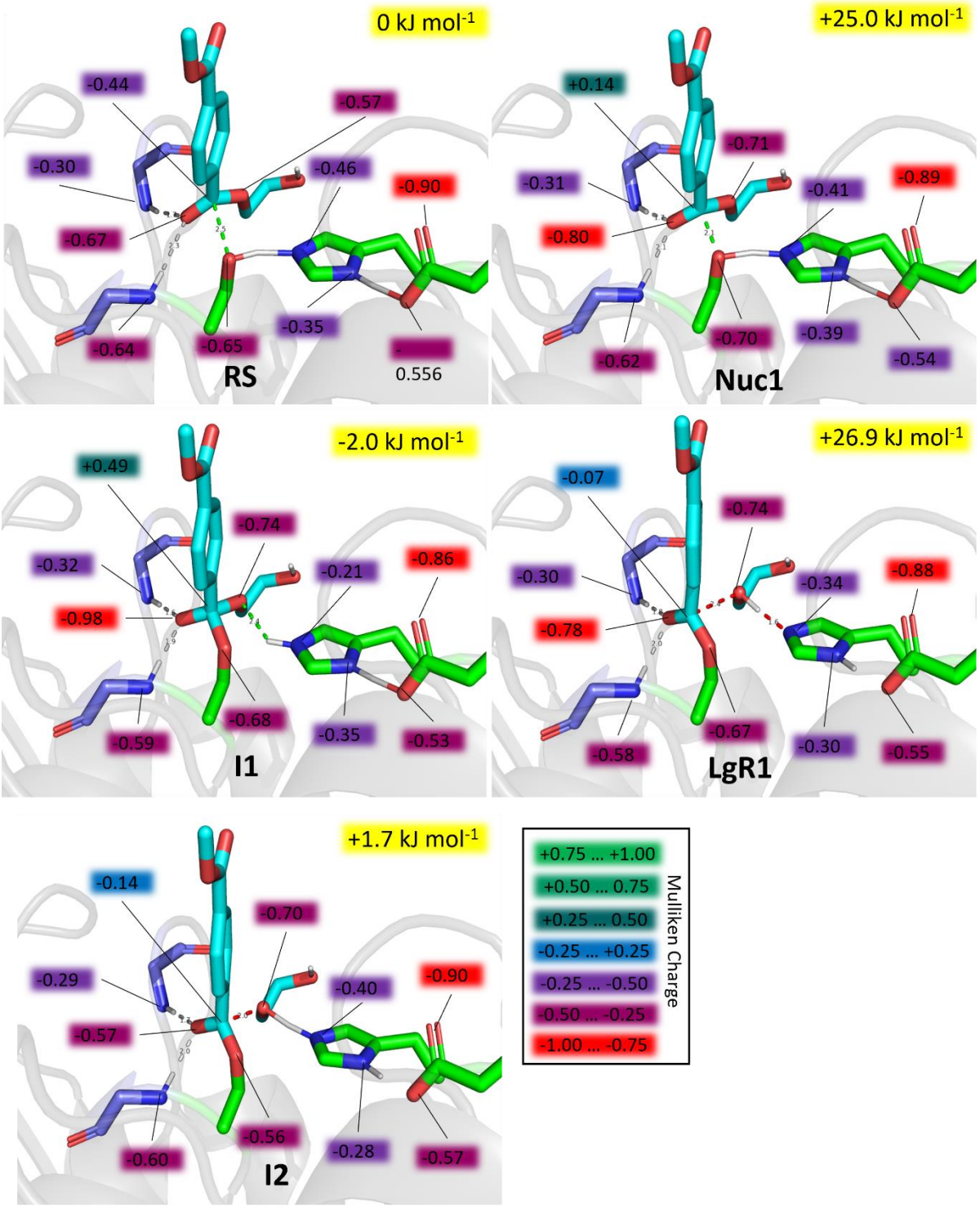


Figure 42: Mulliken charges (shown as labels) calculated for the lowest energy profile found for reaction NS1. For each geometry, the single-point energy relative to intermediate RS is shown in yellow.

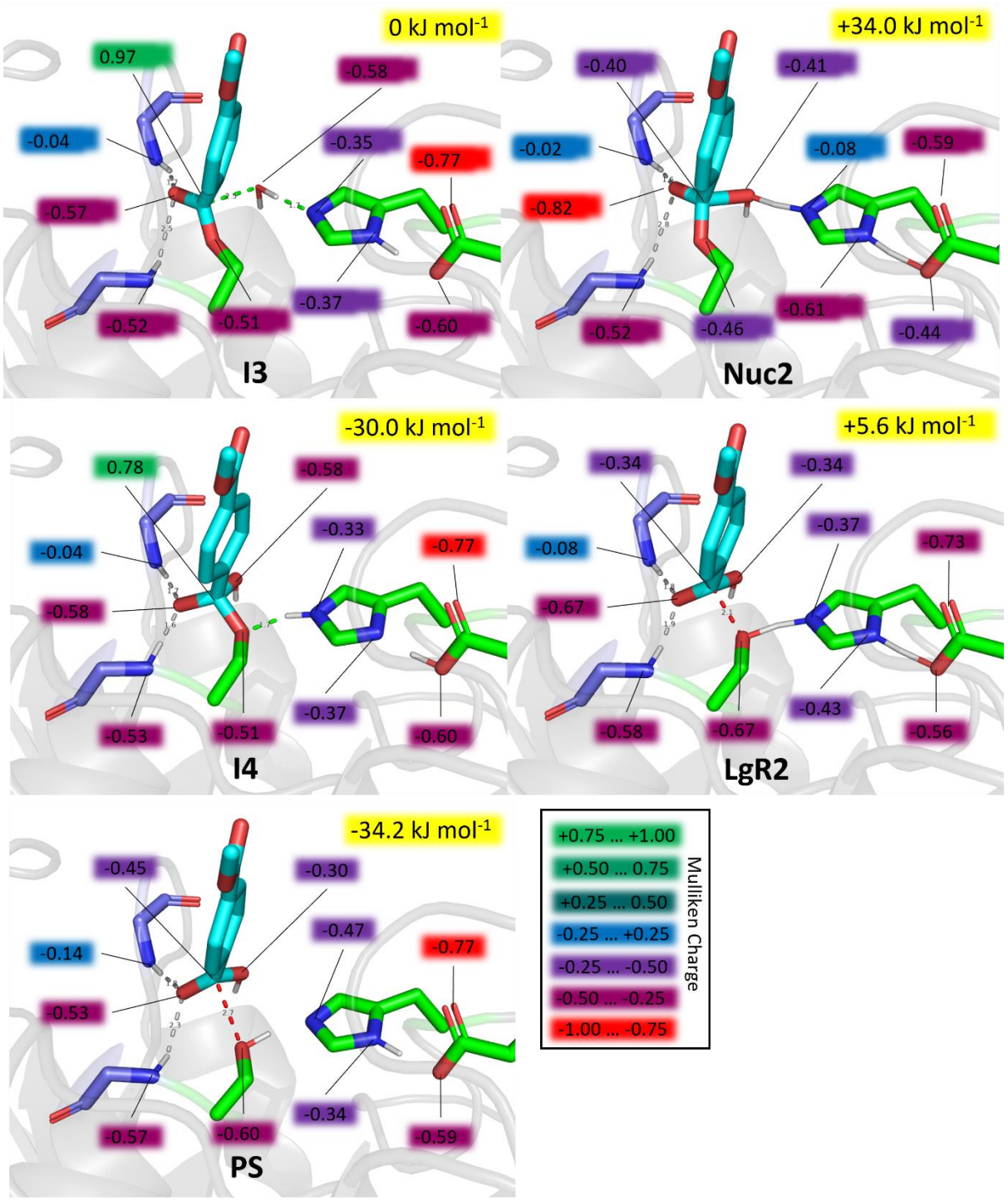


Figure 43 Mulliken charges (shown as labels) calculated for the lowest energy profile found for reaction NS2. For each geometry, the single-point energy relative to intermediate I3 is shown in yellow.

In the first intermediate RS the most negative q_A value in the QM region is located on $O\delta2_{D177}$, with a q_A value of -0.9. Atom $O\delta_{D177}$ has a less negative q_A value of -0.55, perhaps due to the hydrogen bonding interaction with $N\varepsilon_{H208}$. While these values do not agree with the 'curly arrow' mechanism (where $O\delta_{D177}$ has a formal negative charge and $O\delta2_{D177}$ is a neutral carbonyl), their q_A values can be

rationalised easily. As the reaction progresses however, the q_A values for $O\delta_{D177}$ and $O\delta2_{D177}$ do not significantly change. This result is at odds with the proposed mechanism as, according to the 'curly arrow' mechanism, the formal charge localised on $O\delta_{D177}$ should transfer to OC_{LIG} during the Nuc1 step and back to $O\delta_{D177}$ in the LgR1 step. It is likely that this is indeed occurring (although not as a simple formal negative charge) and the method for generating Mulliken charges is providing spurious results.

More physically reasonable q_A values can be observed for the charge centred on OC_{LIG}. The q_A value on OC_{LIG} reduces from -0.66 in RS to -0.98 in I1 during the Nuc1 step, then increases to -0.57 in the LgR1 step. This is commensurate with the proposed mechanism, as a formal negative charge is proposed to build up on OC_{LIG} in I1. In contrast, the q_A value on OC_{LIG} does not follow the proposed mechanism for NS2. In the Nuc2 step, the q_A value is -0.57 in I3 and -0.58 in I4 with the q_A value lowest at the transition state Nuc2 with a value of -0.82. If we are to take these values at face value, this suggests that there is no negative charge localised upon OC_{LIG} in I3. Based on the valency of CO_{LIG} (tetrahedral) this is very unlikely, further shedding doubt on the validity of these Mulliken charges.

In both NS1 and NS2, the q_A value located on N_{M132} is significantly more negative than the q_A value on N_{Y58}. As discussed in section 3.3.5, the hydrogen bonding distance with OC_{LIG} is consistently longer for N_{M132} than for N_{Y58}. A plot of these hydrogen bonding distances against their corresponding q_A values for each revealed no clear correlation between q_A values for N_{Y58} and N_{M132} and their corresponding hydrogen bonding distances with OC_{LIG}.

3.4 Conclusions

In this project, several reaction profiles for the degradation of a PET substrate mimic HEMT via the enzyme *Is*-PETase have been generated through a series of QM/MM methods. This reaction has been treated as two separate, sequential nucleophilic substitution reactions, NS1 and NS2. The conversion from the final intermediate of NS1 (intermediate I2) to the initial intermediate of NS2 (intermediate I3) has been assumed to have negligible energy barriers and thermodynamic driving forces. This assumption is supported by MD simulations of this process[30], which involves the dissociation of an EG molecule from the active site.

The rate-determining steps for reactions NS1 and NS2 were determined. This was achieved through selecting the reaction profile with the lowest overall energy barrier for each reaction. From these lowest energy reaction profiles, the reaction step with the highest energy barrier associated with it was determined to be the rate-determining step for its relevant nucleophilic substitution. Using this method, the rate-determining step for NS1 was determined to be the LgR1 step, with an activation energy of 28.9 kJ mol^{-1}. For NS2, the rate-determining step was found to be either the Nuc2 step or the LgR2 step with activation energies of 34.1 kJ mol^{-1} and 35.6 kJ mol^{-1} respectively. Due to the insignificant difference in the activation energies for the steps Nuc2 and LgR2, we decided that both steps should be considered to be overall rate-determining steps for the degradation of HEMT via *Is*-PETase.

Our calculated activation energy is significantly lower than the activation energies calculated for this reaction using different methods[29–31]. As discussed in Chapter 3.1, the previous QM/MM studies disagree with each other on a rather fundamental level which is not shown in their values for activation energy. Methodological issues have been identified with each previous study; these are discussed in detail in Chapters 3.1.1.2, 3.1.1.3, and 3.1.1.4. Our single-point energy calculations have also been performed at the highest level of theory for this reaction. Combined, these factors lend confidence that our results are accurate, despite discrepancies with previous studies.

Comparison with QM/MM studies[97–99] of hydrolases reveals that our calculated energy barrier for *Is*-PETase is lower than expected for an enzyme with a relatively slow reaction rate. In other words, enzymes with far faster reaction rates have been calculated to have higher activation energies for their respective reactions. This suggests that it is not the chemical steps in the degradation of PET via *Is*-PETase that is the overall rate-determining step. By process of elimination, the true rate-determining process is therefore either substrate-binding or product dissociation. Our calculations revealed the presence of a highly stable after-product state (discussed in Chapter 3.3.6), caused by a

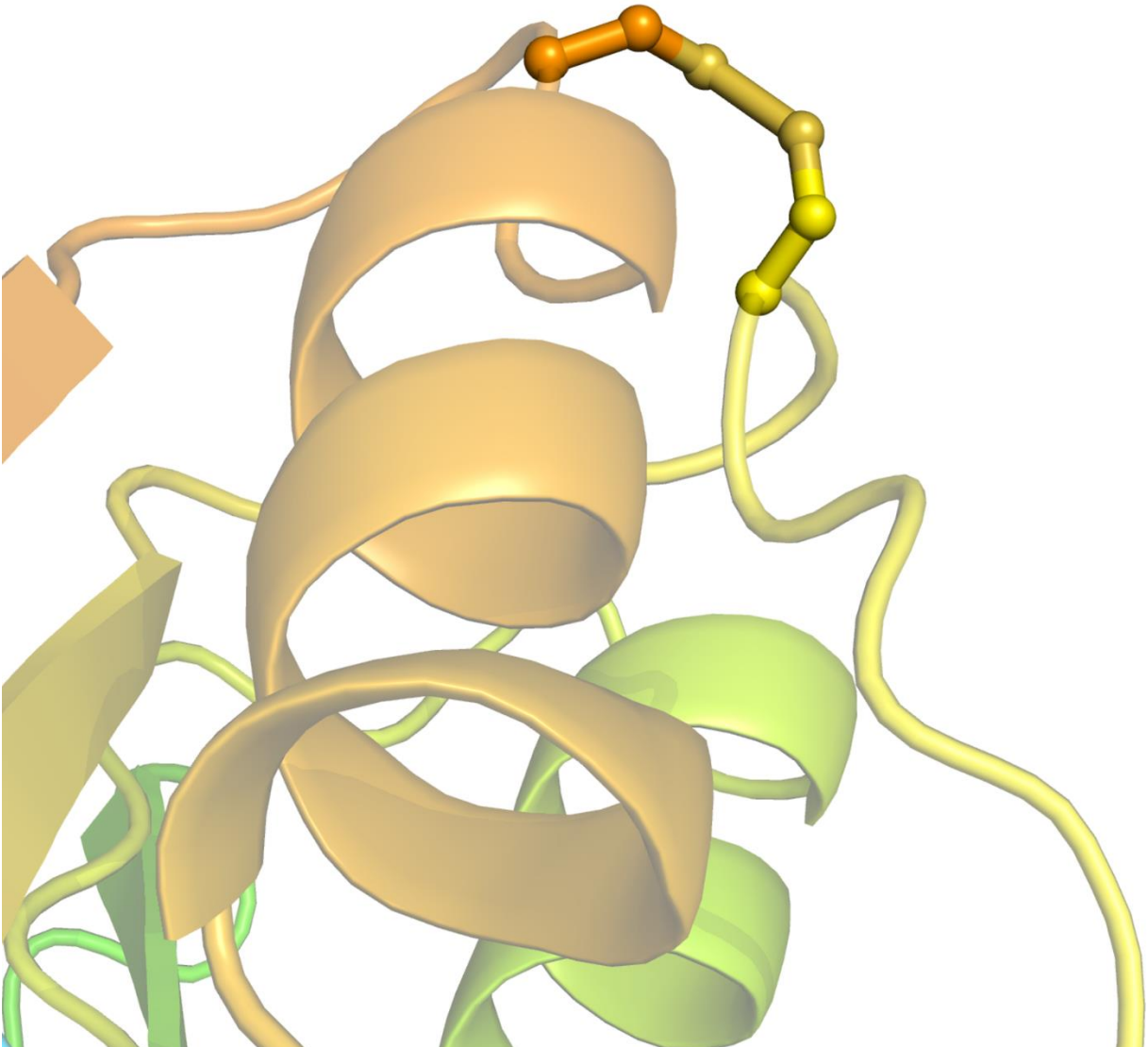
proton transfer from the newly formed carboxylic acid group of the product to the catalytic Ser131 hydroxyl. This proton transfer would temporarily deactivate the enzyme and provides some preliminary evidence for product inhibition.

Early attempts to improve the activity of *Is*-PETase focussed on making single-point modifications to the residues along the enzyme's binding pocket. These mutations generally introduced additional hydrophobic residues to the binding pocket to increase the enzyme's affinity for the highly hydrophobic PET substrate. While the most successful of these modifications yielded only modest improvements to the enzyme's reaction rate, these modifications serve to suggest that the rate-limiting process is substrate-binding, rather than product dissociation.

Greater success in improving the activity of *Is*-PETase has been achieved through improving the enzyme's thermostability. The most popular strategy has been to introduce additional hydrogen-bonding interactions between the enzyme's loop regions that are not too close to the active site. The additional thermostability allows the enzyme to function at higher temperatures where the polymeric PET substrate is more highly mobile. This makes more regions of the PET accessible to the enzyme.

Chapter 4: Exploration of the Thermostability of *Is*-PETase Variants

During her Master's Project, Sylvia Kaempf helped develop some of the methods used in the section of this chapter that explores the thermostability of Is-PETase variants using molecular dynamics simulations. None of Sylvia's data is included in this thesis.



4.1 Background Information

4.1.1 An Introduction to Protein Stability

A protein's stability is defined as its ability to maintain its native folding state under various denaturing conditions. It is important for a protein to maintain its native fold for it to retain its biological function, for example an enzyme will only be catalytically active when in its native folded state. Numerically, a protein's stability is defined by the free energy of unfolding:

$$\Delta G_{unfolding} = -RTlnK_{eq}$$

Equation 83: Gibbs energy of unfolding

In Equation 83, R is the universal gas constant, T is the temperature and K_{eq} is the equilibrium constant between correctly folded and unfolded states for a given protein. The equilibrium constant K_{eq} is defined as the ratio of the concentrations of unfolded and native protein at equilibrium:

$$K_{eq} = \frac{[Unfolded]_{eq}}{[Native]_{eq}}$$

Equation 84: The equilibrium constant for protein unfolding

Given the importance of protein stability for biological function, it is perhaps surprising that most proteins are only marginally stable with $\Delta G_{unfolding}$ values as low as 20 kJ mol^{-1}[100]. Since the number of unfolded and incorrectly folded (if the native fold is considered to be 'correct') states greatly outnumber the native state, folding in the native state has a significant entropic penalty. This entropic penalty has been calculated to be up to 300 kJ mol^{-1} (assuming an ambient temperature of 298 K) for a 100-residue protein and would be significantly higher for larger proteins such as enzymes[100]. The entropic penalty must be overcome by a series of favourable enthalpic interactions in the native state that are greater than those in unfolded or incorrectly folded states.

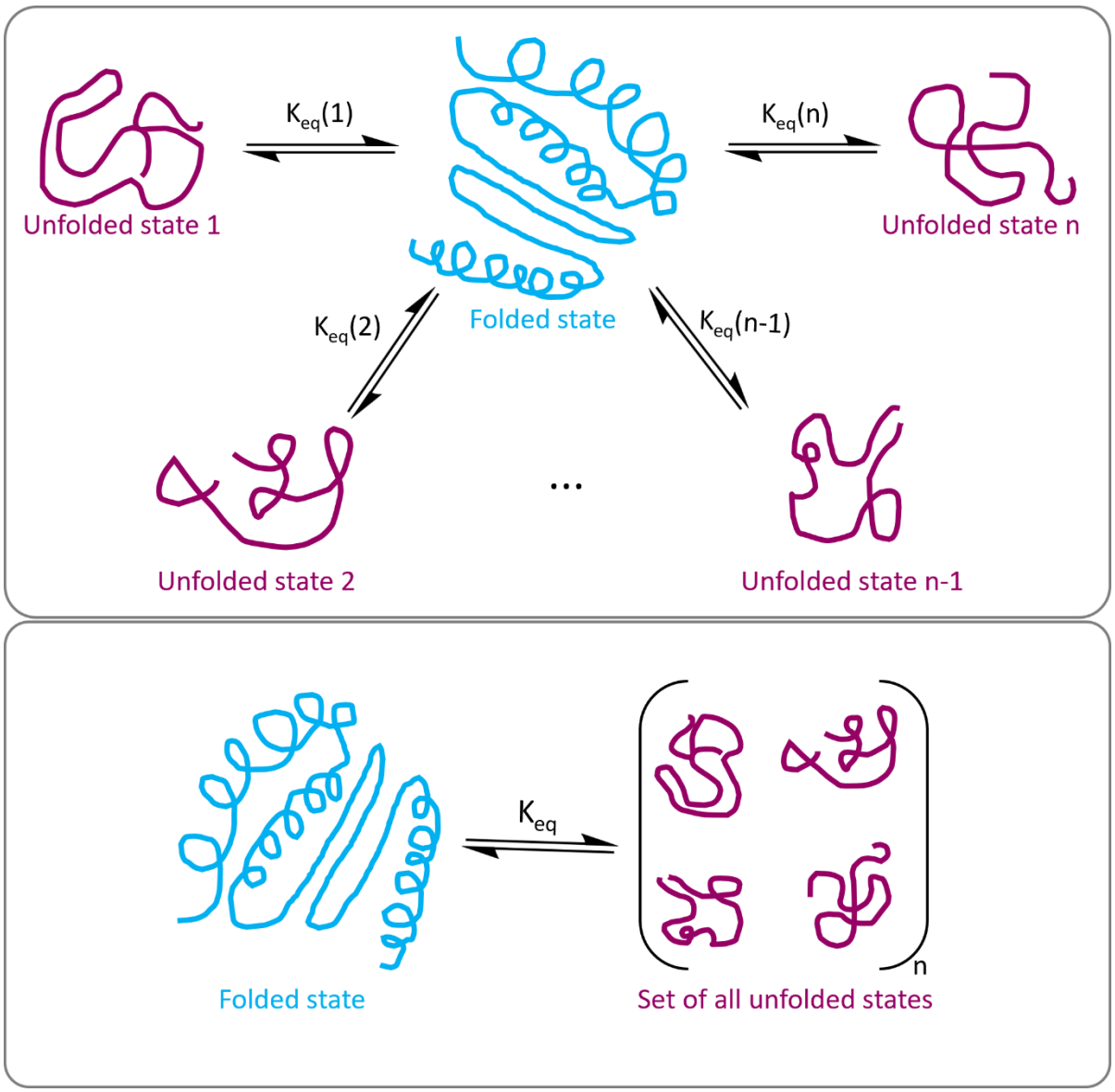


Figure 44: Scheme showing the equilibrium that exists between a protein in its correctly folded state and a large number of unfolded and incorrectly folded states (top). The equilibrium constant for folding and unfolding therefore describes the equilibrium that exists between a protein in its correctly folded state and the set of all possible unfolded states.

There are several reasons why proteins have evolved to be marginally stable. One reason is that a protein's biological function is rarely to simply have high protein stability; the function of a protein may in fact reduce its stability. For example, an enzyme will have reduced stability due to a solvent exposed, hydrophobic active site (as is the case with *Is*-PETase). A compromise is therefore made between protein stability and biological function. Another reason for the marginal protein stability is that it may be beneficial for a protein to exist in several states and have these states energetically accessible. For example, a protein may have to unfold from its native state to pass through a cell membrane, then re-form to its native state afterwards. Another benefit of marginal protein stability

is that for regulatory purposes. If a protein is too stable, it may persist in a cell for a longer period than is necessary and may build up to a point that it negatively impacts the functions of its cell.

Another reason marginal protein stability has evolved is that below a certain threshold of $\Delta G_{unfolding}$, the vast majority of proteins will be folded in their native states. This creates a minimal amount of selection pressure for proteins to become more stable. In the case of proteins produced by thermophiles (organisms that live in high temperature environments, such as hydrothermal vents[101]), higher barriers to unfolding are required to prevent denaturation at the elevated temperatures of their surroundings. Several structural features exist that enhance the thermal stability of such proteins. This section will detail the most important interactions that contribute towards protein thermostability, most of which are exploited by thermophilic organisms.

Whilst there may be reasons for marginal stability to have evolved in nature, for industrial applications of enzymes' increased tolerances to heat, non-physiological pHs, or solvents other than water are desirable. Ideally, we would be able to find enzymes from thermophilic hosts capable of catalysing any desired chemical transformation. In reality, it is often the case that the desired catalytic function is found in enzymes originating from non-thermophilic hosts. Modifications to these enzymes are often required to ensure that they operate efficiently at the elevated temperatures often present in industrial bioreactors.

To engineer proteins with useful activities (such as enzymes) to have improved stability, an understanding of the interactions that contribute to stability is required. The following sections will detail the key factors that determine the stability of proteins and how these factors can be used to engineer more stable enzymes.

4.1.1.1 Hydrogen Bonding

Hydrogen bonds involve a hydrogen bond donor (an electronegative atom with a bound proton) and a hydrogen bond acceptor (an electronegative atom with a lone pair of electrons). Hydrogen bonding is highly directional and requires a strict geometry to minimise (i.e., increase the strength) its interaction energy (see Figure 45). High-accuracy DFT calculations performed on a series of small hydrogen bonding molecules has shown that this interaction occurs at a distance of between 2.3 Å and 3.3 Å (measured heteroatom to heteroatom)[102]. In proteins patterns of hydrogen bonding interactions are used to enforce secondary structures such as α-helices and β-sheets as well as to form short-range interactions between adjacent secondary structures[103]. Due to the tight requirements imposed upon the orientations of the donor and acceptor residues, hydrogen bonds have a significant entropic penalty associated with them resulting in small contributions to the overall $\Delta G_{unfolding}$ of between 2 kJ mol^{-1} and 9 kJ mol^{-1}[104]. Despite the relatively small contribution

of each hydrogen bond, hydrogen bonding is highly important to the stabilisation of proteins due to their sheer numbers. In proteins, hydrogen bonding interactions are predominantly intramolecular interactions involving backbone carbonyl and amide functional groups as well as hydrogen bond accepting and donating groups on various amino acid sidechains. Another type of hydrogen bonding interaction that can contribute to protein stability is the water-mediated hydrogen bond (WMHB)[105]. In this interaction, two residues form hydrogen bonds with a single water molecule, situated in between (see Figure 45). The addition of this bridging water molecule allows hydrogen bonding interactions to occur over longer distances.

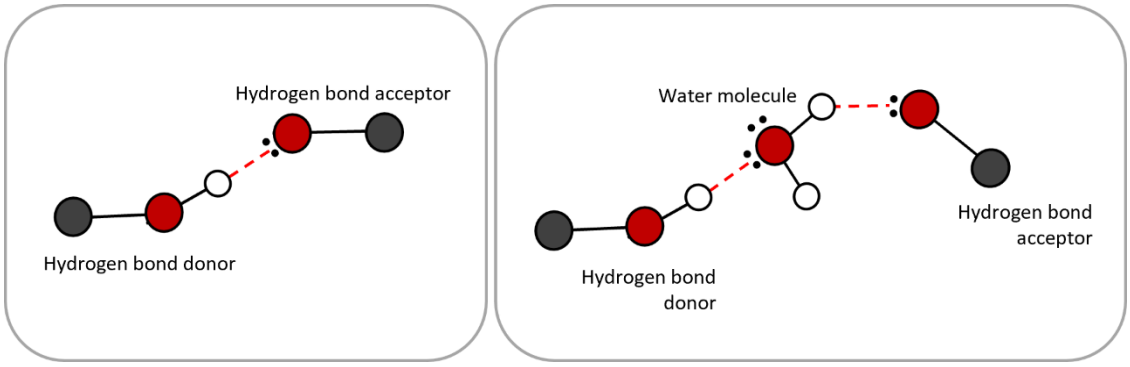


Figure 45: Scheme demonstrating the hydrogen bonding between a hydrogen bond donor and a hydrogen bond acceptor (left). Scheme demonstrating a water-mediated hydrogen bond between a hydrogen bond donor and a hydrogen bond acceptor (right). In these schemes the heteroatom (red) is an oxygen atom, although this scheme could apply to other electronegative heteroatoms (in proteins this would just be nitrogen atoms).

A perfect example of the use of hydrogen bonding interactions to increase the thermostability of enzymes shows up in PETase research (the benefits of which are discussed more fully in section 1.1.3). Son *et al.* introduced the double mutation S121D/D186H into wild-type *Is*-PETase. The expectation was that a new Asp121-His186 hydrogen bond would form. However, X-ray crystallography revealed a distance between these residues of 3.9 Å. This distance is too long for a strong hydrogen bonding interaction to form. With this information, the double-mutant S121E/D186H was made. The longer side chain of Glu121 was expected to allow its carboxylate functional group to get close enough to the sidechain of His186 to form a WMHB. X-ray crystallography of this mutant revealed that Glu121 formed a WMHB with the more distal Asn172 residue instead (see Figure 46). While the initial S121D/D186H mutations provided some additional thermostability to the enzyme (a rationale for this is not provided by the paper), a greater degree of stabilisation was afforded from the S121E/D186H mutations. This example outlines the complexities and inherent unpredictability of improving enzyme thermostability using rational design techniques.

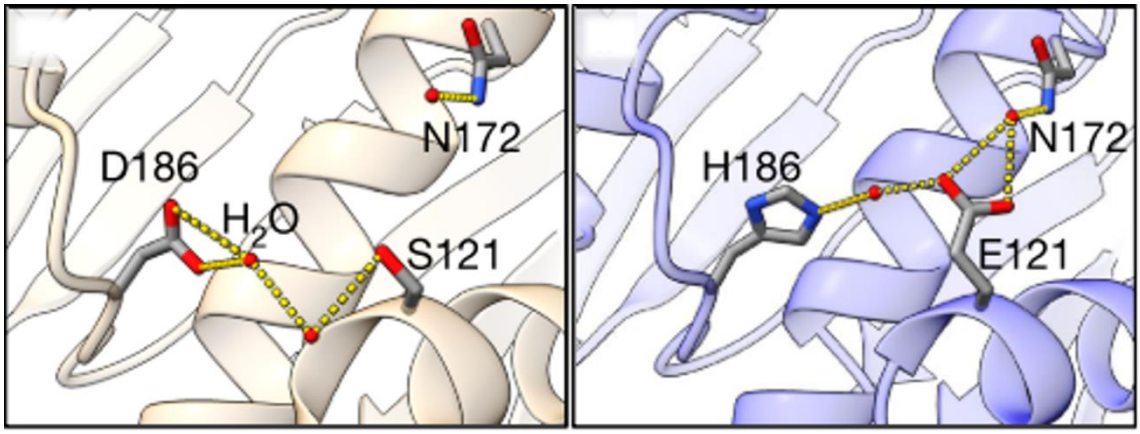


Figure 46: Water-mediated hydrogen bonding networks for residues 186, 121 and 172 in WT-PETase and FAST-PETase, obtained from X-ray crystallography. Figure reproduced from the Lu et al.[37].

4.1.1.2 Hydrophobic Effect

The dominant driving force contributing to protein stability is the hydrophobic effect, which is estimated to account for ~60% of the driving force to fold a globular protein into its native state[104]. The hydrophobic effect describes the entropic and enthalpic stabilisation gained by burying hydrophobic portions of a protein in its core. The most obvious contribution to this stabilisation effect is the increased Van der Waals interactions between non-polar moieties within this core. This however is not the dominant factor in the hydrophobic effect. The liberation of 'frustrated' water molecules from interactions with non-polar portions of the protein has been shown to have a far larger effect[106]. When solvent water molecules interact with a non-polar portion of a protein, they are unable to form the normal four hydrogen bonds with other water molecules in the bulk solvent. When a protein folds to bury its non-polar moieties, these 'frustrated' water molecules can then form hydrogen bonding interactions with either polar portions of the protein or other solvent water molecules; this results in more overall hydrogen bonds being made, stabilising the system. This is the primary enthalpic component of the hydrophobic effect. The hydrogen bonds that 'frustrated' water molecules form are stronger than those of water molecules in bulk solvent. This has an immobilising effect on these 'frustrated' water molecules. When the 'frustrated' water molecules are liberated by burying non-polar moieties of a protein through folding they revert to forming more, weaker hydrogen bonds which results in increased mobility. This is the primary entropic component to the hydrophobic effect.

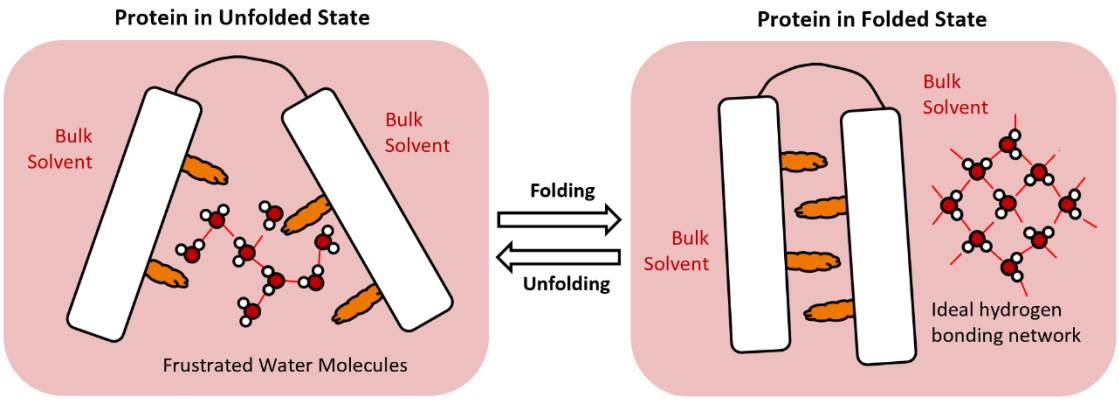


Figure 47: Scheme demonstrating the hydrophobic effect. Regions of a protein are shown in white, while hydrophobic residues are shown in orange. Red dashes in between water molecules indicate hydrogen bonding.

Protein engineers can use the hydrophobic effect to increase and enzyme's thermostability. Unfortunately, it is difficult to predict mutations that will provide stabilisation by exploiting the hydrophobic effect. Instead, the consensus approach identifies highly conserved residues amongst an enzyme's homologs, then reintroduces any of these residues that are not present in the enzyme. The rationale behind this method is that highly conserved residues are likely to have stabilising effects. A modification of this method is to only consider thermophilic homologs when identifying conserved residues. All homology approaches have the inherent weakness that conserved residues may be due to relatedness of the enzymes considered, rather than any inherent stabilisation effect they might provide[104]. An example of this is a study that identified nine conserved residues through homology modelling with a pectase lyase. Of these nine conserved residues, the introduction of only one resulted in any appreciable stabilisation effect[107]. This example shows both the utility and limitations of the consensus method for stabilising enzymes.

4.1.1.3 Salt Bridges

A salt bridge is defined as the favourable interaction between a positively charged, protic atom and a negatively charged aprotic atom. The energetically favourable interaction can be viewed as a combination of a hydrogen bonding interaction (giving the interaction directionality) and a Coulombic interaction. In proteins salt bridges are formed between positively charged lysine and arginine residues (and in some cases the doubly protonated state of histidine) and one of the two negatively charged residues, aspartate, or glutamate (see Figure 48).

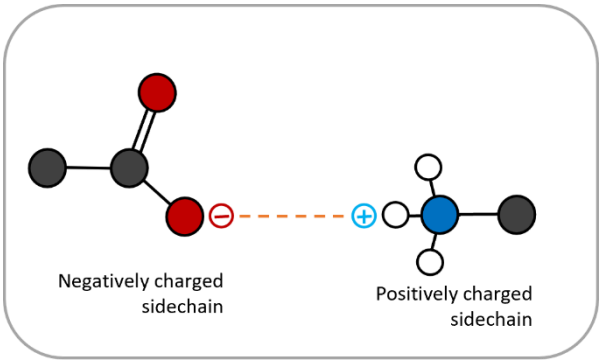


Figure 48: Scheme showing a salt-bridge type interaction between the negatively charged sidechain of either an aspartate or glutamate residue and the positively charged sidechain of a lysine residue.

Salt bridges contribute a small but significant stabilisation effect, increasing the magnitude of $\Delta G_{unfolding}$ by around 3 kJ mol^{-1} for solvent exposed salt bridges[108] and by up to 20 kJ mol^{-1} for salt bridges buried in the core of the protein[109]. Buried salt bridges are however rare due to the large penalty for housing a pair of charged residues in a largely hydrophobic protein core. The low stabilisation effect of salt bridges is due to the favourable interactions that charged residues can form with solvating water molecules in the protein's unfolded state or misfolded states. Thus, the presence of salt bridge forming residues stabilises both the desired native state and the undesired state and has a small overall effect on $\Delta G_{unfolding}$. Interestingly, salt bridges have a stronger stabilising effect at higher temperatures. This is because at elevated temperatures the interaction between the solvating water and the charged residues weakens, while the Coulombic interaction of the salt bridge in the native state is unaffected. The result of this is that the unfolded states receive less stabilisation at higher temperatures than the native state, increasing the magnitude of $\Delta G_{unfolding}$. Due to this effect, salt bridges are more often found in proteins from thermophilic organisms.

4.1.1.4 Cation-π Interactions

The cation- π interaction is a strong, non-covalent, electrostatic interaction between a cationic species and a system with π-bonding, typically occurring when the cation is within 6 Å of the centroid of an aromatic ring. Whist this effect can be observed in non-aromatic double bonds, this effect is most commonly observed in aromatic systems. While aromatic rings such as benzene have no permanent dipole moment, they do have a permanent quadrupolar moment (two dipole moments oriented to produce no net dipole moment). This results in electron density being localised above and below an aromatic ring in the π-bonding orbitals. This results in a partial negative charge above and below the plane of an aromatic ring. This partial negative charge forms an attractive electrostatic interaction with nearby positively charged cations. In proteins, the aromatic residues phenylalanine, tyrosine and tryptophan can all participate in cation- π interactions. It has been found

that small cations with high electron density have stronger interactions with π-systems, than large electron-diffuse cations. Due to the broadly hydrophobic nature of aromatic residues, cations that are similarly hydrophobic have stronger overall interactions with these residues (although the electrostatic interaction itself is unchanged by the polarity of the ion, association with an aromatic residue will incur greater enthalpic costs for a hydrophilic cation)[110].

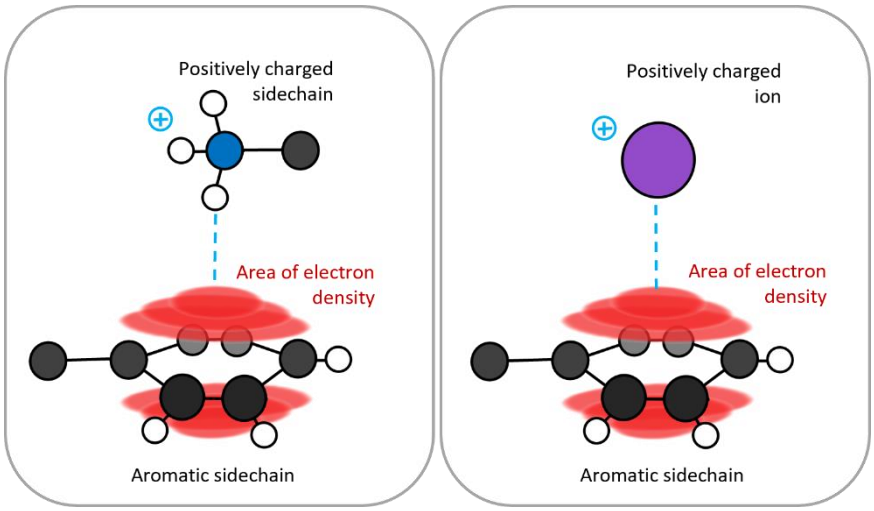


Figure 49: Scheme showing cation-π interactions between the aromatic sidechain of a phenylalanine residue and the cationic sidechain of a lysine residue (left) and a sodium cation (right).

The cation-π interaction is often utilised in nature by receptor proteins. The most studied of these receptor proteins is acetylcholine esterase[111] (AChE) which is responsible for the hydrolysis of the neuro-transmitter acetylcholine (ACh). ACh has a quaternary trimethyl ammonium which binds to the Trp84 residue in AChe through a cation- π interaction. The active site of AChe is situated at the bottom of a deep pocket which is lined with aromatic residues. These are thought to guide ACh into the active site.

While cation-π interactions are most often discussed in relation to ligand binding, intra-molecular cation-π interactions can contribute to protein stability. Interactions between the cationic side chains of lysine and arginine residues or the doubly-protonated state of histidine can form cation-π interactions with any of the three aromatic amino acid side chains. The cationic states of lysine and histidine are highly polar and hydrophilic. While the actual cation-π interaction between an aromatic ring and these two residues is fairly strong, a significant enthalpic penalty must be paid to form this interaction, incurred by breaking hydrogen bonding interactions with solvating water molecules.

Interestingly, despite being generally considered as the most hydrophilic of the naturally occurring amino acids, arginine forms stronger interactions with aromatic residues than lysine. This is due to several structural features of the guanidinium ion that makes up the side chain of arginine. The guanidinium ion on the side chain of arginine consists of a single sp^2 hybridized carbon atom bound

to three sp^2 hybridized nitrogen atoms; this leads to a planar structure. The orientations of protons bound to the nitrogen atoms lead to hydrogen bond formation with solvating water molecules occurring only in-plane, relative to the guanidinium ion. This allows for dehydration at the 'faces' of the ions to occur with a relatively small energy barrier. Additionally, the guanidinium 'head' of the arginine sidechain is sp^2 hybridised, containing six π-electrons. As such, the sidechain of arginine behaves as a Y-aromatic system. This π-system can interact with aromatic residues through π-π stacking type interaction, further increasing the strength of this interaction. It has been theorised that this effect leads to the majority of cation-π interactions found in the protein data bank to be involving arginine residues[112]. The strength of a cation-π interaction is estimated to be between 6.3 kJ mol^{-1} and 20 kJ mol^{-1}[113].

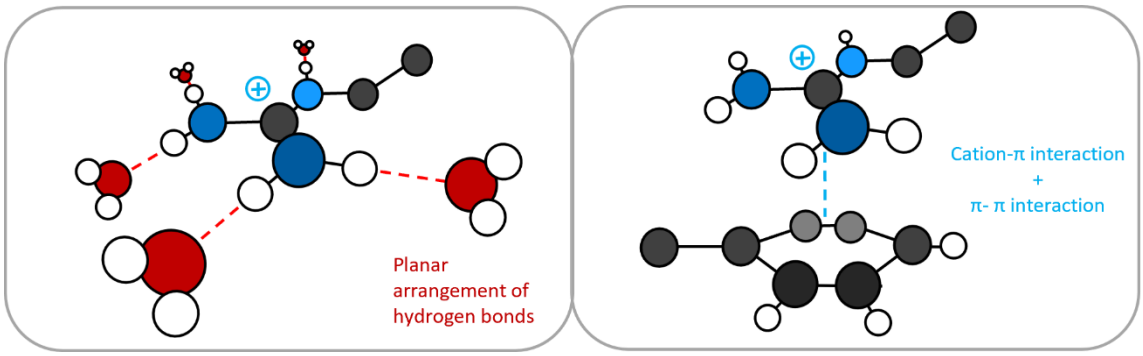


Figure 50: Scheme showing modes of interaction for the sidechain of asparagine as it forms hydrogen bonding interactions with water molecules (left) and cation-π interactions with aromatic rings.

4.1.1.5 Disulfide Bonds

Disulfide bonds are a special case when it comes to protein stabilisation as they are strong chemical bond between two cysteine residues. This bond is formed by the oxidation of the thiol group of two proximal cysteine residues by redox active enzymes, creating a new sulfur-sulfur covalent bond.

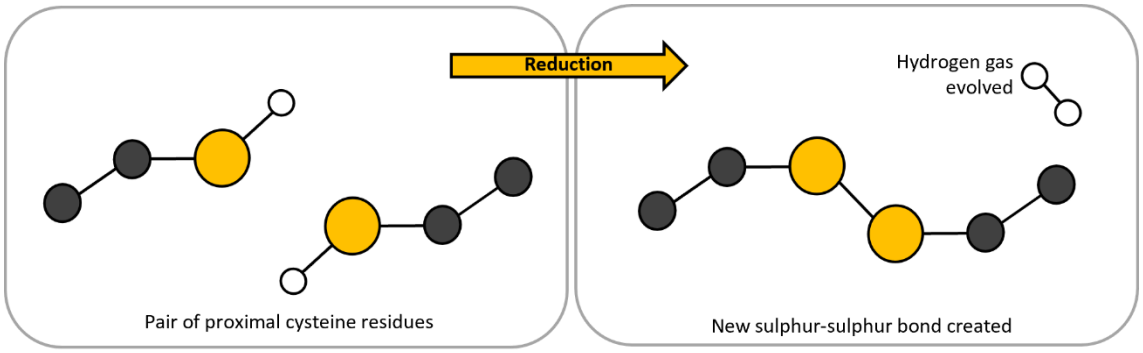


Figure 51: Scheme showing the formation of a disulfide bond under reducing conditions.

Disulfide bonds are formed during the protein folding process and reduce the mobility of residues in the disulfide loop (i.e., residues in-between the two cysteine residues). This reduces the degrees of freedom of the disulfide loop and thus reduces the entropic penalty for the adoption of the native

fold. As a result of this effect, greater stabilisation of the native fold is found when the disulfide loop region is more disordered prior to the formation of a disulfide bond. Based on experimental evidence, an expression for the approximate entropic stabilisation has been proposed[114]:

$$\Delta S = -2.1 - \frac{3}{2} R ln(n)$$

Equation 85: A simple model to calculate the entropic stabilisation afforded by a single disulfide bond.

In Equation 85, R is the universal gas constant and n is the number of residues in the disulfide loop. This is a very simple model and ignores the effect of different residues in the disulfide loop. The stabilisation the native fold provided by a disulfide bond has found to be between 9.6 kJ mol^{-1} and 21.7 kJ mol^{-1}[115]. A more detailed analysis reveals that the flexibility of disulfide loops (as opposed to simply the chain length) prior to the introduction of a disulfide bond can indicate the degree of stabilisation conferred by the disulfide bond. Two explanations are used to rationalise this effect: either this flexibility is useful to accommodate the new bond, or these flexible regions are involved in the early stages of protein unfolding. Under different circumstances, both explanations have been found to be appropriate[114].

In addition to the entropic stabilisation effect, disulfide bonds can provide a kinetic stabilisation effect to protein folds. In cases where a disulfide bond inhibits an early unfolding event, the rate of unfolding is significantly reduced. It has however been observed through analysis with the GeoFold algorithm that disulfide bonds located at regions that are involved in late stages of unfolding can have destabilising effects. In many cases, the kinetic stabilisation of protein fold has been found to be more important that the entropic effect described above. This can be rationalised by considering that proteins do not exist in a state of rapid folding and unfolding. Instead, the unfolding process can be considered to be irreversible, resulting in the permanent denaturing of the protein. In this case where an unfolded state cannot recross to a folded state, the most important parameter becomes the rate of unfolding.

To protein engineers who wish to enhance the thermal stability of an enzyme, it is important to select the ideal site to introduce a novel disulfide bond to maximise its stabilising effect. A study used β-factors determined by X-ray crystallography to select several sites for novel disulfide bonds in the enzyme *Candida antarctica* lipase B (CalB). β-factors are a term used to express a 'smearing' of electron density in X-ray crystal structures, where large β-factors indicate regions of high flexibility. The algorithms MODIP[116] and Disulfide by Design (DbD)[117] were also used to select locations for novel disulfide bonds. MODIP is an algorithm that uses geometric data from the protein data bank (PDB) to predict sites where novel disulfide bonds will provide the most stability, based upon

putative orientations and steric parameters of the newly introduced cysteine residues. DbD uses a similar method, using a predicted torsion angle based on the distance between the β-carbons to assess the best sites to place novel disulfide bonds. It was found introduction of disulfide bonds at sites with high β-factors resulted in the greatest stabilisation effect. Molecular dynamics simulations of CalB with computationally introduced disulfide bonds at various positions were performed at high temperatures (600 K). It was found that a disulfide bond in mutant CalB A162C-K308C reduced the RMSD of residues within 5 Å when compared to the wild-type enzyme.[118] The stability of this new mutant enzyme was determined experimentally. It demonstrated a half-life 4.5-fold longer than the wild-type enzyme. This study shows the utility of disulfide bonds to improve the thermostability of enzymes.

Upon reviewing the several methods that can be employed to enhance the thermostability of enzymes, we decided to introduce a novel disulfide bond to the enzyme *Is*-PETase. Details of this project are provided in section 4.3.1.

4.1.2 Rational Design of Enzymes to Improve Thermostability

When designing biocatalysts for an industrial process, thermostability is considered an important feature. High kinetic stability allows the enzyme to function at higher temperatures. This provides benefits such as increased reaction rates, enhanced catalyst lifetimes, improvements to substrate solubilities and reduces the risk of microbacterial contamination[115].

There are two main routes to improving an enzyme's thermostability: Directed evolution and rational design. Directed evolution is a technique where a large library of mutant enzymes is generated using error-prone peptide-chain-reaction (PCR) methods, then assayed for a desired property (in our case, this would be protein thermostability). The most successful mutations may be combined, then used as a starting point for successive rounds of directed evolution. This technique has proved to be extremely effective in enhancing reaction rates and enzymatic thermostability as well as in modifying substrate specificities. The success of directed evolution was recognized formally by the awarding of the Nobel Prize in Chemistry to Francis Arnold in 2018[38]. Directed evolution does not require any *a priori* knowledge of protein structure and may result in beneficial mutations that would be difficult to predict or even seem counter-intuitive to a human. Instead, the challenges in directed evolution tend to centre around creating a fast-throughput assay that can rate the desired property of each mutant produced as quickly as possible. The faster the assay, the larger the libraries of mutants that can be tested at the same time. This increases the chance that beneficial mutations will occur.

By contrast, rational design is an approach that uses our knowledge of protein structure (discussed in the previous section) to predict mutations that will have beneficial effects to a desired property. At the most basic level, this can be done visually by a researcher with knowledge of protein structure and the enzyme in question. Computational methods such as molecular docking, molecular dynamics simulations and QM/MM techniques can be used to guide the rational design process.

This section will discuss several methods where the rational design approach has been used to enhance the thermostability of enzymes.

4.1.2.1 Sequence-Based Methods used to Enhance the Thermostability of Enzymes

Multiple sequence alignment (MSA) is a technique where close structural homologs are found for a target enzyme. The sequences of these homologs are then compared to that of the target enzyme. Residues that are conserved in the homologs, particularly if they are from thermophilic organisms, and are not present in the target enzyme are then identified. Often, these conserved residues have a stabilising effect in their native enzymes, so it is assumed that they may have the same effect in the target enzyme. Thus, the conserved residues are introduced into the target enzyme.

An example of MSA being used can be seen in the enhancement of the thermostability of the enzyme *Bacillus lipase*[119]. By the introduction of a single P247S mutation, a 60-fold increase in thermal stability at 60°C and a 12-fold increase in catalytic efficiency was reported. A sequence comparison search called basic local alignment search tool (BLAST) was performed and revealed that *B. lipase* shared a 96% sequence similarity with the thermostable enzyme *B. stereothermophillus*, with only 16 residues that were not in common. Of these 16 non-conserved residues, a surface exposed proline residue Pro247 in *B. lipase* was identified as a candidate for mutation. In *B. stereothermophillus* this residue is a serine, so the P247S mutation was made in B*. lipase*. A 3D structure of wild-type *B. lipase* and the P247S variant was generated using homology modelling techniques. Analysis of the structures revealed that the proline residue in the wild-type enzyme formed a single hydrogen bond with an adjacent Thr251 residue, while the novel serine in the P247S variant was able to form three hydrogen bonds with Thr251. This results in a net-gain of two hydrogen bonds through one single-point mutation.

Another example of the use of MSA is the improvement of the thermal tolerance of a fungal GH11 xylanase[120]. Analysis of β-factors in the X-ray crystal structure of the fungal GH11 xylanase XynCDBFV revealed a particularly disordered region between residues 87 and 90 (with sequence Glu-Arg-Ser-Ser). MSA identified a highly conserved Arg-Gly-His-Thr pattern of residues at this position in enzymes with homologous sequences. Four mutant enzymes were expressed by making the single point mutations Q87R, N88G, S89H, and S90T. The thermostability of each mutant was assessed,

with each mutant retaining higher levels of activity at temperatures up to 65°C than the wild-type XynCDBFV enzyme. Despite this increase in protein stability, each mutant enzyme displayed reduced catalytic efficiency compared to the wild-type enzyme. Increased K_M values for each of the mutants suggest that these mutations may have interfered with substrate binding. Molecular dynamics simulations were performed to rationalise the improvements in thermostability caused by each mutation. It was found that the Q87R mutation created a pair of strong hydrogen bonding interactions between the guanidinium side chain of Arg87 and an adjacent Asp215 residue. In the N88G mutant an additional localised hydrogen bond between the side-chain hydroxyl of Ser89 and the backbone carbonyl of G88 was observed; this was not observed in the wild-type enzyme. For the S89H variant, the bulky His89 residue is significantly displaced compared to the wild-type Ser89 residue. This resulted in two additional hydrogen bonds forming between the side-chain carbonyl of Asn88, the backbone nitrogen of Ser90 and the side-chain hydroxyl of Ser90. The S90T variant resulted in a hydrogen bonding network involving the side chains of Thr90 and Asn88 and the backbone nitrogen atoms of Thr90 and Ser89. This study shows the use of molecular dynamics simulations *post hoc*, allowing us to rationalise successful mutations predicted by sequence-based methods.

4.1.2.2 Structure-Based Methods to Enhance Enzyme Thermostability

While sequence-based methods have been proven to be effective in enhancing an enzyme's thermostability, they do not allow us to apply our *a priori* knowledge of protein structures and cannot produce novel stabilising motifs. By analysing a target enzyme's 3D structure before making mutations, we can predict modifications to the enzyme's sequence that may have stabilising effects without relying on these modifications being present in homologous enzymes.

The enzyme *Rhizomucor miehei lipase* (RML) had its thermostability greatly improved through a series of mutations introduced by structural rational design methods[121]. The T18K/T22I/E230I/S56C-N63C/V189C-D238C mutations were introduced and resulted in a 14.3°C increase in T_m and a 12.5-fold increase in half-life at 70°C. Three computational algorithms were used to identify mutations that could be improve RML's thermostability (Rosetta ddg_monomer, FoldX, and I-Mutant). Mutations that were predicted by all three algorithms were further processed. Of the 60 mutations predicted by all three algorithms, 24 mutations were discarded due to predicted interference with salt-bridges, disulfide bonds, or the active site of the enzyme. The remaining 36 mutants were made experimentally. Mutant enzymes with at least a 1°C improvement in T_m were then further characterised. Mutations that resulted in particularly large improvements in T_m were combined to create double and triple mutants. Two disulfide bonds at S56C-N63C and V189C-D238C were introduced to RML. The positions of these disulfide bonds were predicted by using four specialist

algorithms (DbD2, SSBOND, MODIP and BridgeD). Those too close to the active site were discarded while the rest were retained for further processing. The mutants that were retained were then subjected to analysis with the FoldX algorithm. The mutants with the greatest predicted stabilisation were then analysed visually, resulting in the S56C-N63C and V189C-D238C disulfide pairs being selected for expression. The introduction of these two disulfide bridges to the most thermostable triple mutant resulted in the most thermostable mutant found by this study. This work shows the power of combining multiple different computational methods to improve an enzyme's thermostability.

Another example of this approach can be seen in the work improving the thermostability of *Paenibacillus pasadenensis chitinase*. The mutations S224C-I319C/T259P resulted in an increase in T_m of 7.9°C and a 26.3-fold increase in half-life at 50°C. The mutation T259P was predicted using sequence-based methods; it was found that this Pro259 is conserved in homologous thermophilic enzymes. Potential locations of disulfide bridges were predicted using two algorithms (DbD 2.0 and SSBOND), while the FoldX algorithm was used to select the most promising candidates. Combination of the most promising disulfide bond and the T259P mutation yielded a more thermostable enzyme.

The thermostability of the industrially important microbial transglutaminase enzyme (MTG), which is used as a binder in the food industry, has been improved using a variety of rational design methods. Over the past decade MTG has been used in different industrial applications, but its function has been limited by poor thermostability. To address this shortcoming, one group used rational design methods to introduce a disulfide bond into the N-terminal region of the enzyme. Residue pairs were selected based on the following conditions: They must be at least 10 residues apart from each other in the primary protein sequence, their β-carbon atoms must be at most 5 Å from each other and their mutation to cysteine residues must not remove key hydrogen bonding interactions. Residue pairs that fitted these criteria were then evaluated using three computational algorithms (CNX, QUANTA and SSBOND). Six potential candidates resulted from this process. Of these six candidates, four were selected based on their high β-factors in an X-ray crystal structure; these were expressed and assayed for activity experimentally. The rationale for this is that a greater stabilisation effect can be achieved by inserting the disulfide bridge into a more disordered region of the protein. Each of the mutant enzymes expressed displayed increased stability with respect to the wild-type enzyme although moderate reductions in specific activity were observed for all mutants. The T7C-G58C mutant resulted in the greatest improvement in thermostability. A following study further modified the T7C-G58C mutant through a process of random mutagenesis, resulting in several mutants with higher thermostability than before. Of these mutants the S4P mutation yielded the greatest increase in stability[122]. A different study took this new T7C-G58C/S2P mutant and, using a gene-shuffling

protocol added three further single-point mutations. The resultant variant S2P/S23V/Y34N/S119A/K294L (now called TGm1) displayed greater thermostability than the previous disulfide-containing mutant[123]. The thermostability of TGm1 was further improved by performing a proline scan: Using a custom script combined with Rosetta Cartesian_ddg (a script within the Rosetta[124] suite), proline substitutions were made *in silico* to along chain of TGm1; those mutations that had predicted stabilisation effects were expressed and assayed for activity. The TGm1 mutants with A166P, A256P, and A287P substitutions resulted in increased T_m values. Further improvements to the activity of TGm1 were made using a molecular docking approach: The DS suite of algorithms was used to predict a series of mutations that would improve substrate binding. The thermostability of these mutants was assessed using Rosetta Cartesian_ddg. The most thermostable of these mutants were made experimentally and assayed for activity; the E28T mutant was the only substitution that resulted in a meaningful increase in activity. The E28T mutation was combined with A256P/A287P mutations to yield TGm1-E28T/A256P/A287P (now named TGm2). This new variant displayed roughly three-fold increase in half-life at 60°C and an increase in specific activity when compared to TGm1. The surface charge of TGm2 was increased to reduce protein aggregation at high temperatures. This was achieved by manually selecting surface residues that could be substituted for negatively charged aspartate or glutamate residues then using Rosetta Cartesian_ddg to predict whether each mutation would result in greater enzyme stability. The mutations that were predicted to improve thermostability were combined to produce the final variant TGm2-N96E/S144E/N163D/R183E/R208E/K325E (named TGm3). The TGm3 variant was expressed and assayed, displaying a two-fold increase in half-life at 60°C relative to TGm3 and a further increase in specific activity, resulting in a roughly two-fold increase in specific activity relative to TGm1[125].
This set of studies exemplifies the diverse range of rational design methodologies that are used to enhance enzyme thermostability. Whilst many separate methods have been used, a pattern to the workflow can be clearly seen: Generate a library of potential mutants, screen this library for the desired property, then combine successful mutations to create a more effective variant. This cycle can be repeated until they yield no more improved variants.

4.1.2.3 Computational methods for exploring protein thermostability

In the previous section, techniques for improving an enzyme's thermostability through introduction of mutations have been discussed. To determine the extent to which a mutation increases the thermostability of the enzyme (if at all), experimental techniques such as differential scanning fluorimetry (DSF)[126] can be used to directly measure the mutant enzyme's T_m, which can be compared to that of the wild-type. Methods such as DSF provide us with quantitative values to assess whether mutations are beneficial to an enzyme's thermostability. What these methods do not

provide is any qualitative description as to why a mutation might contribute to an enzyme's thermostability. Knowing why a mutation produces a beneficial effect contributes to our overall understanding of the relationship between protein structure and function and can inform future enzyme engineering.

Molecular dynamics simulations have proven to be useful for describing features that contribute to an enzyme's thermostability. One representative study applied molecular dynamics simulations to a set of tryptophan synthase enzymes (TRPS)[127]. TRPS enzymes from a hyperthermophile (functions at high temperatures), a mesophile (functions at moderate temperatures) and psychrophile (functions at low temperatures) were studied using 30 ns MD simulations at various temperatures. Initially, a simple analysis of the RMSD evolution over time and RMSF by residue was performed. The RMSD analysis revealed that, for all temperatures, the psychrophilic and mesophilic enzymes exhibited steady-state evolution of RMSD. By contrast, the hyperthermophilic TRPS exhibited a gradual rise in RMSD in the highest temperature simulation (375 K). For all TRPS enzymes, an increase in average RMSD was observed as the temperatures of each simulation was increased. RMSF analysis revealed that for each TRPS enzyme two highly flexible loops were found, identified by their high RMSF values. These flexible loops contain active site residues. It has been theorised that the flexibility of an enzyme's active site is crucial for substrate binding and product dissociation events to occur. At low temperatures, the psychrophilic TRPS enzyme exhibited greater flexibility in its active site loops than the other two enzymes. This shows that this enzyme's adaptations to cold environments allow for active site flexibility at low temperatures. By contrast, the mesophilic TRPS had low flexibility in its active site at low temperatures, indicating that it would be inactive at these temperatures. The hyperthermophilic TRPS enzyme exhibited the most flexible active site overall, exhibiting high flexibility at low, moderate, and high temperatures.

More elaborate analysis was performed using principal component analysis (PCA). This method represents the conformational space of each simulation as principal components. By plotting the most dominant principal components against each other, the conformations adopted by each enzyme in each simulation can be examined. For each enzyme, the general trend was that a tighter clustering in conformational space was observed at lower temperatures, becoming more dispersed at higher temperatures. This shows that at low temperatures, the conformation of each TRPS enzyme is stable. At these low temperatures the psychrophilic TRPS exhibited a more diffuse pattern due to its higher flexibility at low temperatures. As the temperature of the simulations increased, the psychrophilic TRPS showed an abrupt increase in dispersion of its conformational space (i.e., a larger number of significantly different conformations were visited throughout the duration of each simulation); this was in agreement with the RMSD data. The mesophilic TRPS showed the tightest

grouping in its PCA plot for all temperatures, indicating that it is the least flexible of the three enzymes. For the hyperthermophilic TRPS, a progressive transition from three to four distinct subspaces was observed. This was explained as thermal motion along a shallow energy landscape.

The free-energy landscape (FEL) of each simulation was generated by Boltzmann-inverting multi-dimensional histograms generated during the PCA analysis. A FEL is constructed using the first two principal components as reaction coordinates. Analysis of the FEL for the lowest temperature simulations revealed that all enzymes remained very close to their initial structure. As the temperature of the simulations increased, the psychrophile developed a broader single minimum. By contrast, the hyperthermophile developed multiple minima, each significantly separated in reaction coordinate space. This could be interpreted as the enzyme having a greater control in its flexibility at high temperatures, resulting in a higher T_{opt} for the enzyme.

So far, the analyses performed have been focused upon the idea of enzyme 'flexibility'. Further analyses of the intramolecular interactions occurring throughout each simulation were performed to provide a more detailed description of factors that contributed towards each enzyme's thermostability. The number of hydrogen bonds present in each simulation was analysed using the concept of duty cycles: a hydrogen bond is considered present if it formed in a pre-defined percentage of a simulation[127]. This pre-defined percentage (or % duty cycle) is then varied to identify short-lived and long-lived hydrogen bonds. It was found that the hyperthermophilic TRPS enzyme maintained more long-term hydrogen bonds at high temperatures than the other two enzymes. Additionally, the hyperthermophilic TRPS enzyme had more overall hydrogen bonds. These two factors are likely to contribute to the greater thermostability of the enzyme.

Upon reviewing the methods used in the studies described in this section, we decided to apply some of these techniques to investigate the thermostability of *Is*-PETase and two of its most thermostable (and catalytically active) mutants from the literature[37,39] (discussed in section 4.3.2).

4.2 Methodology

4.2.1 Introduction of an Additional Disufide Bond to Is-PETase

An additional disulfide bond was introduced to the wild-type *Is*-PETase using the mutagenesis plug-in in the program Pymol[14]. This produced two proximal cysteine residues. During the setup step for our molecular dynamics simulations, the sulfur-bound protons on each cysteine residue were deleted and a new disulfide bond between the two sulfur atoms was created. This process was performed automatically by the pdb2gmx function in the GROMACS program[128].

4.2.2 Molecular Dynamics Simulations

Molecular dynamics simulations were performed to solvate, add counter-ions, and equilibrate the novel *Is*-PETase variant using the same methods described in section 3.1.

For the analysis of the thermostabilised variants of *Is*-PETase HOT-PETase, FAST-PETase and TS-PETase, the following molecular dynamics scheme was used:

4.2.2.1 Solvation, Addition of Counter-Ions and Equilibration of Hydrogens

Following the same scheme as outlined in section 3.2.2.2, each enzyme was placed in a water box containing TIP3P[78] water molecules (ca. 15000 water molecules were used for each enzyme). Counter-ions were added to balance the overall system charge by replacing water molecules with chloride ions. Six, five, and four chloride counter-ions were used to balance the charges of WT-PETase, FAST-PETase and HOT-PETase respectively.

An energy minimisation step followed by short simulations using NVT and NPT ensembles respectively was performed using the same protocol as described in section 3.2.2.3.

4.2.2.2 Production Molecular Dynamics Simulations

Production molecular dynamics simulations were performed using the same protocol as described in section 3.2.2.3. Modifications to this protocol were as follows: For each enzyme, a set of ten MD simulations were performed at 300K, as well as ten MD simulations performed at 360K. No ligands were present in these simulations, as such only two temperature coupling groups were required. An initial 10 ns molecular dynamics simulation was performed to equilibrate the system. This was followed by a long 500 ns molecular dynamics simulation. This longer molecular dynamics simulation was analysed to identify features in the *Is*-PETase variants that contribute to their thermostability.

4.3 Results and Discussion

4.3.1 Introduction of a novel disulfide bond into Is-PETase

From our QM/MM calculations performed on wild-type *Is*-PETase (discussed fully in chapter 2), we concluded that the activation energies within the chemical steps of *Is*-PETase's mechanism were sufficiently small such that they did not form the true rate-limiting step in the degradation of PET. Instead, the physical process of substrate binding or product dissociation must therefore be rate-limiting. As discussed in section 1.1.4, there have been several studies where the activity of *Is*-PETase has been significantly increased by augmenting the enzyme's thermostability. A popular method to achieve this has been to introduce new disulfide bonds into the structure of enzymes.

We decided to introduce a novel disulfide bond into the structure of wild-Type *Is*-PETase. In order to achieve this, a pair of residues must be substituted for cysteine residues. This residue pair must be suitably close in cartesian space such that a disulfide bond can be formed. The orientation of the residues pair's side-chains must also be 'facing' each other (i.e., their β-carbons must be closer to each other than their α-carbons). Another consideration is that the novel disulfide bond must not interfere with the catalytic mechanism of the enzyme; as such it should not be introduced within 10 Å of any catalytic residue. As discussed in section 4.1.1.5, greater stabilisation can be achieved through introduction of disulfide bonds when they are introduced to more disordered regions of a protein. Using this as a guide, the novel disulfide bond should be introduced between two loop sections of *Is*-PETase, rather than to α-helices or β-sheets which are more rigid.

Using the criteria outlined above, a new disulfide bond was created by applying the following mutations to wild-type *Is*-PETase: S163C-S194C (see Figure 52). This new mutant will be referred to as DISU-PETase for the remainder of this thesis. For clarity, for the remainder of this thesis, the wild-type variant of *Is*-PETase will be referred to as WT-PETase.

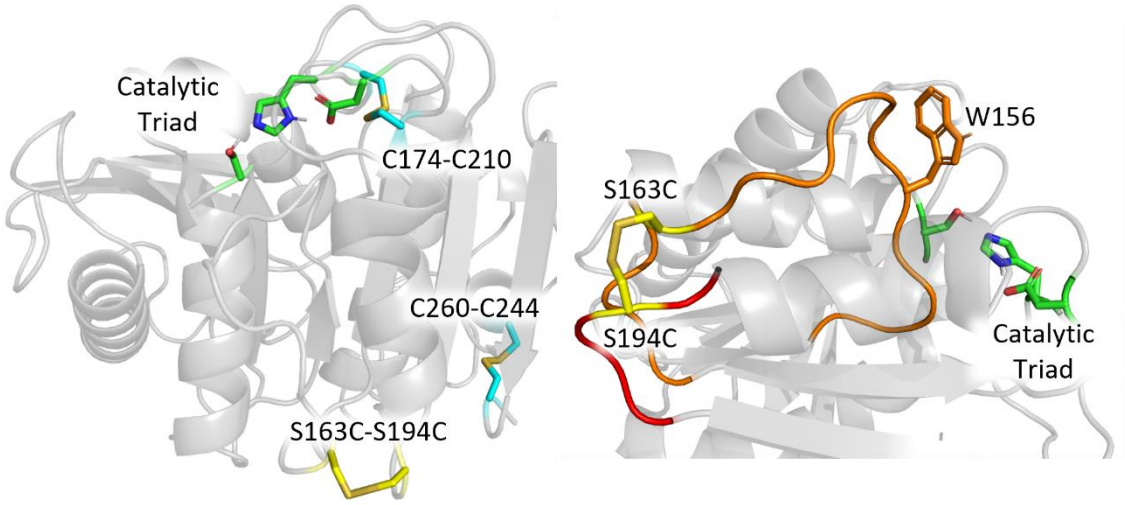


Figure 52: Location of novel S163C-S194C disulfide bond (yellow) in relation to existing disulfide bonds in WT-PETase (left). Loops connected by the S163C-S194C disulfide bond (right). The β5-β6 loop is shown in orange, while the α5-β7 loop is shown in red. The β5-β6 loop contains active site residue Trp156 (shown in orange).

The new S163C-S194C disulfide bond creates a new covalent bond between the loop connecting β5 and β6 sheets (loop 163) and the loop connecting α5-helix and β7-sheet (loop 194). Loop 163 is long, with a length of 13 residues (from residue Gln151 to Pro166) while loop 194 is shorter with a length of only 6 residues (from residue Met193 to Lys199). The loop created by connecting residues 163 and 194 is 31 residues long. Using Equation 85 this should result in an entropic stabilisation of -13.5 kJ mol^{-1}.

As shown in Figure 52, the β5-β6 loop contains the active site residue Trp156. This residue has been shown to be important to the activity of *Is*-PETase and has been suggested to play a crucial role in substrate binding[18,21,22]. Our new disulfide bond may provide additional stability to this loop which could in turn increase the stability of the active site.

4.3.1.1 Creating a Reaction Profile for DISU-PETase

The novel disulfide bond introduced via the S163C-S194C mutation was selected such that it has a sufficient distance from the active site of *Is*-PETase so that it would not interfere with the catalytic mechanism. To test this theory, we generated reaction profiles using the same methods as used for the wild-type enzyme. As the introduction of a disulfide bond distal to the active site is extremely unlikely to change the enzyme's mechanism, we proceeded under the assumption that DISU-PETase

has the same mechanism as the wild-type.

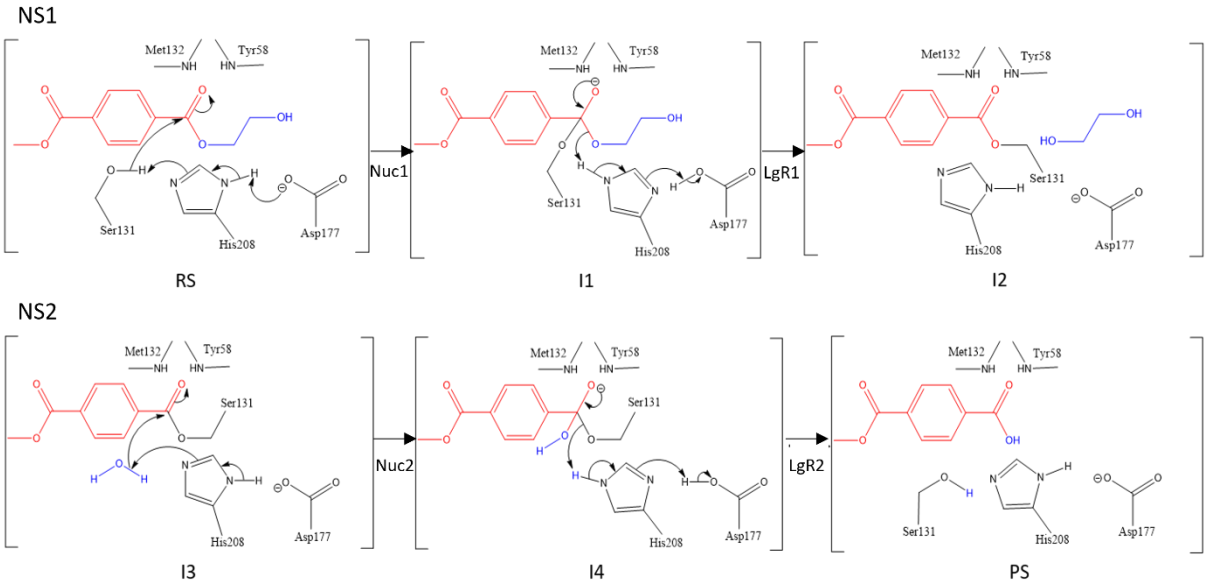


Figure 53: The proposed mechanism for Is-PETase and our mutant enzyme, DISU-PETase. This reaction scheme is identical to the one we provided in chapter 2.

Initially a MD simulation (using methods described in section 4.2.2) in the absence of any ligand was performed. This was done to relax the modified structure. The structure of DISU-PETase after this MD simulation was not significantly different from the wild-type structure. This supports the belief that the introduction of the Cys163-Cys193 disulfide bond has no deleterious effects on protein folding.

Next, the water and ions introduced to the DISU-PETase system were removed using the program Pymol[14]. A molecular docking simulation was then performed using the ligand HEMT [SEE FIGX] into this cleaned structure with the software package AutoDock Vina[68]. The docking pose with the shortest distances between the OC_{LIG} atom and the oxyanion hole nitrogen atoms was selected.

A subsequent MD simulation was then performed with the HEMT ligand present (after re-solvation and counter-ion addition steps were performed). Three snapshots were taken from the trajectory of this MD simulation. These were used as the starting geometries to generate intermediates along the reaction profile of the initial nucleophilic substitution reaction NS1.

The starting geometry for the second nucleophilic substitution NS2 was generated using the following method: Starting with the same binding pose as for NS1, atoms in the ligand that make up the ethylene glycol leaving group were deleted such that the ligand became HMT (with a carboxylic acid functional group instead of an ester bond to ethylene glycol). An MD simulation was then performed with the ligand HMT bound to the active site of *Is*-PETase. Three snapshots were then

taken from the trajectory of this MD simulation. These were used as the starting geometries to generate intermediates along the reaction profile of NS2.

Intermediates along the reaction profiles of NS1 and NS2 were generated using identical methods as applied to the wild-type enzyme (discussed fully in section 3.1). Similarly, NEB calculations were used to generate transition state geometries using the same methods as previously discussed in section 3.2.3.4. Single-point energy calculations were then performed on the geometries of intermediates using methods detailed in section 3.2.3.5.

4.3.1.2 Analysis of the Energy Profiles of DISU-PETase

Of the three starting geometries taken from the MD trajectories, two reaction profiles were successfully generated for NS1 and one for NS2. The remaining reaction profiles were not generated due to poor convergence of our some of our NEB calculations. This precluded the generation of the geometries of transition states along our reaction profiles.

The value of these reaction profiles is that we can compare their energy profiles with the energy profiles previously calculated for the wild-type enzyme. If the activation energy barriers for DISU-PETase are significantly different from the wild-type's, it will support the conclusion that the introduction of the S163C-S194C disulfide bond has had some impact on the chemical mechanism of *Is*-PETase. As the purpose of the introduction of the S163C-S194C disulfide bond was to increase the enzyme's thermostability not to affect the energetics of its chemical mechanism, it is expected that no significant difference will be found between the reaction profiles of DISU-PETase and WT-PETase.

For the reaction profile of NS1, the majority of the single-point energies calculated for DISU-PETase fit within the range of single-point energies calculated for WT-PETase (see Figure 54). Two datapoints are outside the range of energies calculated of WT-PETase: One energy calculated for transition state Nuc1 has an energy of 9.5 kJ mol^{-1} relative to the preceding RS intermediate. This is 12.8 kJ mol^{-1} lower than the lowest energy calculated for a Nuc1 geometry for WT-PETase. The other datapoint out of the range of energies calculated for WT-PETase is for an I2 geometry. The relative energy for this geometry is 32.2 kJ mol^{-1}, which is 6.9 kJ mol^{-1} higher than the highest energy calculated for an I2 geometry for WT-PETase. Given that the most of the single-point energies calculated for DISU-PETase lie within the range of energies calculated for WT-PETase for NS1, and that the outlying datapoints are not outside this range by a large amount, we feel that it is safe to conclude that the introduction of the S163C-S194C disulfide bond has no significant effect on the energetics of the NS1 reaction in *Is*-PETase.

For the reaction profile of NS2, our energy profile for DISU-PETase fits wholly within the range of energies calculated for WT-PETase (see Figure 54). This leads to the conclusion that the introduction of the S163C-S194C disulfide bond has no significant effect on the energetics of the NS2 reaction in *Is*-PETase.

From our reaction profiles calculated for DISU-PETase, we can assign a rate-determining step. Since we have only two reaction profiles for NS1 and one for NS2 with realistic energy barriers, the "lowest energy profile" method for selecting rate-determining steps is appropriate. The highest energy barrier for the lowest energy reaction profile calculated for NS1 is the LgR1 step, with an activation energy of 58.0 kJ mol^{-1}. The highest energy barrier for the lowest energy reaction profile for NS2 is the Nuc2 step, with an activation energy of 62.2 kJ mol^{-1}. This makes the overall rate-determining step the Nuc2 step. This activation energy is roughly twice as high as the activation energy calculated for WT-PETase (35.6 kJ mol^{-1}), although the rate-determining step is the same. As we have only generated two full reaction profiles for DISU-PETase, further sampling by generating more may yield reaction profiles with lower energy barriers. This however was not performed as generation of reaction profiles using our QM/MM techniques is extremely time consuming and we can already make relevant conclusions using the reaction profiles that have already been generated.

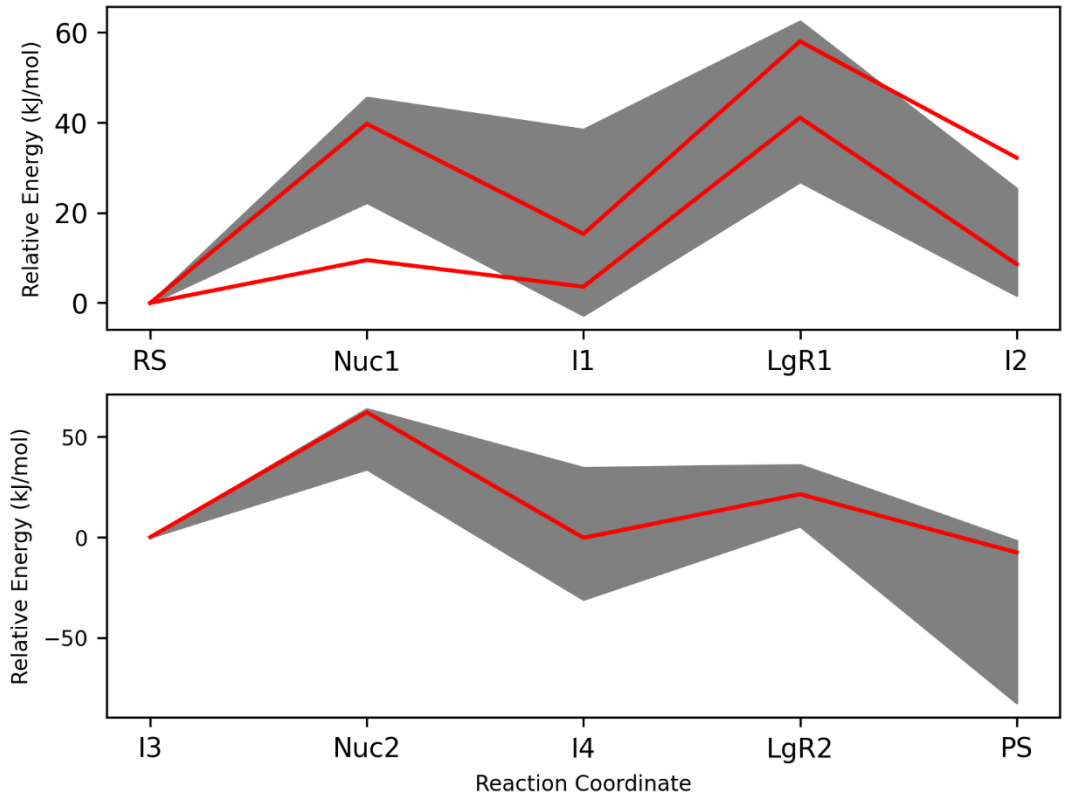


Figure 54: Energy profiles for the degradation of HEMT via our mutant enzyme DISU-PETase are shown as red lines. The range of our previously calculated energy values for WT-PETase is shown as grey fill. The reaction profiles for NS1 (top) and NS2 (bottom). Note that for clarity, different y-axis limits have been used in each plot.

4.3.2 Can Molecular Dynamics Simulations Reveal Routes to Thermostability for Is-PETase Variants?

As discussed in section 1.1.3, thermostability is thought to be the key feature to optimise in PETases as it allows the enzyme to operate at temperatures above the glass transition temperature of the substrate PET. There have been several studies that report *Is*-PETase variants with improved thermostability. At the time of writing the two variants with the highest reported thermostability are FAST-PETase[37] and HOT-PETase[39]. The single-point mutations in FAST-PETase have been predicted by machine learning methods, while the majority of single-point mutations in HOT-PETase have been created through directed evolution. Both of machine learning and directed evolution can be considered 'black box' methods, i.e., no understanding of the protein's structure is required to obtain mutant proteins with improved thermostability. This is beneficial as more thermostable variants can be created within relatively short timeframes. The downside of 'black box' methods is that they provide no explanation for their successes. Thus, further analysis must be performed upon their resulting protein structures in order to understand the factors that contribute to these

mutants' improved thermostability. In the study discussed in section 4.1.2.3, molecular dynamics simulations have been used as a method to probe enzyme thermostability[127]. In the following sections we apply similar techniques to probe the thermostabilities of various mutants of *Is*-PETase.

We performed a series of MD simulations upon FAST-PETase, HOT-PETase and WT-PETase. Geometries for FAST-PETase[37] and HOT-PETase[39] were obtained from crystallographic structures obtained by their original creators. For WT-PETase, the PDB structure from Austin *et al.* (PDB id: 6eqe[18]) was used instead of the PDB structure used in our QM/MM studies of WT-PETase (PDB id: 5xh3[36]). This was because the structure in PDB 6eqe contains additional residues at the C-terminus that are present in the crystal structure of HOT-PETase. In the crystal structure of FAST-PETase, these C-terminal residues have not been resolved, perhaps due to their high flexibility.

For each of these three enzymes, ten 50 ns MD simulations were performed at 300K and at 360K. These temperatures were chosen to provide a point of comparison between the enzymes at low and high temperatures. Multiple simulations were performed per temperature and enzyme as it became clear through preliminary testing that individual MD trajectories produced noisy data. It is our expectation that averaging over several trajectories will result in data that is more interpretable.

4.3.2.1 RMSF analysis of MD simulations

From each of the MD simulations performed, the root mean squared fluctuation (RMSF) for each residue's α-carbon (referred to from now on as ρ_{CA}) was extracted from its trajectory. RMSF is a measure of deviation from the mean:

$$\rho_i = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle}.$$

Equation 86: The equation for calculating RMSF.

In Equation 86, the ρ_i represents the RMSF of each group (in our case single α-carbon atoms). The term x_i denotes the Cartesian coordinates of particle i (including x, y and z values), while the notation ⟨ ⟩ indicates the arithmeticmean of values inside.

Residues that are highly mobile throughout each simulation will have high associated ρ_{CA} values. This data provides us information on which regions of each enzyme have high flexibility. Average ρ_{CA} values were calculated across 10 repeats. This should reduce the noise of the data and reduce the impact of individual MD simulations with outlying behaviour on our ρ_{CA} values. It was decided to perform this analysis on the α-carbons of each residue rather than the entire residue as this removes the effect of side chains upon the RMSF value. This allows our ρ_{CA} values to tell us more about large-scale motions of the protein backbone rather than rotations of residues' sidechains.

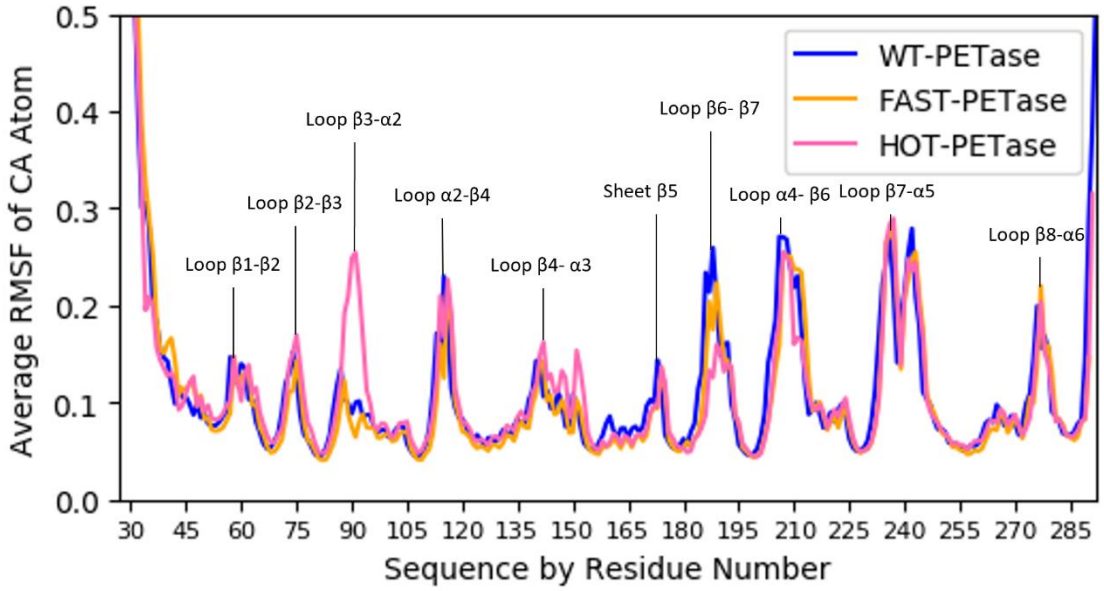


Figure 55: Average RMSF values per residue for the enzymes WT-PETase (blue), FAST-PETase (orange) and HOT-PETase (pink). Residue numbers have been shifted to match the residue numbers in the PDB structure 6eqe[18]. This data has been gathered from simulations performed at 360 K, with 10 repeats of 50 ns MD simulations. Regions with high flexibility have been labelled by secondary structure.

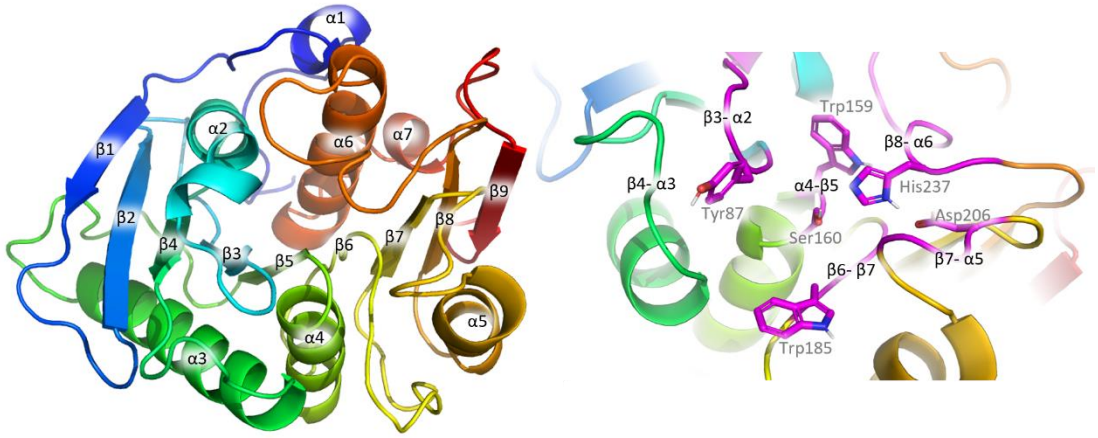


Figure 56: The structure of Is-PETase with labelled α-helices and β-sheets (left). Loops in the active site of Is-PETase with residue side-chains that are important to the enzyme's activity (right).

From our averaged RMSF data we can see a series of 'spikes' that represent regions in each *Is*-PETase variant with high flexibility throughout our simulations (see Figure 55). Visual inspection reveals that these 'spikes' in flexibility occur in the same regions for all three enzymes. This is to be expected as they have very high sequence similarity (FAST-PETase and HOT-PETase are, after all mutants of WT-PETase). The majority of high flexibility regions in our simulations are loop regions. This is to be expected as α-helices and β-sheets are known to have ordered patterns of interactions that contribute to their rigidity.

By comparing the differences in the RMSF values for with WT-PETase, the effects of the mutations in FAST-PETase and HOT-PETase upon the enzyme's flexibility can be revealed. By analysing the protein environment proximal to these key mutant residues, we identified potential interactions that may provide additional stabilisation to the enzyme in question. Once we had identified interactions of interest, we further investigated the presence of these interactions using a combination of distance measurements (see section 4.3.2.2) and visual inspection of key snapshots of our MD simulations (see section 4.3.2.3).

For the enzyme FAST-PETase, the RMSF data overlays almost exactly with that of WT-PETase. A slight reduction in RMSF relative to WT-PETase can be observed for FAST-PETase at active site residue Trp185, located on the β6- β7 loop. Located on the β6- β7 loop is the mutant residue D186H. In the original paper[37], it was predicted that this new histidine residue formed a WMHB with a second mutant residue S121E (located on the end of the α3-helix). This new glutamate residue was predicted to form a further hydrogen bond with the residue Asn172 (located on the end of the α4-helix). If this proposed hydrogen-bonding network is in fact present in FAST-PETase, it could be

responsible for the slight reduction in the flexibility of the active site loop β6- β7. Discussion of our further investigations of these interactions in FAST-PETase can be found in sections 4.3.2.2.1.2 and 4.3.2.3.1.

Additionally, the mutant residue R224Q was predicted to form a new hydrogen bond with the backbone of residue Ser192. This mutation however replaces the hydrogen bond between Arg224 (located on the β6- β7 loop) and Ser193 in WT-PETase. In theory, the replacement of one hydrogen bond with another should not provide any significant stabilisation. If the new R224Q-Ser192 hydrogen bond forms a stronger or longer-lived hydrogen bond however, some stabilisation may have been achieved. Discussion of our further investigations of these interactions in FAST-PETase can be found in sections 4.3.2.2.1.1 and 4.3.2.3.1.

Another slight reduction in RMSF around the catalytic Ser160 residue, located on the α4-β5 loop can be observed for the FAST-PETase variant. As discussed above, the D185H-water-S121E-Asn172 hydrogen bonding network predicted to be present in FAST-PETase may be providing stabilisation to the β6- β7 loop. The residue Asn172 is on the opposite end of the α4-helix as the α4-β5 loop containing Ser160. This novel hydrogen-bonding network may be providing a stabilising effect to the α4-β5 loop as well.

In the original paper for FAST-PETase[37], it was predicted that a salt-bridge interaction between the mutant residue N233K (located on the active site β7-α5 loop) and the residue Glu204 (located on the active site α4-β6 loop) would form. In theory, this novel hydrogen bond between two active site loops should provide these loops some stabilisation. Our RMSF data shows no clear change in flexibility for either loop. Discussion of our further investigations of these interactions in FAST-PETase can be found in sections 4.3.2.2.1.3 and 4.3.2.3.1.

The RMSF data for HOT-PETase shows more dramatic differences than FAST-PETase when compared with WT-PETase. This is unsurprising as HOT-PETase has 21 single-point mutations relative to WT-PETase, while FAST-PETase only has five. The most remarkable feature in our RMSF data for HOT-PETase is the large increase in flexibility in the β3-α2 loop. Intuitively, one might assume that an increase in flexibility would be associated with the reduction of the enzyme's thermostability. Enzyme thermostability is however a far more complex and multifaceted problem. This loop contains the active site residue Trp87. Proximal to the β3-α2 loop are the mutant residues R90T and K95N. This could lead to the conclusion that these two mutations contribute to the flexibility of this region, which in turn may be important to the activity of the enzyme. This data does not however provide any mechanistic information what interactions are responsible for this effect or whether

other, less proximal mutations are responsible for this increase in flexibility. Discussion of our further investigations of these interactions in FAST-PETase can be found in section 4.3.2.2.2.1.

A dramatic reduction in flexibility relative to WT-PETase can be seen in HOT-PETase's β6-β7 loop. This loop contains the active site residue Trp185. It was noted in the original paper[39] that the S214Y mutation creates a new π-stacking interaction between S214Y and Trp185. This new interaction could be responsible for the reduction in flexibility in this region. Additionally, the residue adjacent to Trp185 has the D186H mutation in HOT-PETase. As discussed previously, WMHBs have been observed between D186H and S121E in the crystal structure of FAST-PETase[37], as well as a WMHB between S121E and Asn172. As the S121E mutation is also present in HOT-PETase, it is plausible that the same network of WMHBs is present. Discussion of our further investigations of these interactions in FAST-PETase can be found in section 4.3.2.2.2.2.

A less dramatic, but still noticeable difference between our WT-PETase and HOT-PETase RMSF data can be seen at the β4-α3 loop, where a slight increase in flexibility can be observed for HOT-PETase. This loop does not contain any active site residues but is located directly 'behind' the β3-α2 loop which contains the active site residue Trp87. A change in flexibility for the β4-α3 loop could therefore affect the flexibility of the β3-α2 loop, which could be important to the catalytic activity of the enzyme. In HOT-PETase, the mutation Q119K is located on the β4-α3 loop. We will later investigate whether any novel hydrogen bonding interactions or salt-bridges are introduced through this mutation. Discussion of our further investigations of these interactions in HOT-PETase can be found in section 4.3.2.2.2.3.

Analysis of our RMSF data has not provided any mechanistic information on how each mutation affects the stability of its enzyme. It has however, highlighted the regions of each enzyme that merit further, more detailed, investigation.

4.3.2.2 Distance analysis of MD simulations

Our RMSF analysis has revealed new interactions in the mutant enzymes FAST-PETase and HOT-PETase that may provide stabilisation to these enzymes relative to WT-PETase. While many of these interactions have been observed in X-ray crystal structures, X-ray crystallography provides only a static image of an enzyme and provides little information on the dynamics of the system. By measuring the distances between groups throughout our MD simulations, we can build a more dynamic picture of these stabilising interactions. By comparing the behaviour of these interactions under different temperatures, we may be able to comment on their thermal stability.

We will discuss each distance measurement as a distribution over the span of ten 50 ns MD simulations, performed at both 300K and 360K. In general, we will assess the strength of each interaction as a function of its distance, where the shorter the distance, the stronger the interaction.

It is worth noting that this analysis does not contain any angular component, relying solely upon the Euclidean distance between two predefined atoms. It may appear that this means that we cannot therefore identify highly directional interactions such as hydrogen bonds with this method. However, we have selected atom pairs to measure between that are suitably oriented such that hydrogen bonding interactions could form. Thus, it is reasonable to assume that when these atoms are close enough to form a hydrogen bond (i.e., less than 3.5 Å) a hydrogen bond is likely to be present.

4.3.2.2.1 Distance Analysis for FAST-PETase throughout MD trajectories

Our RMSF analysis identified three clusters of residues where mutations in FAST-PETase relative to WT-PETase seem to have had a small effect upon the flexibilities of their surrounding loops (see Figure 57). These clusters correspond to hydrogen bonding networks that were proposed by the creators of FAST-PETase[37] to be present in FAST-PETase and absent in WT-PETase, resulting in the increase in FAST-PETase's thermostability.

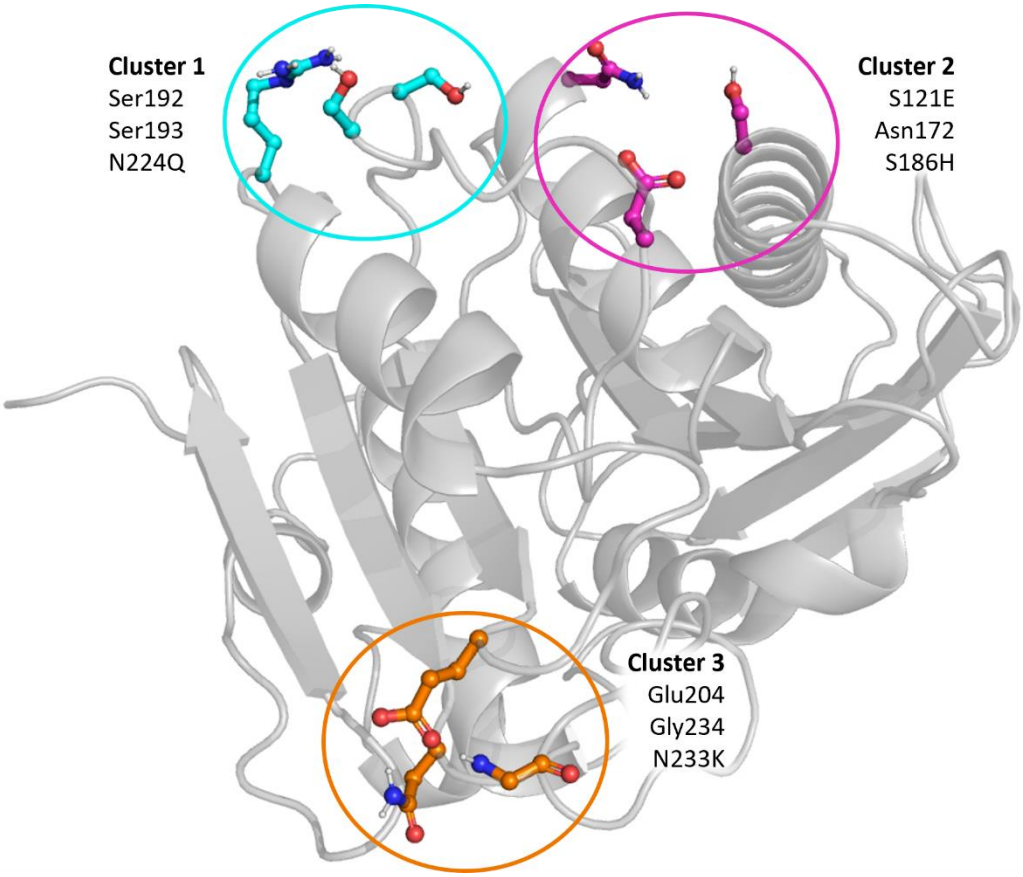


Figure 57: Locations of residue clusters with interactions thought to be important to the increased thermostability of FAST-PETase relative to WT-PETase. The structure shown is that of WE-PETase, residues with three-letter identifiers are present in both WT-PETase and FAST-PETase while residues with one-letter identifiers show mutations from WT-PETase to FAST-PETase (e.g., N224Q is Asn224 in WT-PETase and Gln224 in FAST-PETase).

4.3.2.2.1.1 FAST-PETase Cluster 1

In FAST-PETase, Cluster 1 contains the mutant residue N224Q and the wild-type residues Ser192 and Ser193. It was proposed in the original paper for FAST-PETase[37] that the mutant residue N224Q forms a new hydrogen bond with Ser192. This interaction replaces a hydrogen bond between Asn224 and Ser192 in WT-PETase. In X-ray crystal structures, both of these hydrogen bonds are observed[37]. The distances between the sidechain oxygens of Ser192 and Ser193 and the sidechain nitrogen atoms of Asn224 (in WT-PETase) and N224Q (in FAST-PETase) were measured throughout our MD simulations.

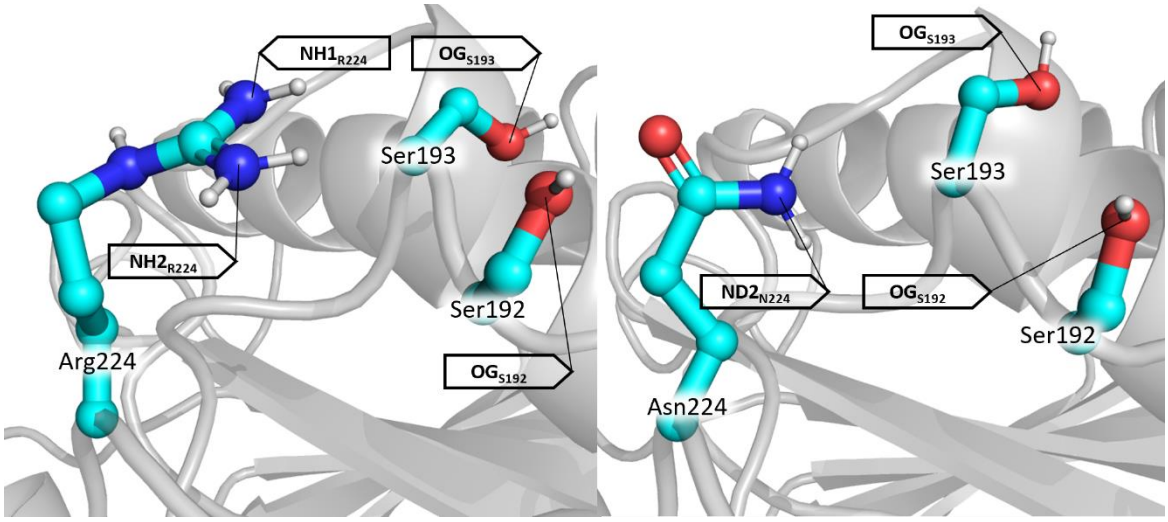


Figure 58: Residues in cluster 1 for WT-PETase (left) and FAST-PETase (right). Key atoms have been labelled.

Our distance measurements show that for the majority of timesteps within our simulations (at both 300K and 360K), the distance between the sidechain nitrogen of N224Q and the sidechain oxygen of Ser192 (see Figure 59) and Ser193 (see Figure 60) remains above ca. 5 Å. This indicates that if a hydrogen bond is formed between N224Q and either of these serine residues, it is very weak. This result is not commensurate with the observations that can be made via X-ray crystallography. One interpretation of this is that these hydrogen bonds can only form under the specific conditions required to obtain protein crystals, whilst under normal conditions the motion of the sidechains involved results in loss of these stabilising hydrogen bonds.

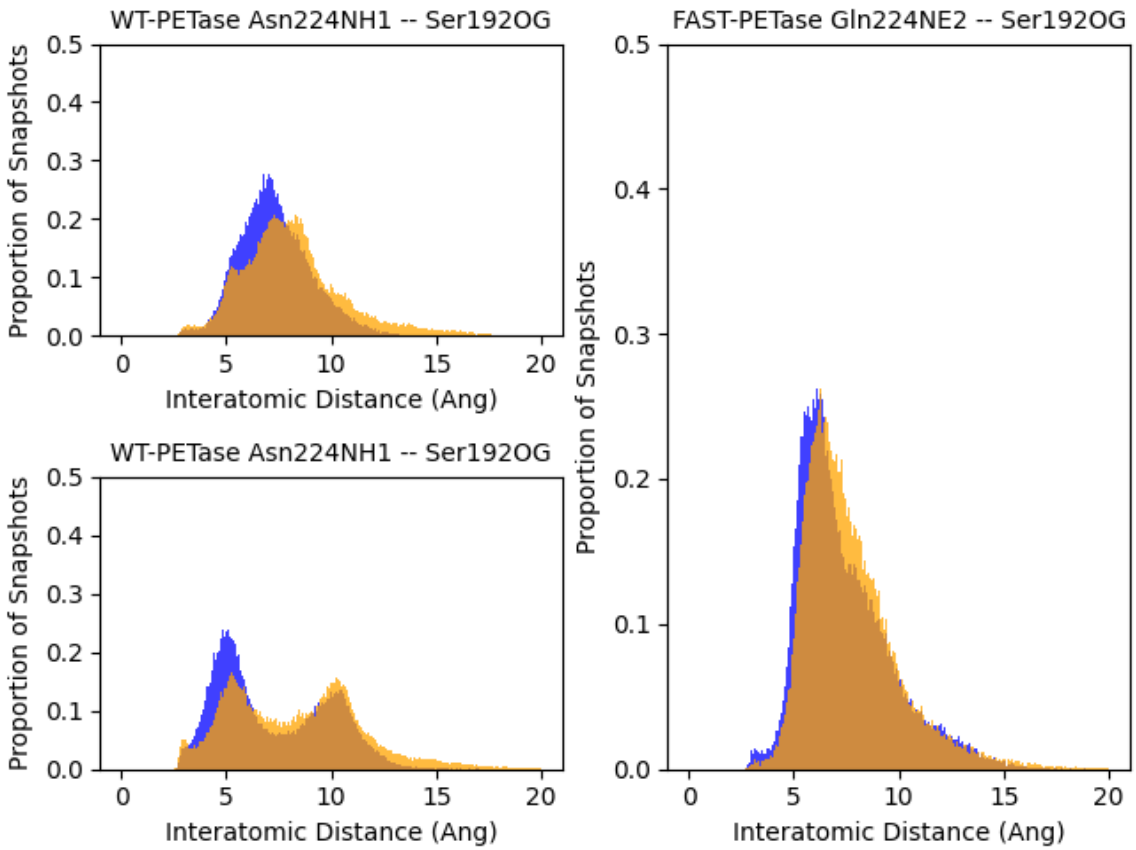


Figure 59: Distance distributions for interactions between Asn224 and Ser192 in WT-PETase (top left, top right) and N224Q and Ser192 in FAST-PETase (right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

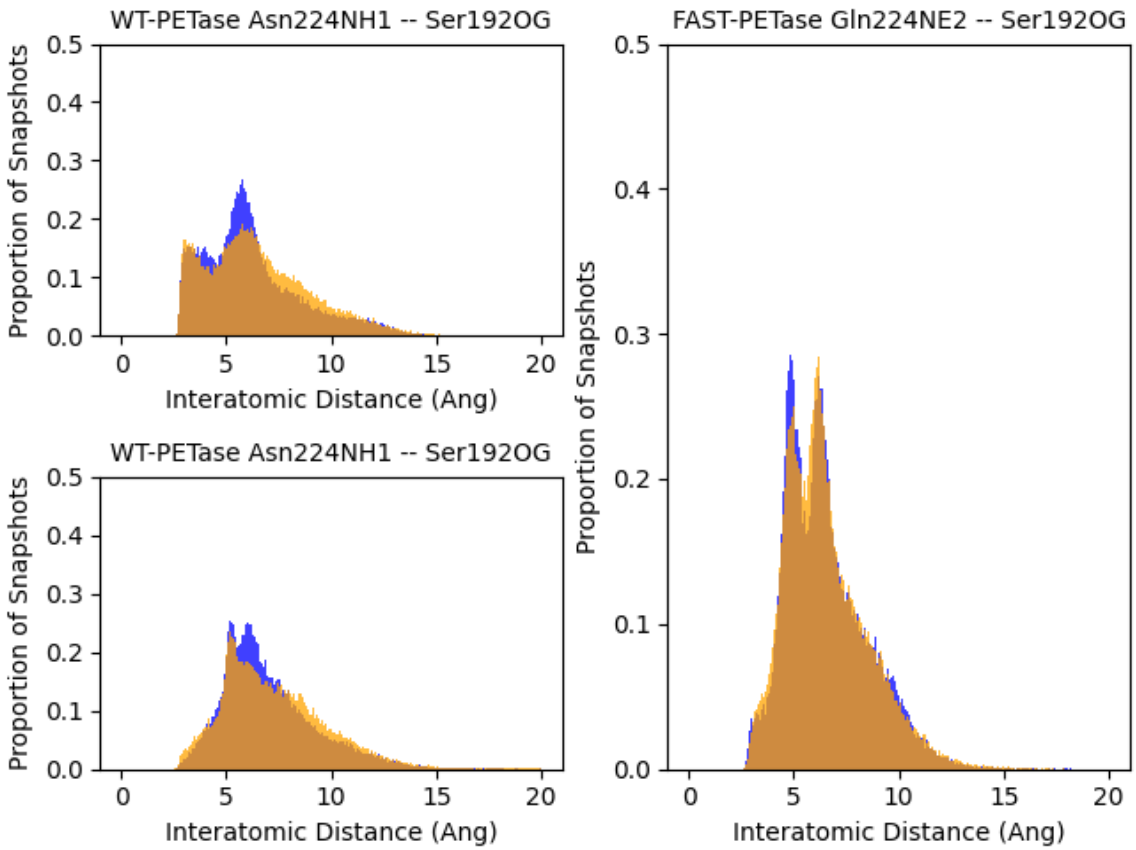


Figure 60:: Distance distributions for interactions between Asn224 and Ser193 in WT-PETase (top left, top right) and N224Q and Ser193 in FAST-PETase (right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

4.3.2.2.1.2 FAST-PETase Cluster 2

In FAST-PETase, Cluster 2 contains the mutant residues D186H and S121E, as well as the natural residue Asn172. In a structure of FAST-PETase obtained via X-ray crystallography[37] ordered water molecules were observed in between His186 and Glu121 and in between Glu121 and Asn172. This indicates that WMHBs can form between these residues. From our set of MD simulations of the enzyme FAST-PETase, the distances $NE2_{H186}$-$OE1_{S121E}$ and $NE2_{H186}$-$OE1_{S121E}$ were measured (see Figure 61). For comparison, we also measured the distances $OD1_{D186}$-OG_{S121} and $OD2_{D186}$-OG_{S121} in our simulations of WT-PETase.

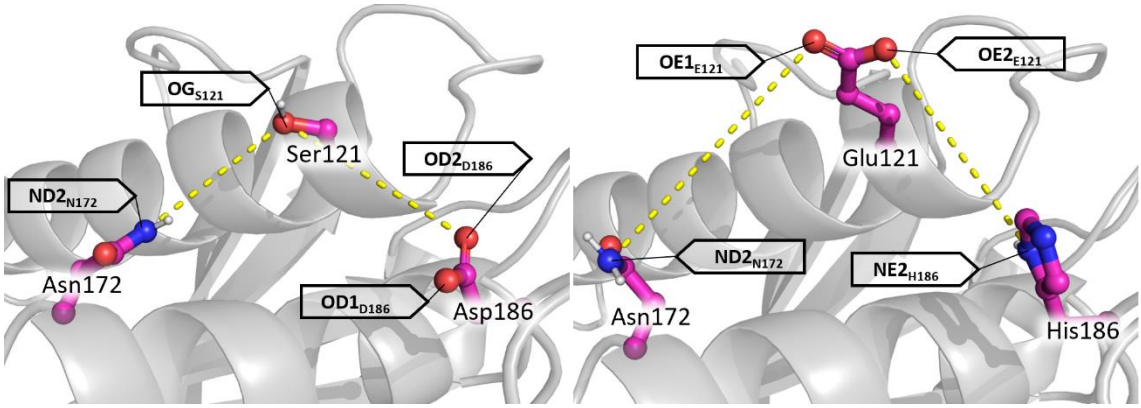


Figure 61: Residues in cluster 2 for WT-PETase (left) and FAST-PETase (right). Key atoms have been labelled.

The carboxylate atoms OE1$_{S121E}$ and OE2$_{S121E}$ and the atoms OD1$_{D186}$ and OD2$_{D186}$ are chemically equivalent. This resulted in no significant difference between distance distributions for distances involving equivalent atoms; as such we will now discuss the interactions as interactions between residues rather than atoms (e.g., when we refer to a distance involving the residue S121E, we will be discussing distances of both of its carboxyl oxygens at the same time). At 300K, the Asp186-Ser121 interaction in WT-PETase appears to be stronger than the equivalent D186H-S121E interaction in FAST-PETase. This can be observed by the tighter distribution of distances around 7.5 Å, which is an appropriate distance for the formation of a WHMB (see Figure 62). While the same distance distribution around 7.5 Å can be observed for the His186-S121E interaction in FAST-PETase, it is broader suggesting a weaker overall interaction. At 360K, both of these interactions become significantly weaker. This can be observed by the broadening of their distance distributions and an increase in the median distance. Overall, our data suggests that a WMHB can be formed between Asp186-Ser121 in WT-PETase and between His186-S121E in FAST-PETase, but this WMHB is not particularly stable with respect to temperature. It is also worth noting that as the sidechain of S121E is negatively charged, its interactions with D186H and Asn172 are likely to be stronger than those of Ser121. Additionally, the mutation D186H can be considered to be synergistic to S121E as it ensures that two negatively charged residues are not proximal to each other.

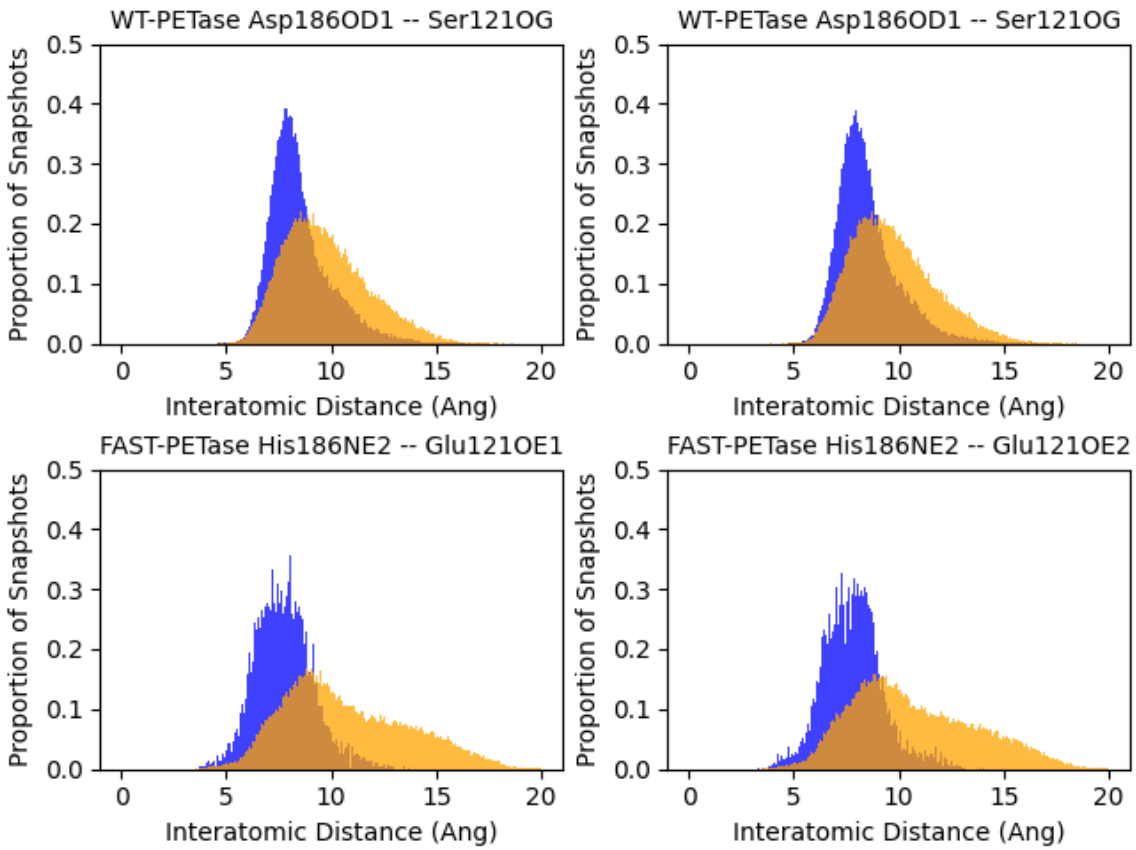


Figure 62: Distance distributions for interactions between Asp186 and Ser121 in WT-PETase (top left, top right) and His186 and Glu121 in FAST-PETase (bottom left, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

The interaction between S121E and Asn172 in FAST-PETase appears to be stronger than the interaction between Ser121 and Asn172 in WT-PETase. At 300K, tighter distributions for the distances $OE1_{S121E}$-$ND2_{N172}$ and $OE1_{S121E}$-$ND2_{N172}$ in FAST-PETase can be observed than for the OG_{S121}-$ND2_{N172}$ distance (see Figure 63). Both distributions have a median distance of ca. 7.5 Å, suggesting that a WMHB can form. It is also worth noting that there are two potential hydrogen bond donors in Glu121 in FAST-PETase compared with one in Ser121 in WT-PETase. This means that when the distance $OE1_{S121E}$-$ND2_{N172}$ is unsuitable for formation of a WMHB, the distance $OE2_{S121E}$-$ND2_{N172}$ is likely to be suitable for one. At 360K, the distribution for the distance OG_{S121}-$ND2_{N172}$ in WT-PETase becomes bimodal, with an additional peak at ca. 10 Å. This distance is too long to facilitate formation of a WMHB, This suggests that the WMHB formed between OG_{S121}-$ND2_{N172}$ in $OE1_{S121E}$-$ND2_{N172}$ and $OE1_{S121E}$-$ND2_{N172}$ in WT-PETase is unstable with respect to temperature. By contrast, the distributions for distances $OE1_{S121E}$-$ND2_{N172}$ and $OE1_{S121E}$-$ND2_{N172}$ in FAST-PETase do not change dramatically from our 300K simulations to our 360K simulations. This suggests that the WMHB between S121E and Asn172 is more thermostable than the WMHB between Ser121 and Asn172 in WT-PETase. Our data

suggests that the stronger, more thermostable bond between S121E and Asn172 may be important for the overall improvement of the enzyme's thermostability.

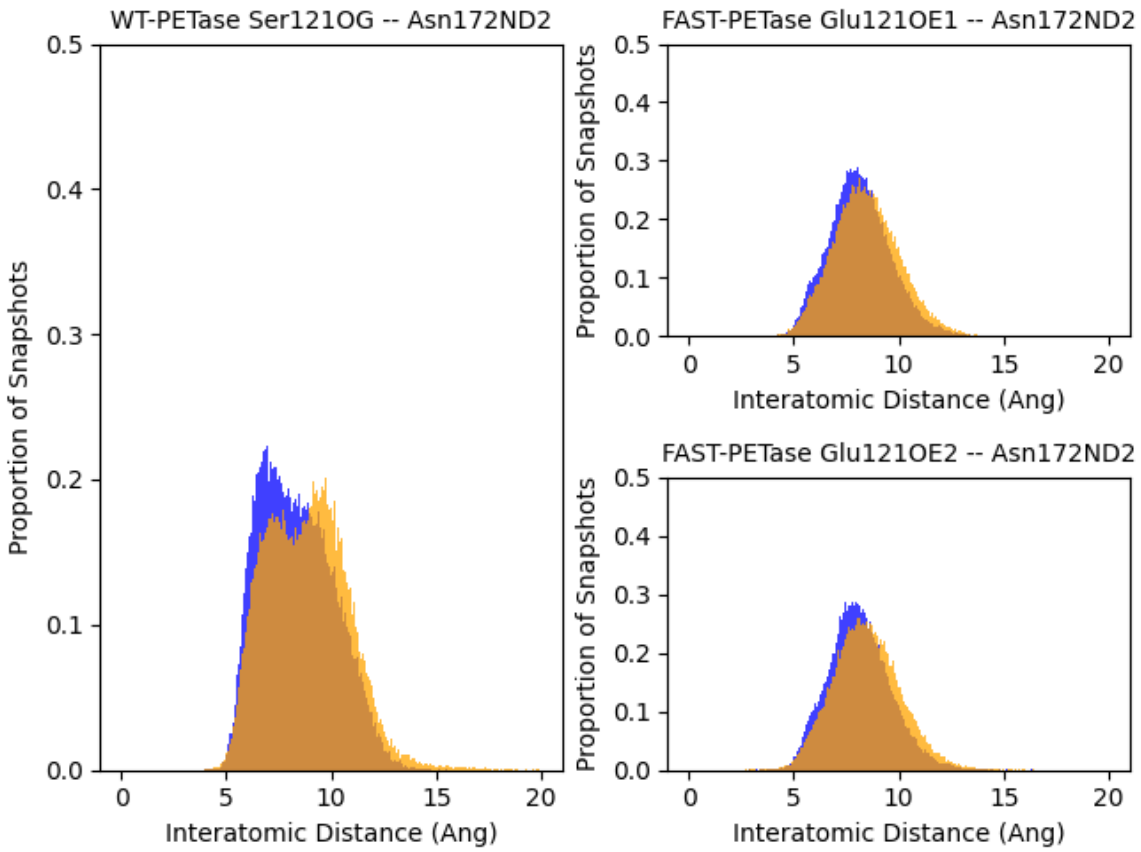


Figure 63: Distance distributions for interactions between Ser121 and Asn172 in WT-PETase (left) and Glu121 and Asn172 in FAST-PETase (bottom left, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

4.3.2.2.1.3 FAST-PETase Cluster 3

In FAST-PETase, Cluster 3 contains the mutant residue N223K as well as wild-type residues Glu204 and Gly234. In the original paper for FAST-PETase[37], X-ray crystallography revealed a hydrogen bond between the carboxyl sidechain of Glu204 and the backbone nitrogen atom of Gly234 in both WT-PETase and FAST-PETase. In the crystal structure of FAST-PETase, a novel salt-bridge interaction was observed between the positively charged sidechain of N233K and the carboxyl side chain of Glu204. In the crystal structure of WT-PETase, no attractive interaction between Asn233 and Glu204 was observed.

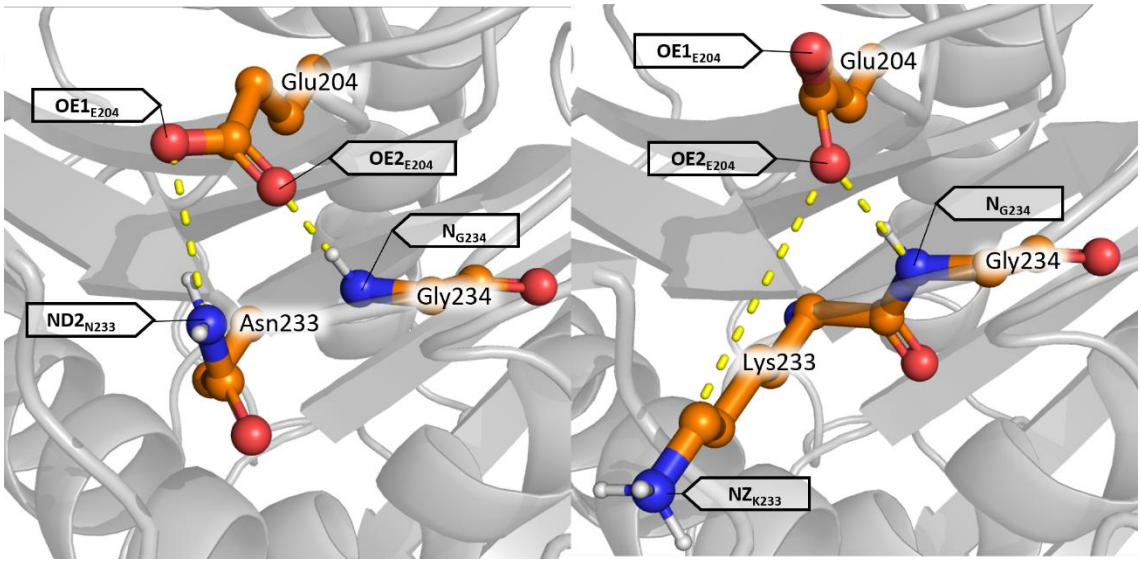


Figure 64: Residues in cluster 3 for WT-PETase (left) and FAST-PETase (right). Key atoms have been labelled.

Distance measurements through our MD simulations reveal that in both WT-PETase and FAST-PETase, a hydrogen bond between the carboxyl side chain of Glu204 and the backbone nitrogen of Gly234 is present for a large proportion of the time (see Figure 65). This is shown by the peak in the distance distribution around 3 Å. The distance distributions for these interactions do not significantly change when the simulation temperature is increased from 300K to 360K. This suggests that these interactions are highly stable. An additional peak at ca. 5 Å is observed in these distance distributions. This is most likely due to the carboxyl oxygen on Glu204 that is not forming a hydrogen bond with the backbone nitrogen of Gly234. The distinct peaks at ca. 3 Å and 5 Å suggest that the sidechain of Glu204 does not rotate freely. This further supports the theory that a strong interaction between Glu204 and Gly234 is present in both WT-PETase and FAST-PETase.

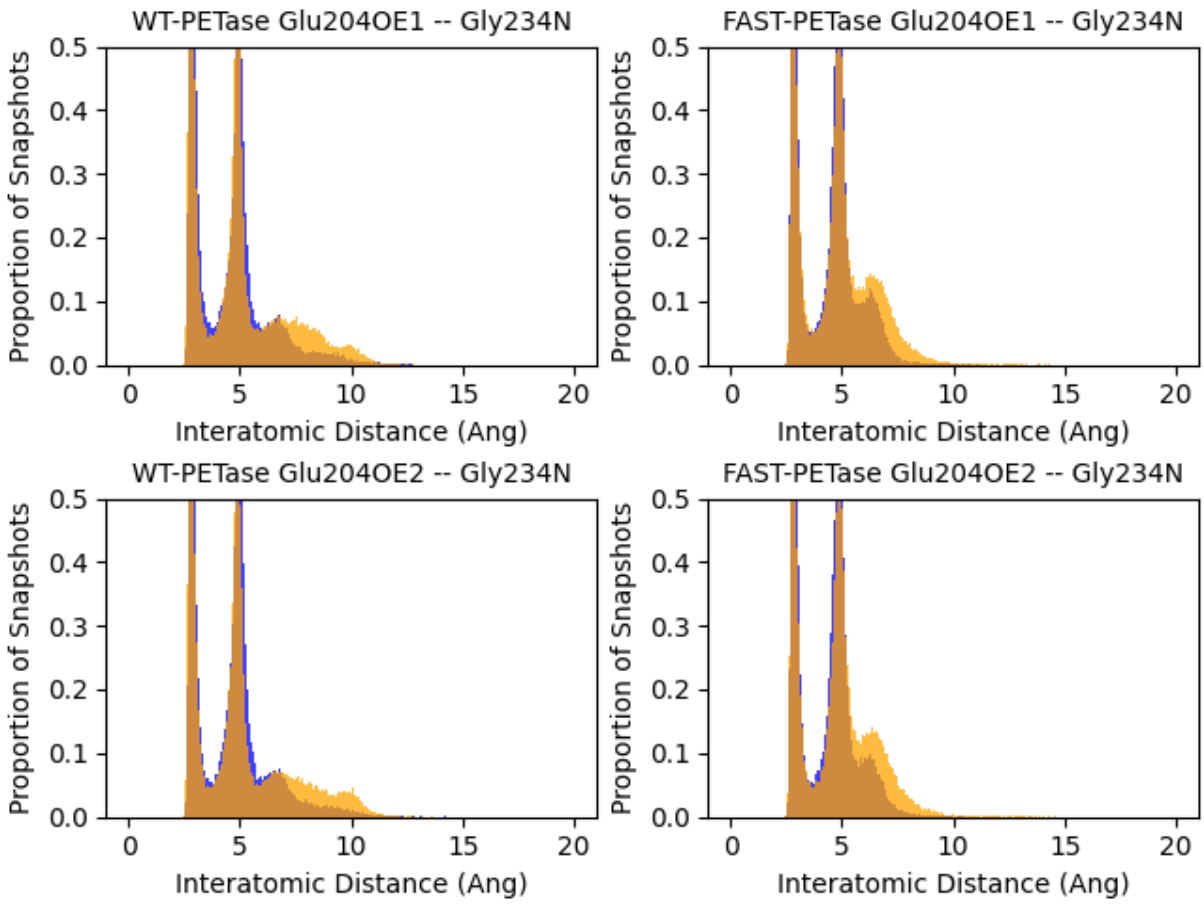


Figure 65: Distance distributions for interactions between Glu204 and Gly234 in WT-PETase (left top, left bottom) and in FAST-PETase (top right, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

Our distance measurements reveal a strong hydrogen bonding interaction between Glu204 and the sidechain nitrogen of Asn233 in WT-PETase. This can be seen through the peak in the distance distribution at ca. 3 Å. Similarly, a peak at ca. 3 Å can be observed in the distance distributions between the carboxyl oxygen of Glu204 and the sidechain nitrogen of N233K in FAST-PETase (see Figure 66). Using these distance distributions alone, we would conclude that no stability enhancement has been made through the N233K mutation. We must however take into account the positive charge carried by the sidechain of N223K. The interaction between N223K and the negatively charged Glu204 can therefore be classified as a salt-bridge rather than a hydrogen bond. This makes it likely that a stronger attractive interaction between N223K and Glu204 is present in FAST-PETase, when compared to the Asn233 – Glu204 interaction in WT-PETase.

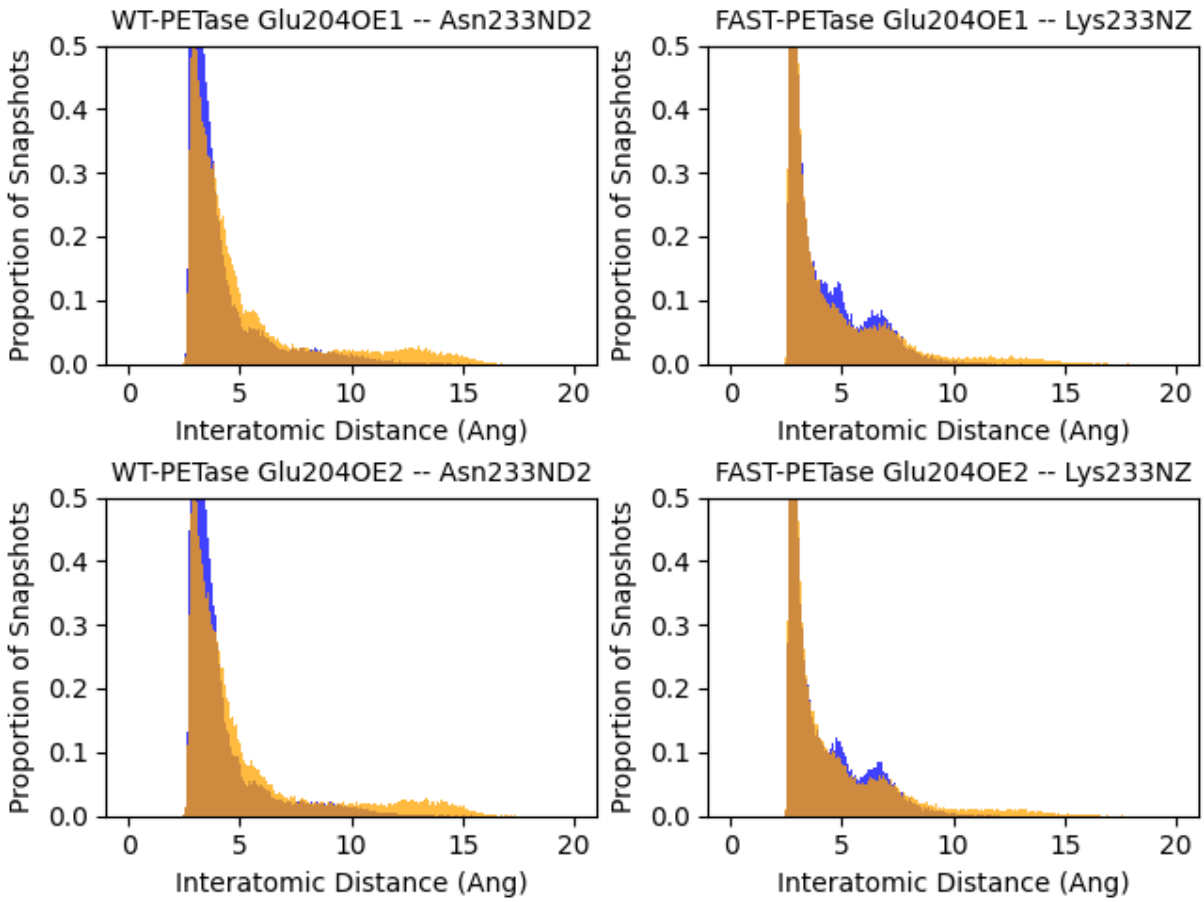


Figure 66: Distance distributions for interactions between Glu204 and Asn233 in WT-PETase (left top, left bottom) and between Glu204 and N233K in FAST-PETase (top right, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

4.3.2.2.2 Distance analysis for HOT-PETase throughout MD trajectories

Our RMSF analysis has identified three key regions in HOT-PETase where its mutations have had a significant impact upon the flexibility of the enzyme compared to WT-PETase. To organise this discussion, we have separated our analysis into three separate clusters. The first of these clusters contains the mutant residues K90T and K95N (see Figure 67). We will use distance measurements to assess whether either of these mutant residues form any novel hydrogen bonding interactions with surrounding residues. The second cluster contains the mutant residues S121E and D187H (see Figure 67); these mutations are also present in FAST-PETase. As discussed previously in section 4.3.2.2.1.2, in FAST-PETase these residues are involved in WMHB interactions with each other and with the residue Asn172. We will use distance measurements to assess whether WMHB formation is plausible within this cluster of residues in HOT-PETase. Additionally, we will use distance measurements to assess whether the mutant residue Q119K (see Figure 67) forms any novel interactions with its surrounding residues.

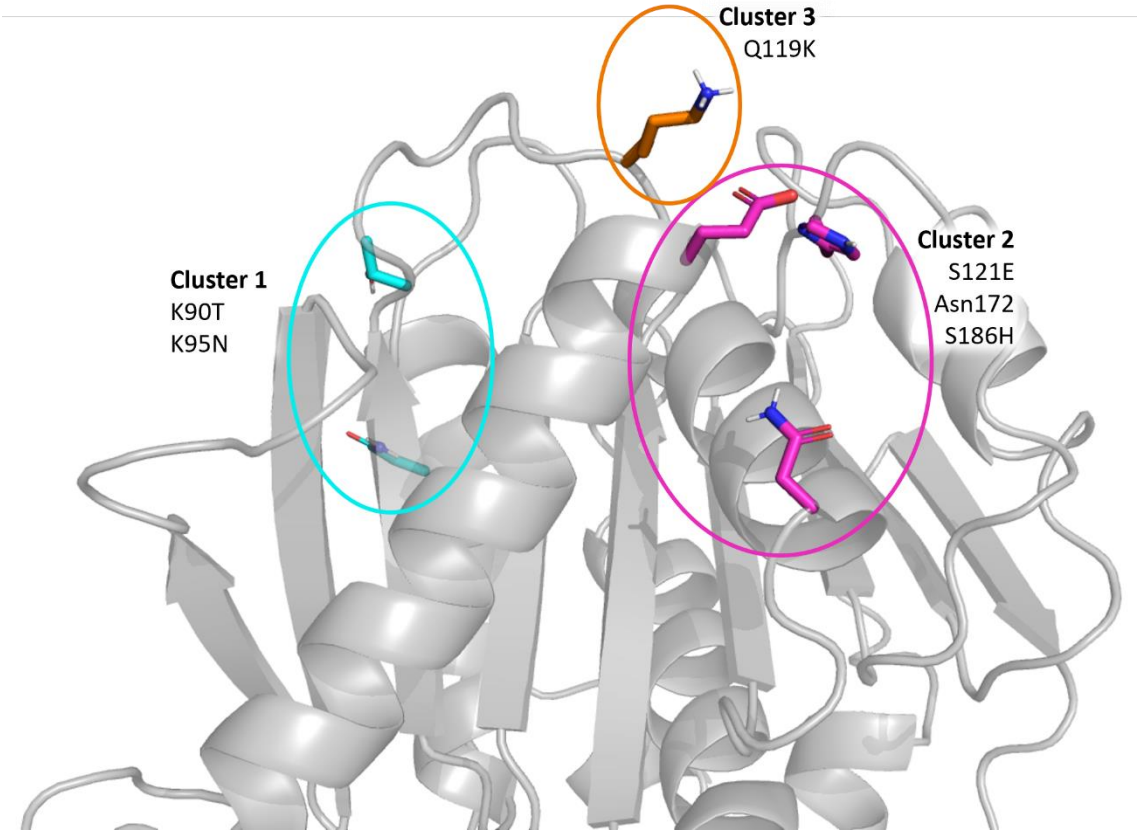


Figure 67: Locations of residue clusters with interactions thought to be important to the increased thermostability of HOT-PETase relative to WT-PETase. The structure show is that of WT-PETase, residues with three-letter identifiers are present in both WT-PETase and HOT-PETase while residues with one-letter identifiers show mutations from WT-PETase to HOT-PETase).

4.3.2.2.2.1 HOT-PETase Cluster 1

Cluster 1 in HOT-PETase contains the mutant residues R90T and K95N. The sidechains of both of these resides have potential to form hydrogen bonds. In both of the single-point mutations we are considering in this cluster, a positively charged residue has been replaced with a neutral residue with a hydrogen bond donor. Visual inspection of the crystal structure of WT-PETase[69] revealed no negatively charged sidechains proximal to Arg90 or Lys95. This means that the wild-type residues could not have formed highly stabilising salt-bridges in WT-PETase; as such their replacement is not obviously deleterious to the thermostability of the enzyme. Visual inspection of the crystal structure of HOT-PETase[39] revealed that the sidechain hydroxyl group of R90T could potentially form hydrogen bonds with the sidechain hydroxyl group of Ser92 (see Figure 68). Relevant measurements were made for our MD simulations performed on HOT-PETase. For comparison, equivalent measurements in WT-PETase have also been made.

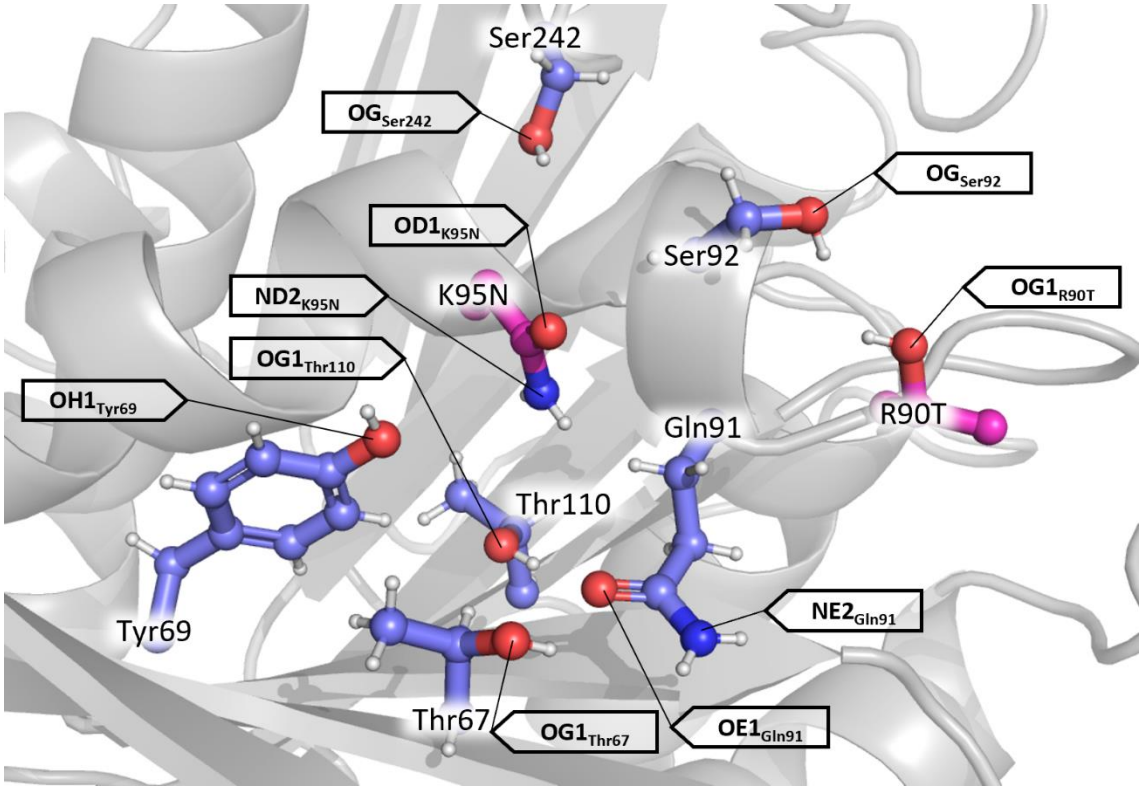


Figure 68: The mutant residues R90T and K95N (shown in magenta). Surrounding residues with hydrogen bond donors/acceptors that had potential to form hydrogen bonds with R90T and K95N (shown in blue).

Our measurements reveal that a new hydrogen bond can form between the sidechain of R90T and the sidechain of the adjacent residue Ser92 in HOT-PETase. In WT-PETase, a weaker interaction (i.e. a longer ranged one) was observed between the sidechain of Lys90 and Ser92.

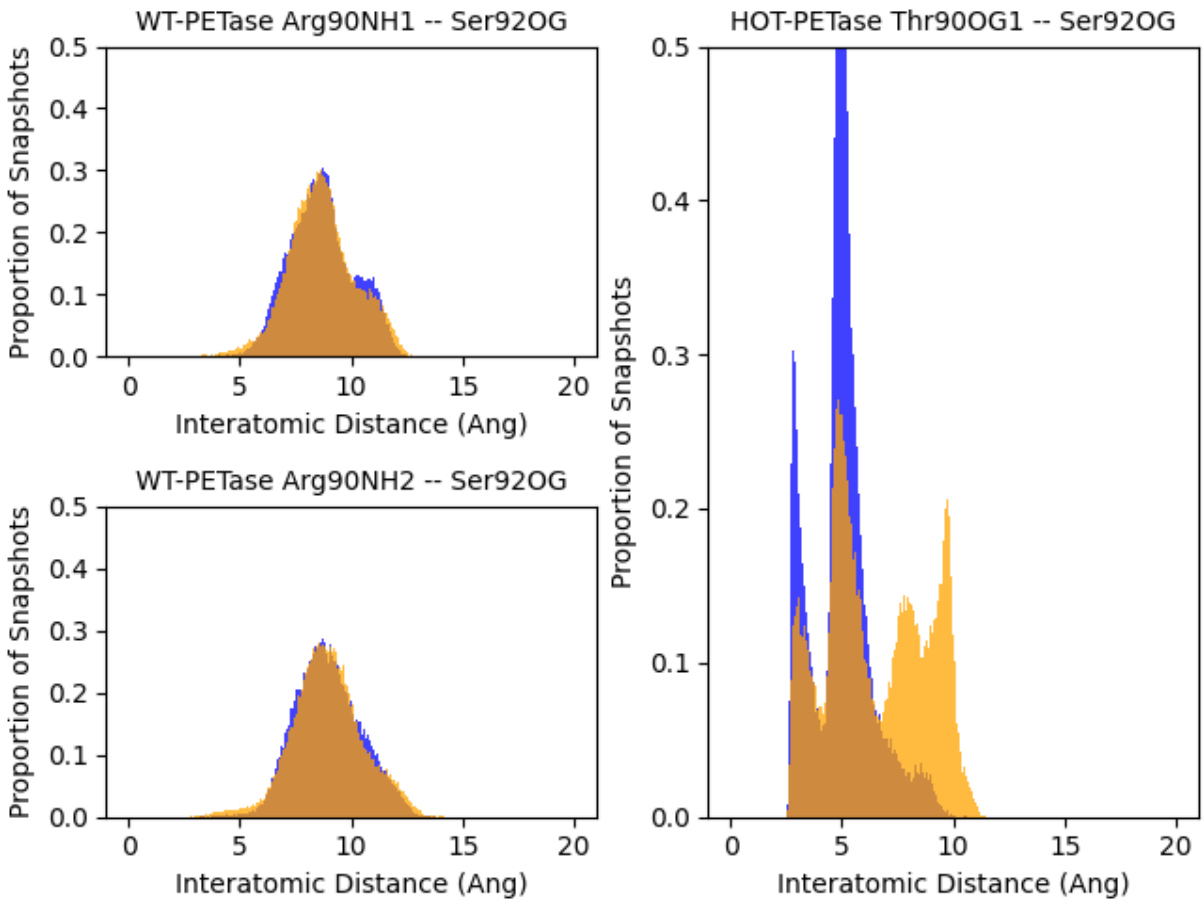


Figure 69: Distance distributions for interactions between Arg90 and Ser92 in WT-PETase (top left, bottom left) and R90T and Ser92 in HOT-PETase (right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

The long sidechain of K95N in HOT-PETase made the formation of hydrogen bonds with many nearby residues possible. The sidechain of K95N contains a hydrogen bond donor ($ND2_{K95N}$) and a hydrogen bond accceptor ($OD1_{K95N}$). This further increases the number of residues that have potential to form hydrogen bonds with K95N (see Figure 68). To identify hydrogen bonds formed by K95N, we selected all hydrogen bond acceptors within 6 Å of K95N and took measurements between these and $OD1_{K95N}$ (see Figure 70). We also took measurements between all hydrogen bond donors within 6 Å of K95N and $ND2_{K95N}$ (see Figure 71).

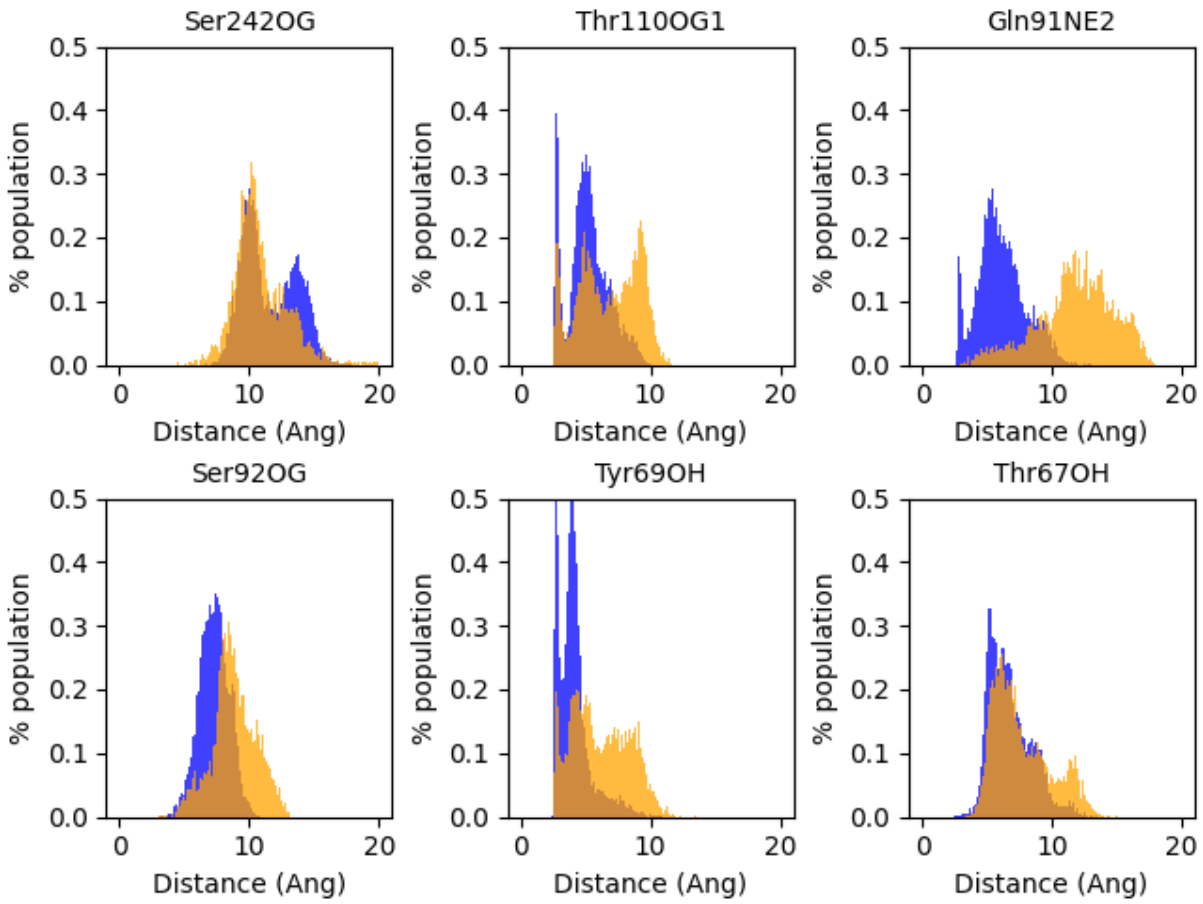


Figure 70: Distance distributions for interactions between $OD1_{K95N}$ and the sidechain hydrogen bond acceptors on nearby residues in Hot-PETase (residue and atom name in plot titles). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

Our measurements for the interactions of $OD1_{K95N}$ suggest that at low temperatures (i.e., in our 300K simulations), $OD1_{K95N}$ is capable of forming strong hydrogen bonding interactions with the sidechain hydroxyl groups of Tyr69 and Thr111 (see Figure 70). These interactions can be observed by peaks in their respective distance distributions at ca. 3 to 4 Å. Through visual inspection of our MD trajectories, we found that the hydrogen bond between $OD1_{K95N}$ and Thr111 can indeed form. The formation of a hydrogen bond between $OD1_{K95N}$ and Tyr69 does not seem likely due to the orientation of each of residue's sidechains. We do however observe a hydrogen bond between $ND2_{K95N}$ and the hydroxyl group of Tyr69 (discussed below, see Figure 71).

A weaker interaction can be observed between $OD1_{K95N}$ and the sidechain amide of Gln91 (see Figure 70) with a peak in the distance distribution at ca. 5 Å. In our simulations performed at 360K, this peak is almost entirely replaced with one at ca. 12 Å. This suggests that if any stabilising interaction between $OD1_{K95N}$ and Gln91 can form, it is not particularly thermostable.

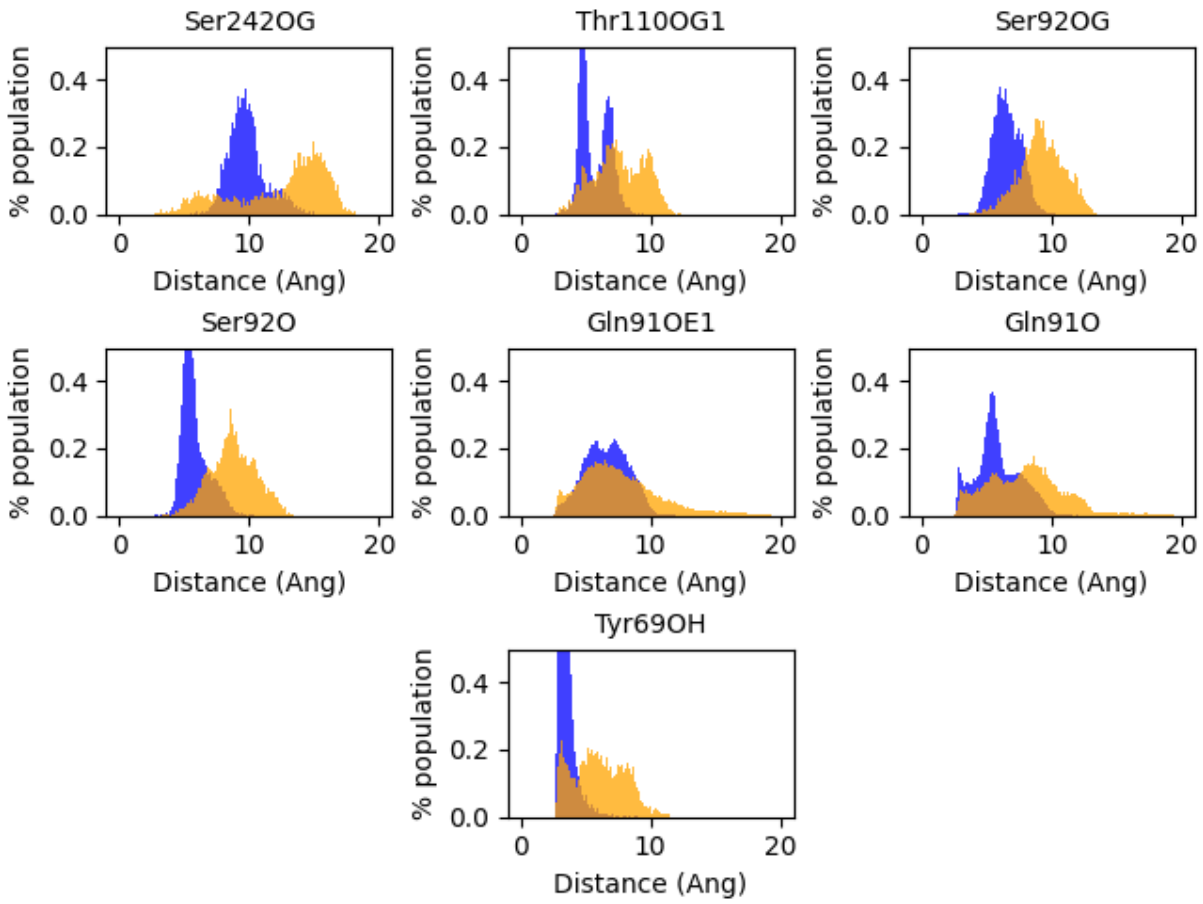


Figure 71: Distance distributions for interactions between $ND2_{K95N}$ and the sidechain hydrogen bond acceptors on nearby residues in Hot-PETase (residue and atom name in plot titles). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

Our measurements for the interactions of $ND2_{K95N}$ suggest that at low temperatures the sidechain hydroxyl of Tyr69 forms the strongest interaction with $ND2_{K95N}$. This can be observed by the peak at ca. 3 Å in the corresponding distance distribution (see Figure 71), indicating the formation of a strong hydrogen bond for a large proportion of timesteps in our MD simulations at 300K. For our higher temperature simulations (360K), the distance distribution broadens significantly. This suggests that the $NS2_{K95N}$-Tyr69 interaction is not particularly thermostable.

4.3.2.2.2.2 HOT-PETase Cluster 2

Cluster 2 in HOT-PETase is analogous to Cluster 2 in FAST-PETase (discussed above). It is comprised the mutant residues D187H and S121E, and the wild-type residue Asn172. The mutations D187H and S121E are also present in FAST-PETase. X-ray crystallography of FAST-PETase revealed that WMHBs can form between D187H and S121E and between S121E and Asn172. It would be reasonable to assume that these WMHBs can also form in HOT-PETase (for atom naming conventions, refer to Figure 61, as these are conserved between FAST-PETase and HOT-PETase). Distance measurements between $NE2_{H187}$ and the carboxyl oxygen atoms of S121E ($OE1_{S121E}$ and $OE2_{S121E}$) as well as

measurements between $OE1_{S121E}$ and $OE2_{S121E}$ and $ND2_{N172}$ in our simulations of HOT-PETase were made. These measurements provide evidence for the formation of WMHBs forming within this cluster of residues. In both WT-PETase and HOT-PETase, in our simulations performed at 300K the median distance between hydrogen bond donating/accepting atoms of residues 186 and 121 is ca. 7.5 Å. This suggests that a WMHB can form for a significant portion of our simulation. At 360K, the median value for this distance increases for WT-PETase to ca. 10 Å. By contrast, the median distance for HOT-PETase does not increase significantly when the simulation temperature is increased from 300K to 360K. This data suggests that the novel WMHB between D186H and S121E in HOT-PETase is more thermostable than the equivalent Asp186-Ser121 WMHB in WT-PETase.

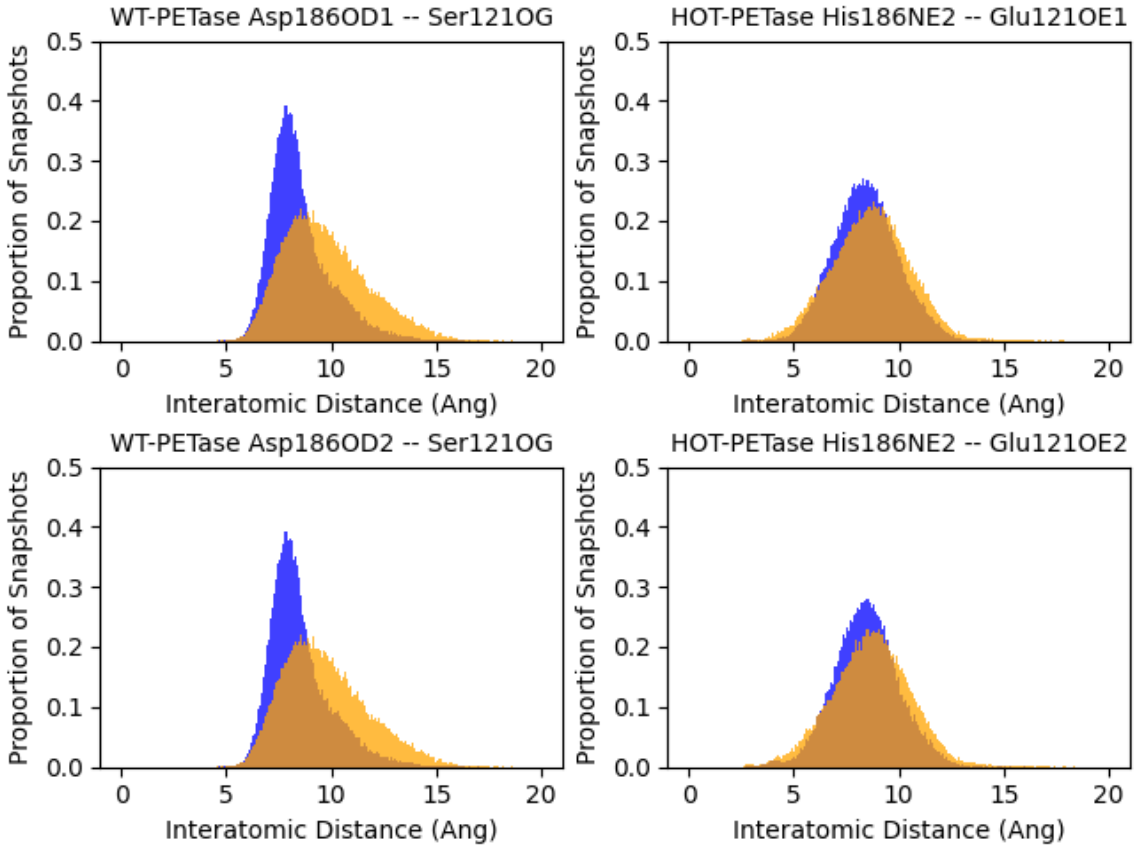


Figure 72: Distance distributions for interactions between Asp186 and Ser121 in WT-PETase (top left, bottom left) and D186H and S121E in HOT-PETase (top right, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

For the measurements between hydrogen bond donors/acceptors in residues 121 and 172, a similar result was found: in our 300K simulations, the median distance is ca. 7.5 Å for both WT-PETase and HOT-PETase. For both enzymes, the median distance increases to ca. 10 Å. This data suggests that at low temperatures, WMHBs may exist for the majority of the time between residues 121 and 172 in both WT-PETase and HOT-PETase. At high temperatures however, the distances between the hydrogen bond donors/acceptors become too long for WMHBs to form.

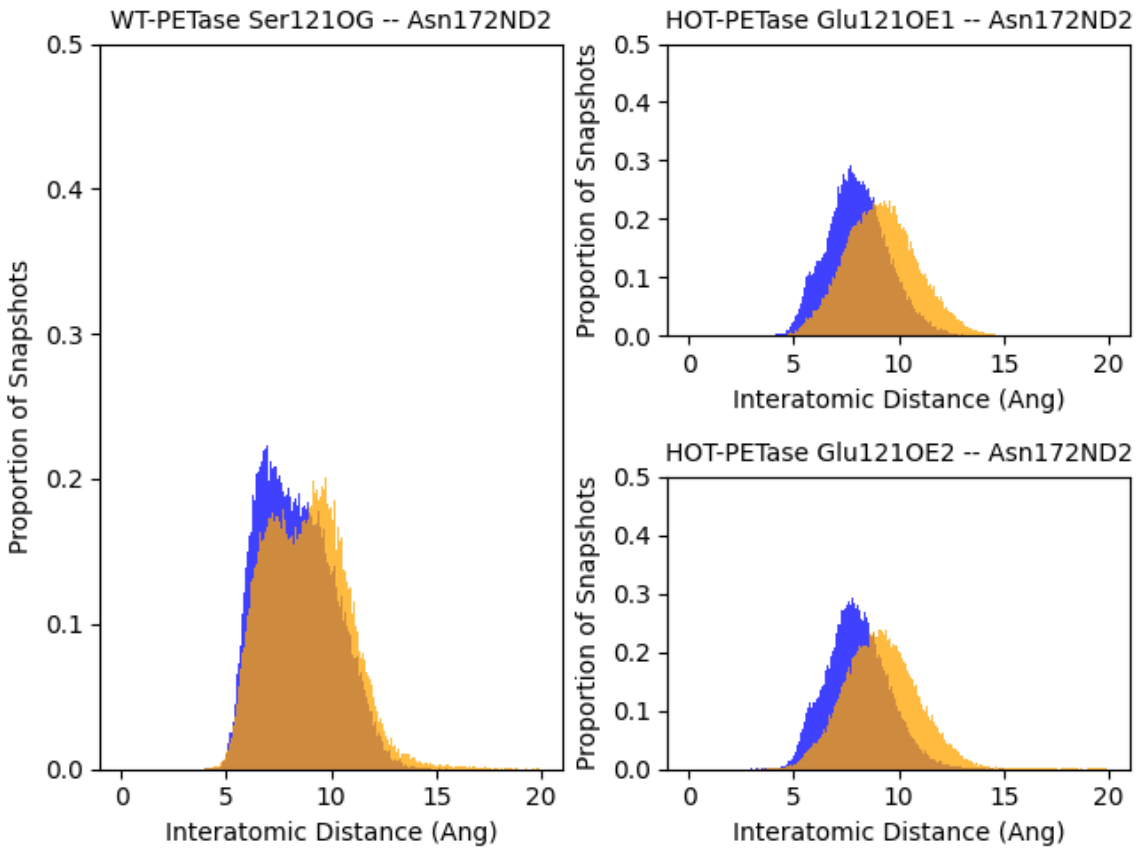


Figure 73: Distance distributions for interactions between Ser121 and Asn172 in WT-PETase (left) and S121E and Asn172 in HOT-PETase (bottom left, bottom right). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

4.3.2.2.2.3 HOT-PETase Cluster 3

Our RMSF analysis highlighted that the β3-α2 loop had slightly reduced flexibility in HOT-PETase compared to in WT-PETase (see Figure 55). The mutation Q119K is located on the β3-α2 loop in HOT-PETase. It is therefore reasonable to believe that this mutation is responsible for this increase in loop rigidity. The most likely reason the Q119K mutation would increase the rigidity of the β3-α2 loop would be if it formed a new hydrogen bond or salt-bridge interaction with a nearby residue. To identify residues with potential to form interactions with Q119K, we selected residues within 6 Å of Q119K with either hydrogen bond acceptors in their sidechains.

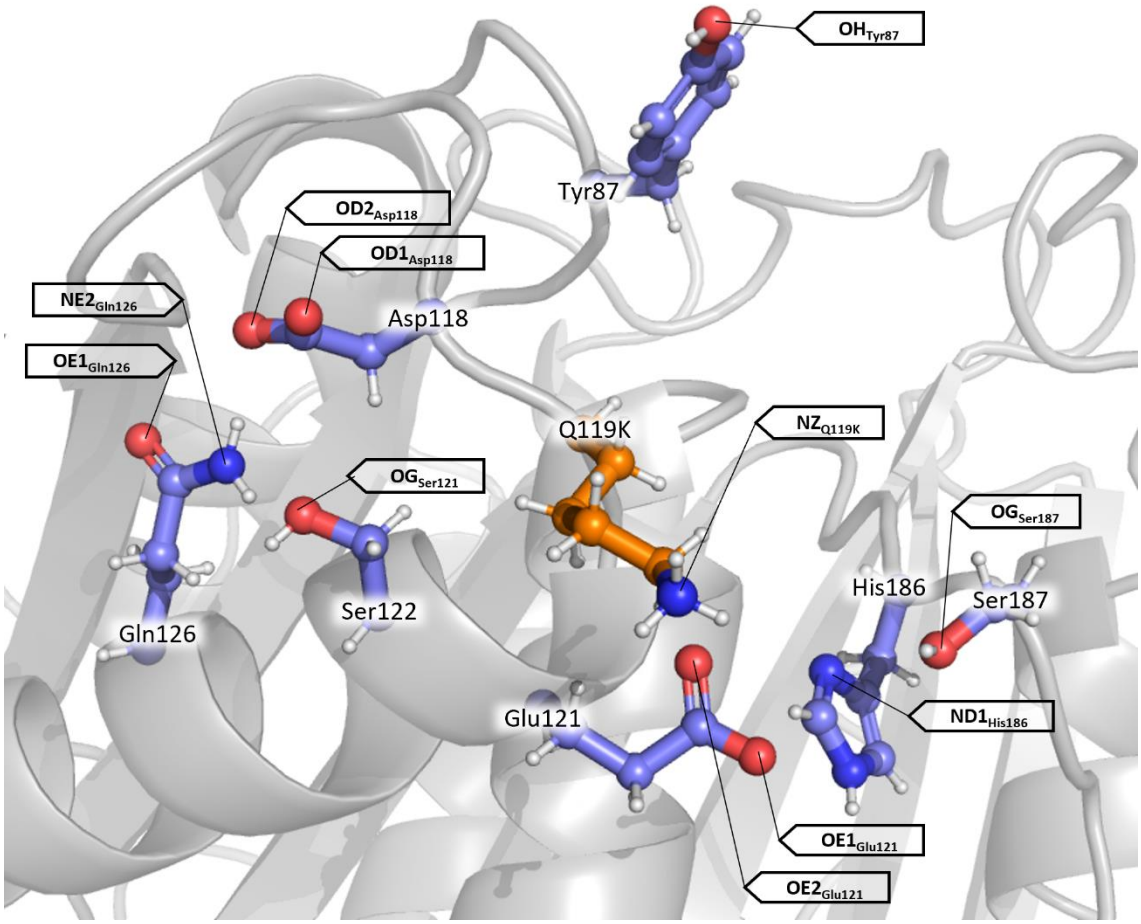


Figure 74: The mutant residue Q119K (shown in orange). Surrounding residues with hydrogen bond acceptors that had potential to form hydrogen bonds or salt-bridges with Q119K (shown in blue).

Through visual inspection of the crystal structure of HOT-PETase[39] we concluded that no backbone oxygens were oriented suitably to form interactions with Q119K. Measurements were taken in our MD simulations for HOT-PETase for the distances between NZ_{Lys119} and the sidechain hydrogen bond acceptors on the following residues: Tyr87, Asp118, Glu121, Ser122, Gln126, His186 and Ser187 (see Figure 75).

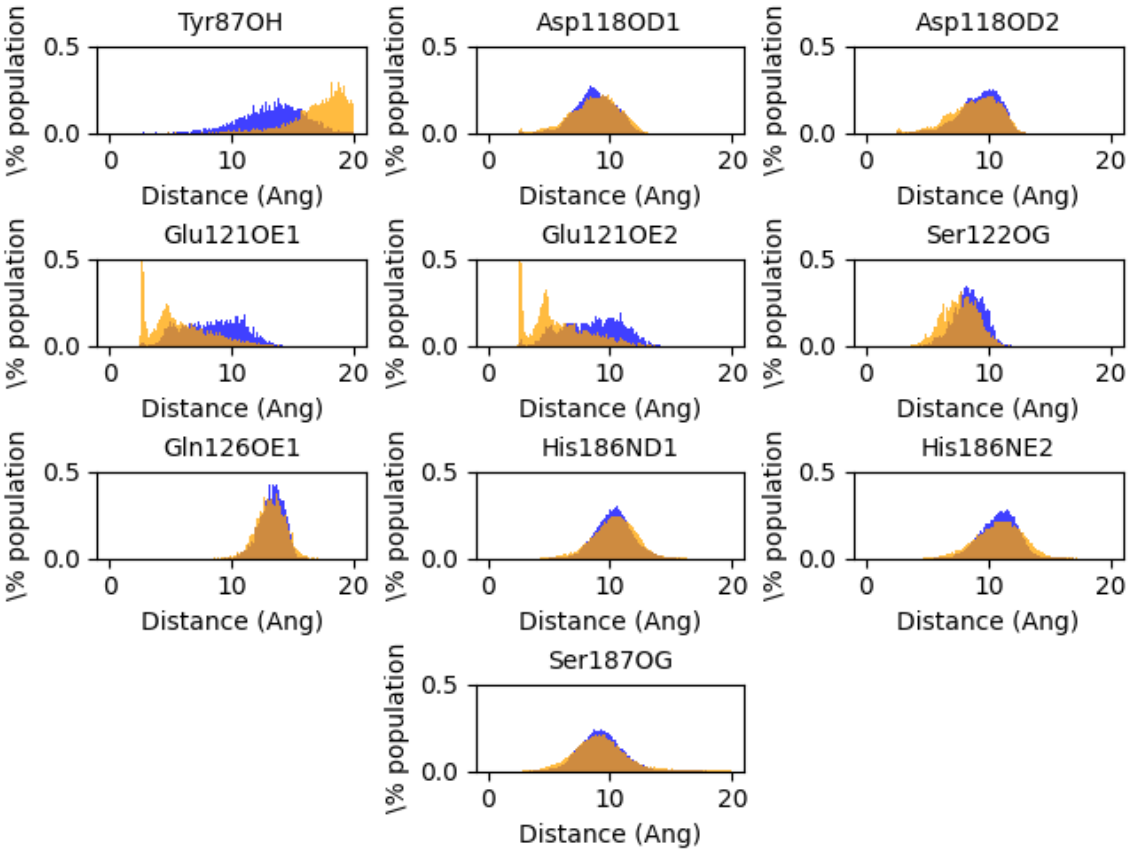


Figure 75: Distance distributions for interactions between Q119K and the sidechain hydrogen bond acceptors on nearby residues (residue and atom name in plot titles). Distance distributions are shown for MD simulations performed at 300K (blue) and 360K (orange).

The majority of the distances we measured between Q119K and nearby hydrogen bond acceptors had median distances of ca. 10 Å or above. These distances would correspond with fairly weak attractive interactions. The interaction between Q119K and Ser122 had a median distance of ca. 8 Å; this may represent a stronger attractive interaction. Most interestingly, at 300K our distance measurements clearly show no (or only a weak) interaction between Q119K and S121E. At 360K however, we observe peaks in the distance distributions at ca. 3 Å and ca. 4 Å. These peaks indicate a strong salt-bridge type interaction between the positively charged sidechain of Q119K and the negatively charged sidechain of S121E. It is unclear why this salt-bridge only forms in our high-temperature simulations. It is however possible that this novel salt-bridge plays a part in stabilising the WMHBs S121E forms with Asn172 and D187H (discussed in section 4.3.2.2.2.2). It is also possible that a significant energy barrier is associated with the formation of the Q119K-S121E salt-bridge; this would explain why it exists in our 360K simulations and not in our 300K simulations.

4.3.2.3 Visual inspections of MD trajectories used to confirm presence of interactions

Our distance measurements discussed in section 4.3.2.2 can be used to provide evidence for the existence of various novel interactions involving mutant residues in the enzymes FAST-PETase and HOT-PETase. In order to confirm that these interactions do in fact form throughout our MD simulations, we visually inspected portions of our trajectories. A short script was written that identified timesteps where each distance of interest was shortest, then converted a small portion of the trajectory around this timestep to a file format that could be visually examined (in practice this involved the conversion of the .trr file to a .pdb file). In cases where the distance of interest was too long for the formation of a strong hydrogen bond, we selected nearby solvating water molecules to assess whether a WMHB was present.

The following section will explore our findings using this visual inspection method for interactions within the enzymes FAST-PETase and HOT-PETase.

4.3.2.3.1 Visual inspection of MD trajectories for FAST-PETase

Our RMSF analysis revealed a small reduction in the flexibility of the β6- β7 loop of FAST-PETase, relative to that of WT-PETase. The mutant residue R224Q is located on the β6- β7 loop. In the original paper for FAST-PETase[37], it was suggested that a novel interaction between the mutant residue R224N and Ser192 was present in FAST-PETase. This conclusion was reached due to the R224Q – Ser192 hydrogen bond being present in the X-ray crystal structure of FAST-PETase[37]. Our distance analysis however showed that the hydrogen bond donors/acceptors in R224Q and Ser192 remain too far apart throughout our MD simulations for a hydrogen bond to form. Visual inspection of our MD simulations confirmed that no hydrogen bond formed between R224Q and Ser192 (see Figure 76). We took measurements between R224Q and Ser193 to assess whether a hydrogen bond existed between these two residues. Again, we found that no hydrogen bond formed between R224Q and Ser193 throughout our MD simulations (see Figure 76).

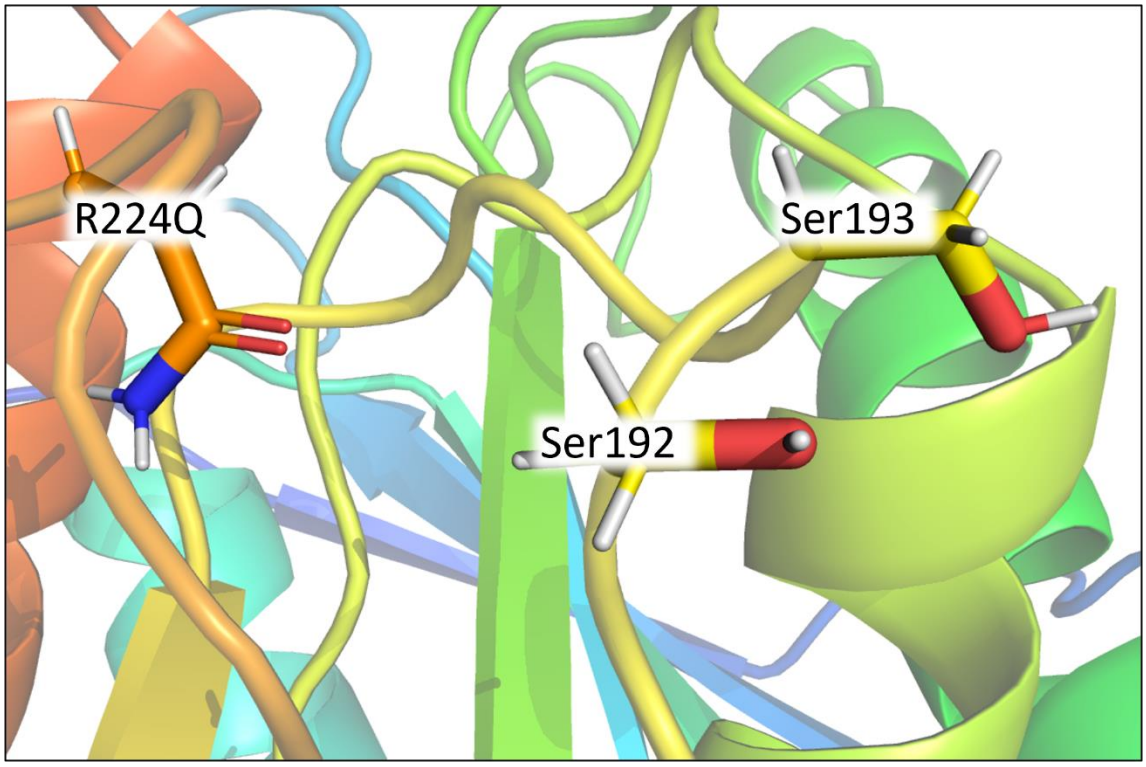


Figure 76: The lack of hydrogen bonding between the mutant residue R224Q and the residues Ser192 and Ser193. Image generated from a snapshot of one of our MD simulations performed at 300K.

The original paper suggested that there were WMHBs in between residues D186H, S121E and Asn172, based on the X-ray crystal structure of FAST-PETase[37]. This cluster of residues is proximal to the β6- β7 loop, which exhibited a small decrease in flexibility in our RMSF analysis. Our distance analysis revealed that, throughout our simulations, these residues remain within a suitable distance for WHMBs to form (discussed in section 4.3.2.2.1.2). Through visual inspection of our MD trajectories, we found no cases where a WMHB formed between D186H and S121E or between S121E an D186H (see Figure 77).

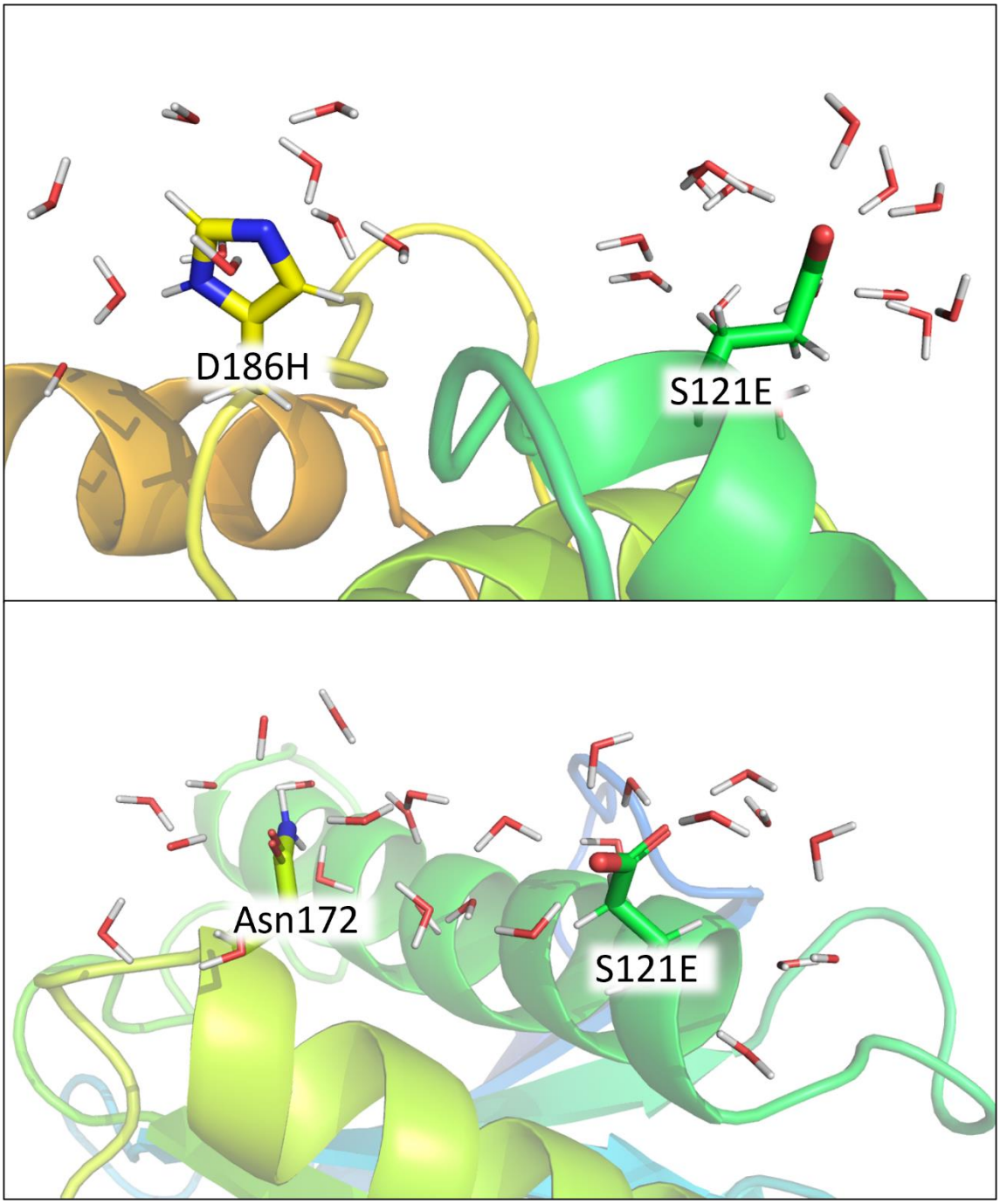


Figure 77: The residues D185H and S121E (top) and S121E and Asn172 (bottom) with surrounding solvent water molecules. Images were generated from snapshots where the relevant hydrogen bond donors/acceptors were closest.

We also investigated the presence of a salt-bridge between the mutant residue N233K and Glu204K. This interaction can be observed in the X-ray crystal structure of FAST-PETase[37]. Although no significant changes in the flexibility of nearby loops was observed in our RMSF analysis, we still investigated the presence of this interaction using distance measurements and visual inspection of our MD trajectories. Through these techniques we were able to confirm that a salt-bridge between N233K and Glu204K does indeed exist in FAST-PETase (see Figure 78).

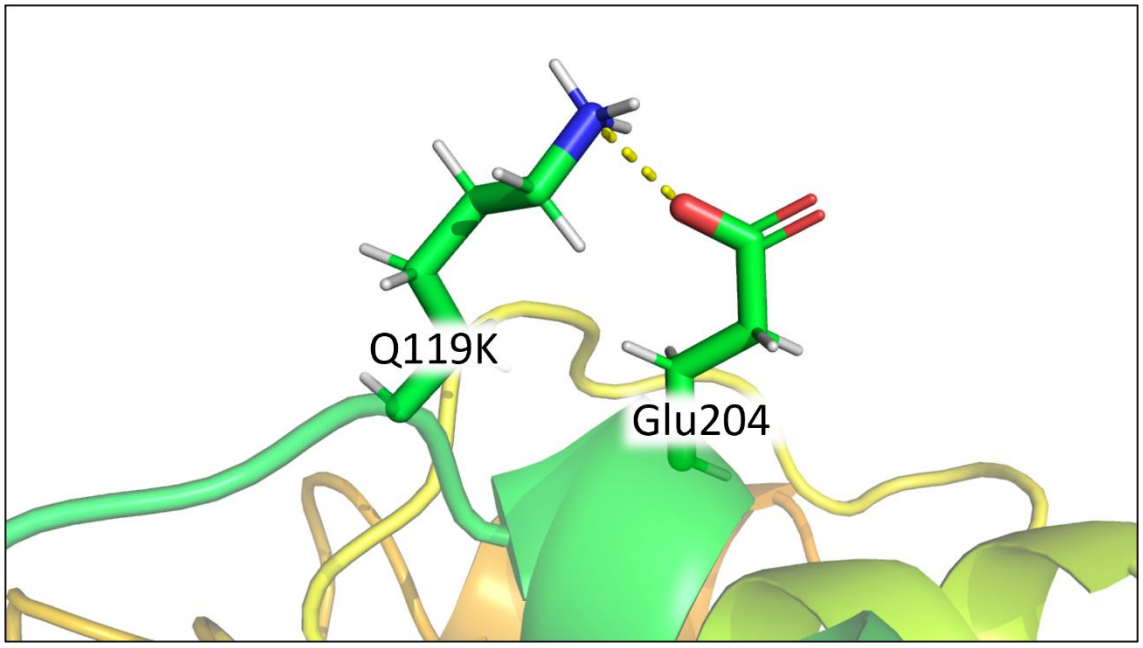


Figure 78: A salt-bridge interaction between Q119K and Glu204 in FAST-PETase. Image generated from a snapshot of one of our MD simulations of FAST-PETase at 300K.

4.3.2.3.2 Visual inspection of MD trajectories for HOT-PETase

Our RMSF analysis identified the residues K90T and K95N as potentially important for the dramatic increase in flexibility of the nearby β3-α2 loop of HOT-PETase compared to that of WT-PETase (discussed in section 4.3.2.1). Our distance analyses of these residues revealed that K90N has the potential to form a strong hydrogen bond with the sidechain of Ser92. Our visual inspection of our MD trajectories revealed that this interaction is indeed possible.

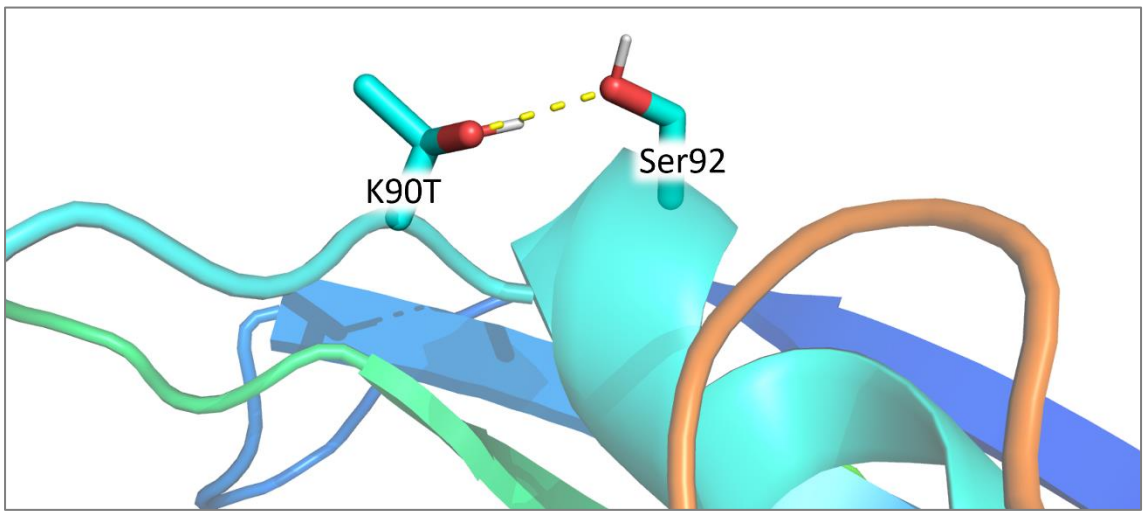


Figure 79: The hydrogen bonding interaction between K90T and Ser92, observed in a snapshot taken from one of our MD simulations of HOT-PETase performed at 300K.

Our distance analysis of K95N revealed a more complex picture of its non-covalent bonding interactions with its neighbouring residues. The length of the sidechain of K95N as well as the

presence of both a hydrogen bond donor ($ND2_{K95N}$) and acceptor ($OD1_{K95N}$) gives the residue the potential to form hydrogen bonds with a large number of nearby residues. Through visual inspection of each interaction with a suitably short distance (discussed in section 4.3.2.2.2.1), we found that $OD1_{K95N}$ can form hydrogen bonds with the residues Gln91 and Thr110, as well as a WMHB with Thr67.

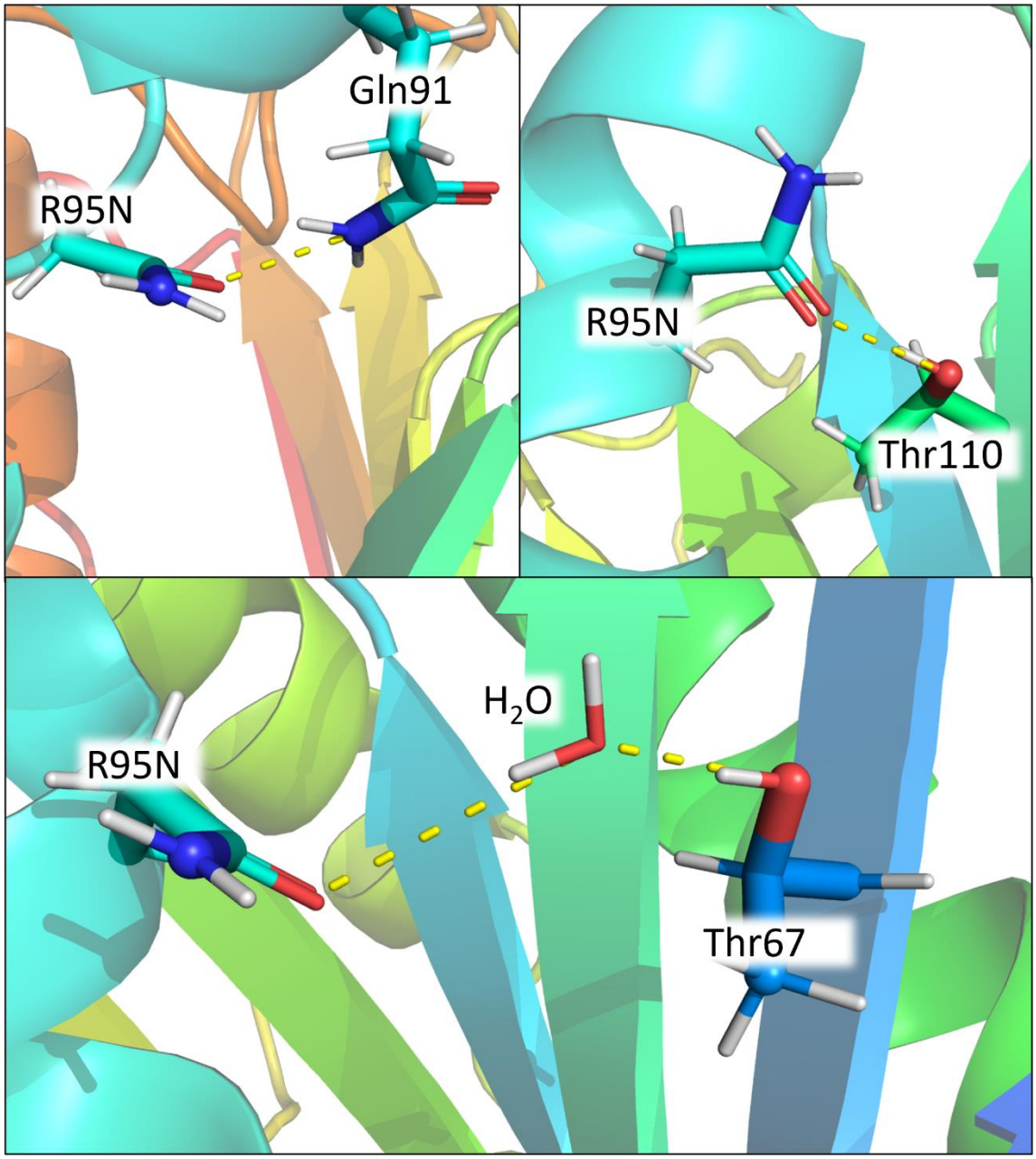


Figure 80: The hydrogen bonding interaction between $OD1_{K95N}$ and the following residues: Gln91 (top left), Thr110 (top right) and Thr67 (bottom). Images were generated from snapshots taken from one of our MD simulations of HOT-PETase performed at 300K.

A similar analysis revealed that K95N is capable of forming hydrogen bonding interactions with the residues Tyr69, Thr110 and Gln91 via its $ND2_{K95N}$ atom (see Figure 81).

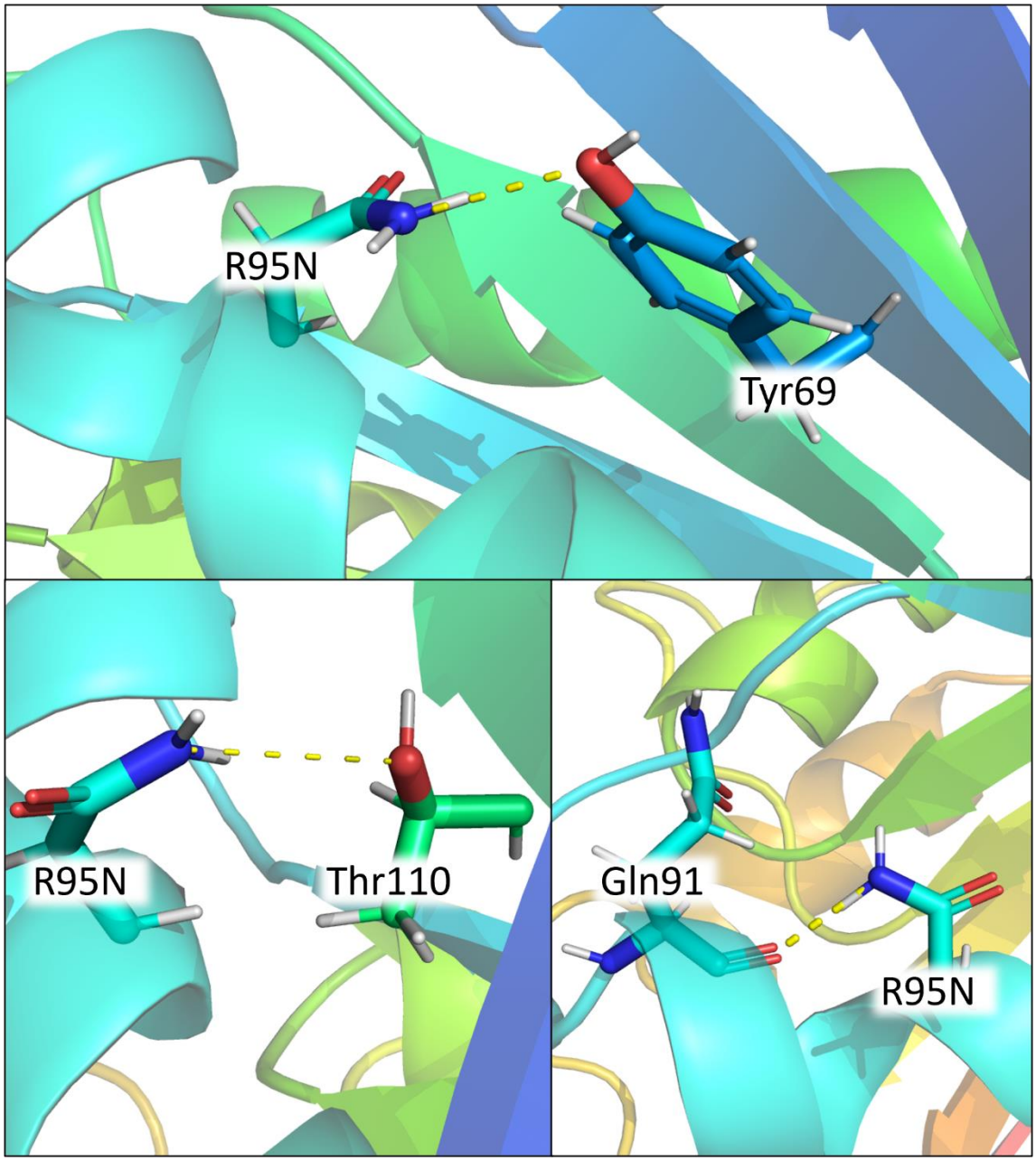


Figure 81: The hydrogen bonding interaction between $ND2_{K95N}$ and the following residues: Tyr69 (top), Thr110 (bottom left) and Gln91 (bottom right). Images were generated from snapshots taken from one of our MD simulations of HOT-PETase performed at 300K.

Our RMSF analysis revealed that the residues D187H, S121E and Asn172 could be of importance in reduction of the flexibility of the β6-β7 of HOT-PETase. The mutations D187H and S121E are also present in the enzyme FAST-PETase. The creators of FAST-PETase[37] observed WMHBs between D187H and S121E, and between S121E and Asn172 (discussed in section 4.3.2.2.1.2). For HOT-PETase. Our distance measurements revealed that the distances between hydrogen bond donors and acceptors within this cluster of residues were in a suitable range for the formation of WMHBs. Through visual inspection of our MD trajectories, we observed these WMHB interactions (see Figure 82).

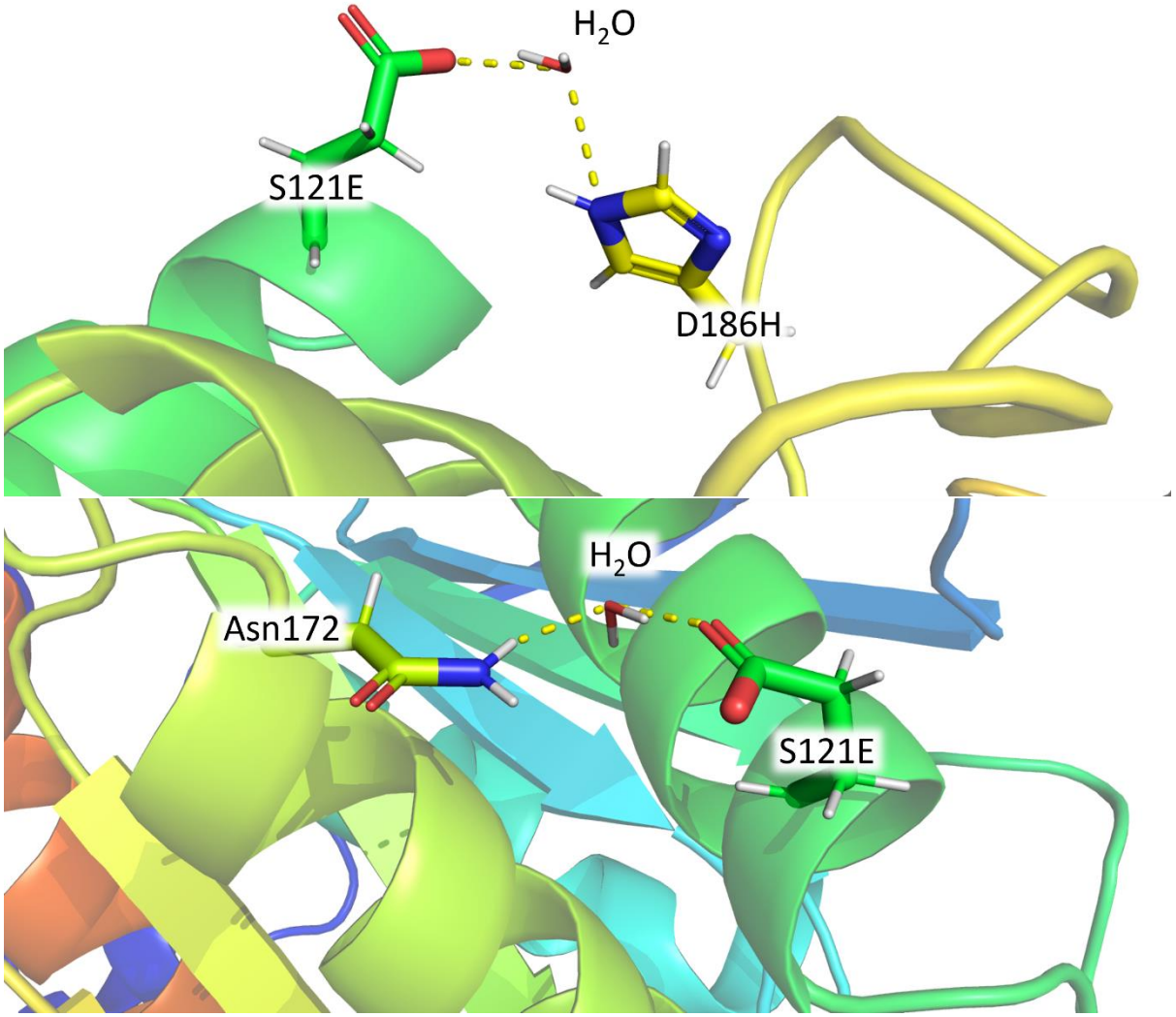


Figure 82: Water-mediated hydrogen bonding interactions between D186H and S121E (top) and between S121E and Asn172 in HOT-PETase. Images were generated from snapshots taken from one of our MD simulations of HOT-PETase performed at 300K.

Our RMSF analysis also identified the mutant residue Q119K as having potential importance for the dramatic increase in the flexibility of the nearby β3-α2 loop. Our distance analyses of potential interactions of Q119K with nearby residues identified a potential salt-bridge between the positively charged sidechain of Q119K and the negatively charged sidechain of S121E. Interestingly, this interaction does not seem to occur in our 300K MD simulations of HOT-PETase but occurs in our 360K simulations. We suggest that this is due to the breaking of the WMHBs between S121E and the residues D187H and Asn172 (discussed above) at higher temperatures.

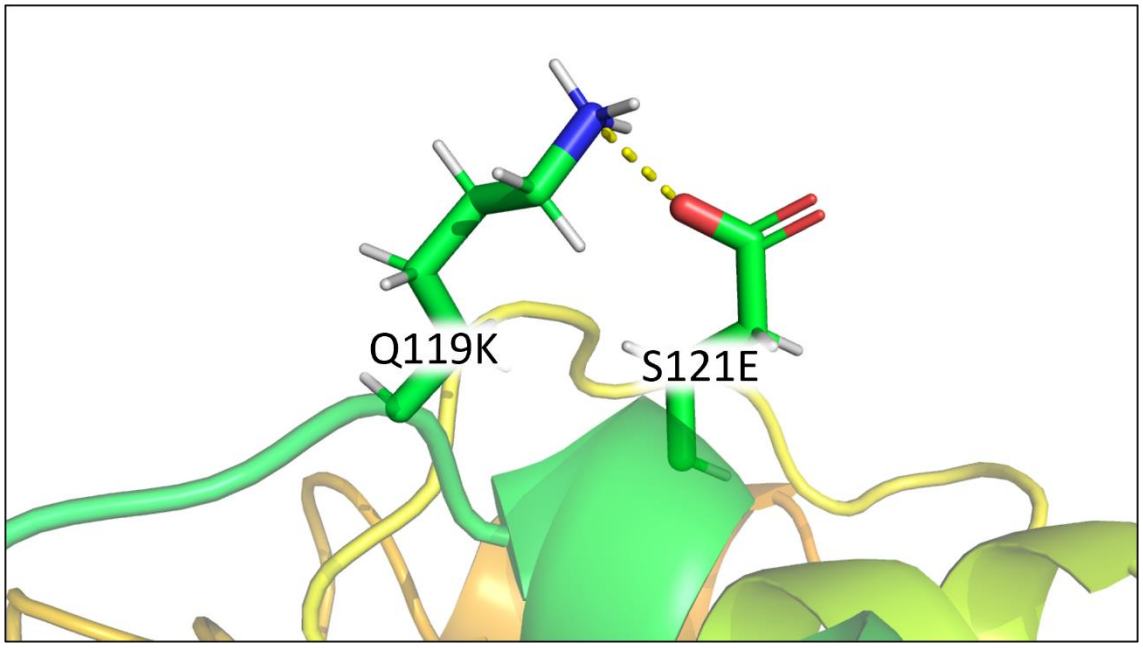


Figure 83: A salt-bridge between the residues Q119K and S121E. Image was generated from a snapshot taken from one of our MD simulations of HOT-PETase performed at 360K.

4.4 Conclusions

4.4.1 Introduction of Disulfide bond to Is-PETase has no significant effect upon reaction profiles generated by QM/MM methods

The objective of the introduction of an additional disulfide bond into *Is*-PETase was to increase the enzyme's thermostability without compromising the energetics of its chemical mechanism. In order to assess this mutant enzyme's thermostability, a long MD simulation of the mutant enzyme was compared to an equivalent simulation of WT-PETase as well as MD simulations of highly thermostable *Is*-PETase mutants created by other groups[34,39,129]. To ensure that the introduction of an additional disulfide bond did not increase energy barriers in the enzyme's chemical mechanism, QM/MM methods were applied to calculate reaction profiles for the degradation of a PET substrate mimic by the mutant enzyme. These reaction profiles were then compared to those previously generated for WT-PETase.

From our energy profiles calculated for DISU-PETase, we can conclude that the introduction of a novel disulfide bond into *Is*-PETase through the mutations S163C-S194C has no significant impact on the energetics of its chemical mechanism. This conclusion has been reached as the majority of energies calculated for each intermediate and transition state calculated for DISU-PETase's reaction pathways lie within the energy range of equivalent geometries calculated for WT-PETase.

By confirming that the S163C-S194C double mutation has no deleterious effects upon the catalytic mechanism of *Is*-PETase, we can therefore recommend that this mutant variant of *Is*-PETase be created experimentally and assayed for activity against PET. It is our assumption that the introduction of this novel disulfide bond will improve the enzyme's thermostability, which has been shown to be an important characteristic for PETase enzymes.

4.4.2 MD simulations provide some rationalisation of effects of mutations on thermostability in Is-PETase variants

We conducted several MD simulations on the enzymes WT-PETase, FAST-PETase, and HOT-PETase at temperatures of 300K and 360K. Analysis of the average RMSF per residue in these simulations allowed us to identify regions in each FAST-PETase and HOT-PETase where mutations had made an effect upon their flexibility.

In FAST-PETase, we identified two regions with slight reductions in their flexibility: the α4-β5 loop and the β6-β7 loop. We selected three clusters of residues situated near these loops and investigated their non-covalent interactions by analysing the evolution of a series of interatomic distances. Our distance measurements revealed no strong hydrogen bonding interactions between

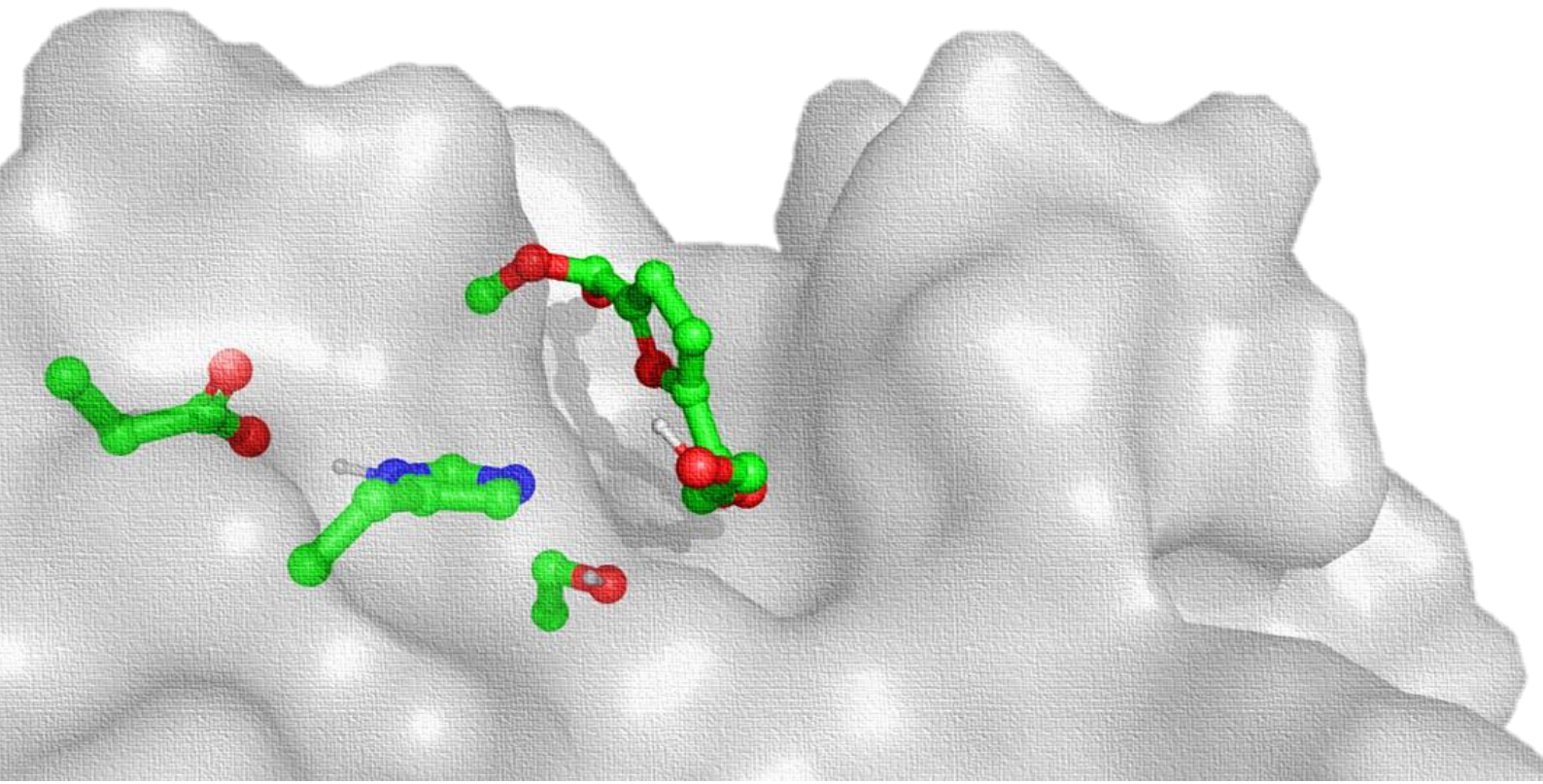
the mutant residue R224Q and either Ser192 or Ser193. This result is in contrast with the findings from the crystal structure of FAST-PETase[37], which showed a hydrogen bond between R224Q and Ser193. Our distance measurements for the cluster of residues containing D186H, S121E, and Asn172 support the notion that water-mediated hydrogen bonding interactions occur between D186H and S121E and between S121E and Asn172. Our distance measurements also provide evidence for the formation of a novel salt-bridge between the mutant residue N223K and the residue Glu204. This interaction was also observed in the crystal structure of FAST-PETase.

In HOT-PETase, through our RMSF analysis we found that the mutations in HOT-PETase resulted in a reduction in flexibility in the β6-β7 loop. Surprisingly, we also observed increases in flexibility in the α2-β3 loop as well as in the β3-α3 loop. From this analysis, we identified several mutant residues in HOT-PETase that were likely to be responsible for these changes in flexibility. We then investigated these residues' non-covalent interactions via the time evolution of a series of distance measurements. Our distance measurements have been used to identify several possible stabilising interactions involving various mutant residues in HOT-PETase. We have provided evidence for the existence of novel interactions between the following residue pairs: K90T and Ser92, K95N and Tyr110, K95N and Tyr69, D186H and S121E, S121E and Asn127, and Q119K and S121E. We propose that these novel interactions play some part in the remarkable increase in stabilisation observed experimentally in HOT-PETase.

Overall, our analysis of MD simulations of WT-PETase, FAST-PETase and HOT-PETase provides a far more detailed (and complex) picture of each enzyme's intramolecular non-covalent interactions than X-ray crystallography can provide. We reveal that interactions that are present in X-ray crystallography are only transient in MD simulations and thus may provide considerably less stabilisation than the static X-ray structures suggest. We also have identified interactions that are not present in the X-ray crystal structure, including cases where one residue forms attractive interactions with multiple adjacent residues.

Chapter 5: Exploration of the Reaction Profile for the degradation of Poly(ethylene) Furanoate via Is-PETase

During her Master's Project, Chloe Martin helped develop some of the methods used in this chapter. None of Chloe's data is included in this thesis.



5.1 Background Information

5.1.1 Poly(ethylene) Furanoate as an Alternative to Poly(ethylene) Terephthalate

Despite its myriad uses and excellent physical properties, the negative environmental impact of the plastic PET has made its replacement with a more benign plastic desirable. The current frontrunner to replace PET is poly(ethylene) furanoate (in this work PEF, IUPAC name: poly(ethylene) 2,5-furandicarboxylate). PEF has similar physical properties to PET and in some cases that will be discussed later, has improved physical properties when compared PET. The primary draw of PEF however is our ability to derive it from plant matter. This greatly reduces the environmental impact of its production compared to the petrochemically derived PET.

Structurally, PEF is very similar to PET (see Figure 84); it is a copolymer formed from ethylene glycol (EG) and 2,5-furandicarboxylic acid (FDCA). The resulting chain is therefore repeating units of furan rings separated by ethylene linkers, joined together with ester bonds. The sole difference between PET and PEF therefore is the type of aromatic ring in the di-acid monomer (a furan ring in PEF and a phenyl ring in PET). The furan ring is similarly rigid to a phenyl ring, conferring generally similar properties to PEF as PET.

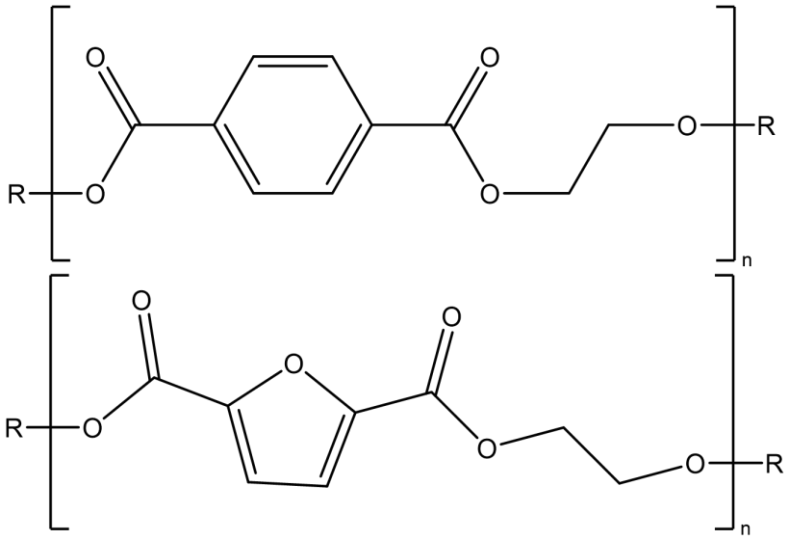


Figure 84: The chemical structure of the polymers PET (above) and PEF (below).

The idea of circular life-cycles has become increasingly popular over the past decade. This concept describes a material or product that can be broken down into its component parts and then recreated. In an ideal system, this would eliminate any waste products associated with the material and provide a sustainable source for its production. PEF provides an opportunity to produce a polymer with a circular life-cycle as, like PET, it can be hydrolysed into its monomers EG and FDCA which can then be repolymerised into PEF. This process can be performed enzymatically and will be

discussed in section 5.1.3. Where PEF becomes more interesting is that PEF can be biodegraded more indiscriminately using composting techniques. The resulting compost can be used to grow plants or algae; these in turn can produce sugars which can be used to synthesise virgin PEF. Mechanical recycling techniques can also be performed upon commercial PEF products, although some losses in desirable physical properties in the recycled material would be expected. A scheme for the potential circular life-cycle for PEF products is shown in Figure 85.

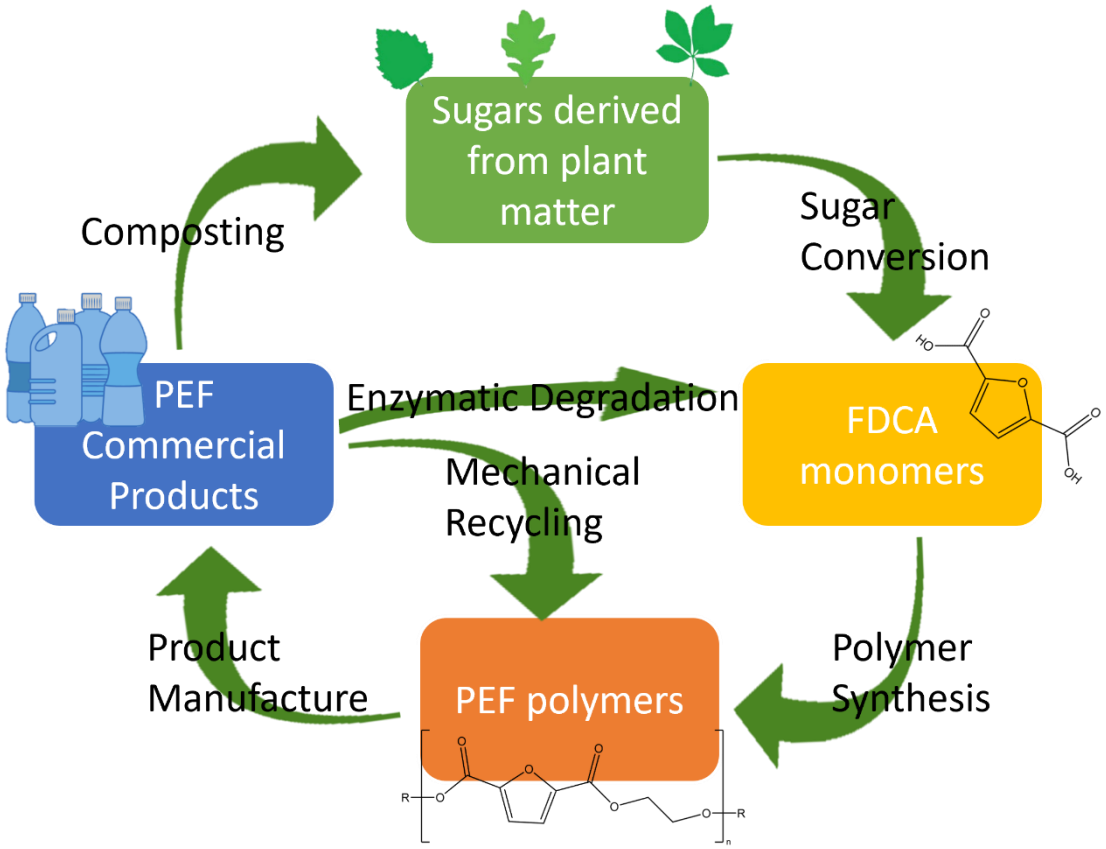


Figure 85: Scheme showing a potential circular life-cycle for PEF products.

5.1.2 Synthetic Pathways for Poly(ethylene) Furanoate

The first synthesis of PEF was reported in 1946. This was a bulk polyesterification process performed at 200°C and under high vacuum[130]. For over 50 years, the lack of demand for PEF products resulted in little improvement in this rather energy intensive synthesis method. It wasn't until 2009 that a novel synthetic method using a polytransesterification reaction of *bis*-(hydroxyethyl)-2,5-furanducarboxylate was introduced using an antimony (III) oxide catalyst[131]. Since then, several polycondensation and polytransesterification reactions with various catalysts have been investigated for the production of PEF. These classical chemical methods are performed at high temperatures and use various metal catalysts. It has been shown that these conditions can adversely affect the physical properties of the PEF material produced. Examples of this are discolouration, thermal instability and decreased electrical performance (e.g., reduced conductivity)[132]. Contamination of the metal catalysts in commercial products could also present a significant health risk to consumers. An alternate method using a rapid ring-opening polymerisation technique has been developed[133]. This method produces polymers with more desired physical properties (higher glass transition temperature, lower melting point, higher tensile strength and higher Young's modulus) without any discolouring. The cyclic oligomers of EG and FDCA required for the ring-opening polymerisation reaction can be prepared using non-metal catalysts; this eliminates the risk of metal contamination in the final product[134]. A range of environmentally benign ionic liquids have also been shown to catalyse the production of PEF via a polycondensation-type mechanism[135].

The enzymatic production of PEF and similar FDCA-derived copolymers has been investigated[136] using the enzyme *Candida antarctica* Lipase B (CALB). It was found that CALB has a preference for longer (at least three carbons long) chain aliphatic linkers between FDCA units. This makes it a non-ideal enzyme for the production of PEF as the EG linker has only 2 carbons. This study does provide a proof-of-concept for the biocatalytic production of PEF. This would be desirable as it would be more environmentally friendly than classical chemical techniques as it could be performed at ambient temperatures and pressures and would eliminate the need for metal-based catalysts[136].

5.1.3 Production of 2,5-Furandicarboxylic Acid from Sugars

Part of what makes PEF such an attractive material is that its monomer FDCA can be derived from plant matter. This would make a welcome departure from the petrochemically-derived terephthalic acid (TA) monomers used in PET production.

Starting from raw biomass, hexoses (sugar molecules containing six carbons) such as glucose, fructose and mannose must be extracted. These hexoses are then subjected to a dehydration process, whereby water molecules are removed from the sugar molecule by subjecting it to

dehydration under acidic conditions. These dehydration reactions have a large number of products, the primary product being 5-hydroxymethylfurfural (HMFF). There has been a large amount of work in optimising this process[132], including the use of an extremely wide array of catalysts and solvent systems. A major challenge in the production of HMFF is its instability under acidic conditions as it has a tendency to degrade into levulinic acid (LA) and formic acid (FA). A biphasic method using water and dimethylsulfoxide (DMSO) has been developed to alleviate this problem: The dehydration reaction occurs in the aqueous phase while the presence of DMSO suppresses the unwanted degradation of HMFF[137]. Another method to prevent the degradation of HMFF products is through the use of HMFF ethers, which form when alcohols are introduced into the solvent mixture[138].

It has been found that a greater proportion of HMFF is produced when fructose is used as the feedstock for this reaction. The problem with this is that fructose is significantly more expensive than glucose. For this reason, glucose is usually used as the initial feedstock and isomerised into fructose prior to the hydrogenation reaction. Alternatively fructose can be obtained from corn starch with relative ease as this process has been developed extensively for the production of high-fructose corn syrup in the food industry[139].

Once HMFF has been obtained through hexose dehydration, it can be oxidised to produce FDCA. While HMFF can be oxidised to FDCA using common industrial oxidants such as nitric acid and homogeneous metal bromides, yields have been reported to be fairly low (~60%)[138]. By contrast, a more tailored system using a platinum-based catalyst has been reported to have up to quantitative yields (~100%). Further improvements to this process (such as catalyst lifetime) were made using a 5% lead dopant. It is worth noting that although this lead dopant does provide some improvement to this reaction, it does little for the environmental sustainability of the overall process. Recently, an *in-situ* method for HMFF oxidation has been reported starting from a fructose feedstock[140]. A platinum/bismuth catalyst suspended on carbon was used to achieve this; unfortunately a low FDCA yield (~25%) is obtained using this method.

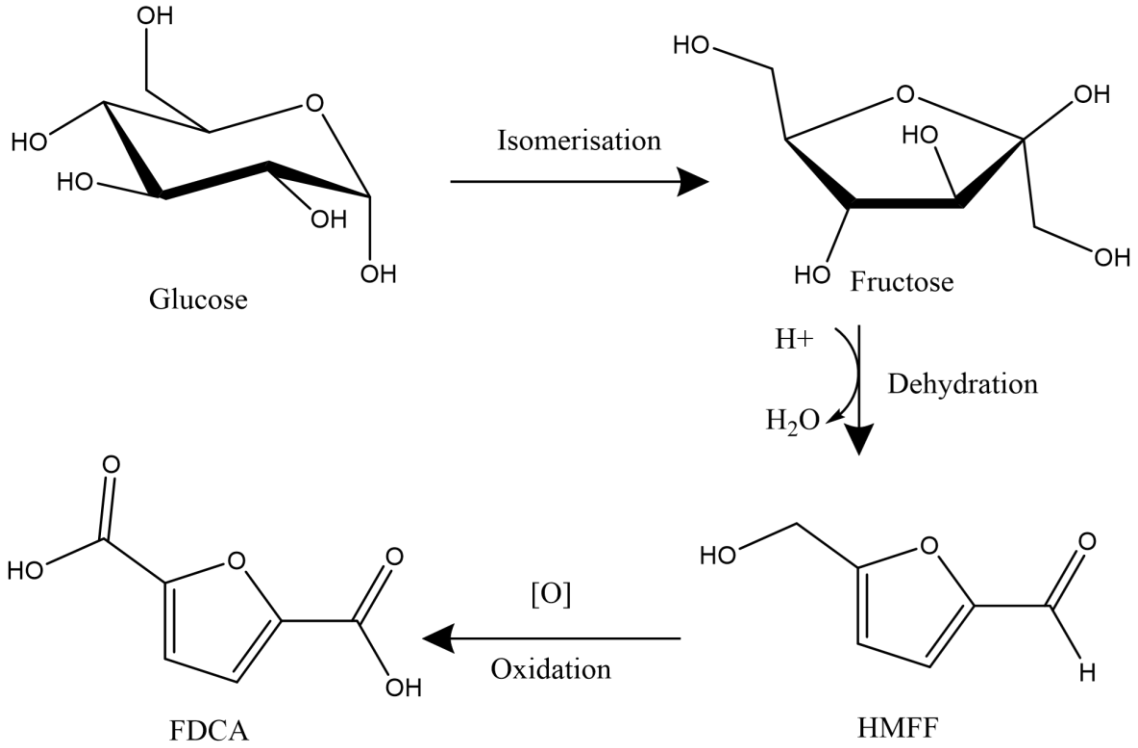


Figure 86: Scheme showing production of FDCA from glucose. Initially it is isomerised to fructose. Next a dehydration reaction converts it to HMFF. Finally, HMFF is oxidised to FDCA. The resulting FDCA can then be co-polymerised with ethylene glycol to produce PEF.

5.1.4 Enzymatic Hydrolysis of Poly(ethylene) Furanoate

If PEF is to be used as a drop-in replacement in PET products, it would be desirable to be able to efficiently biodegrade PEF into its constituent monomers (see Figure 85). Biodegradation would be preferable to mechanical recycling as it is more energy efficient and results in virgin product, eliminating concerns over loss in material properties in recycled products. Enzymatic degradation of PEF into its monomers for subsequent repolymerisation may be more efficient than simple composting of PEF products (although this would also require the use of PEF-degrading enzymes) because it would limit the number of steps between product disposal and creation of new PEF polymers (see Figure 85).

Shortly after the discovery of the enzyme *Is*-PETase[17], its ability to degrade PEF was reported[18]. In this study, scanning-electron microscopy (SEM) imaging of a synthesised PEF coupon (a small disc of convenient size to perform assays with) displayed more extensive pitting and scarring patterns than the equivalent PET coupon when treated with wild-type PETase. This non-quantitative measurement was supported by differential scanning calorimetry (DSC) measurements which showed a reduction in relative crystallinity of 15.7% for the PEF coupon, with a value of 10.1% for the PET coupon. This data suggests that not only is *Is*-PETase capable of degrading PEF, but it can degrade it at a higher rate than its 'natural' substrate PET. Molecular docking experiments were performed[18] for the tetramers of PET and PEF into the active site of *Is*-PETase. The nucleophilic distance (between OG_{S131} and CO_{LIG}) was found to be slightly shorter for the docked PEF tetramer (5.1 Å and 5.0 Å for PET and PEF tetramers respectively). While it may be tempting to interpret this result as a causal factor for *Is*-PETase's preference for PEF over PET, the minimal difference in distance and the poor accuracy of molecular docking protocols limits the value of this observation. To investigate this further, higher level *in silico* calculations (such as QM/MM geometry optimisations of the docked conformers) should be performed.

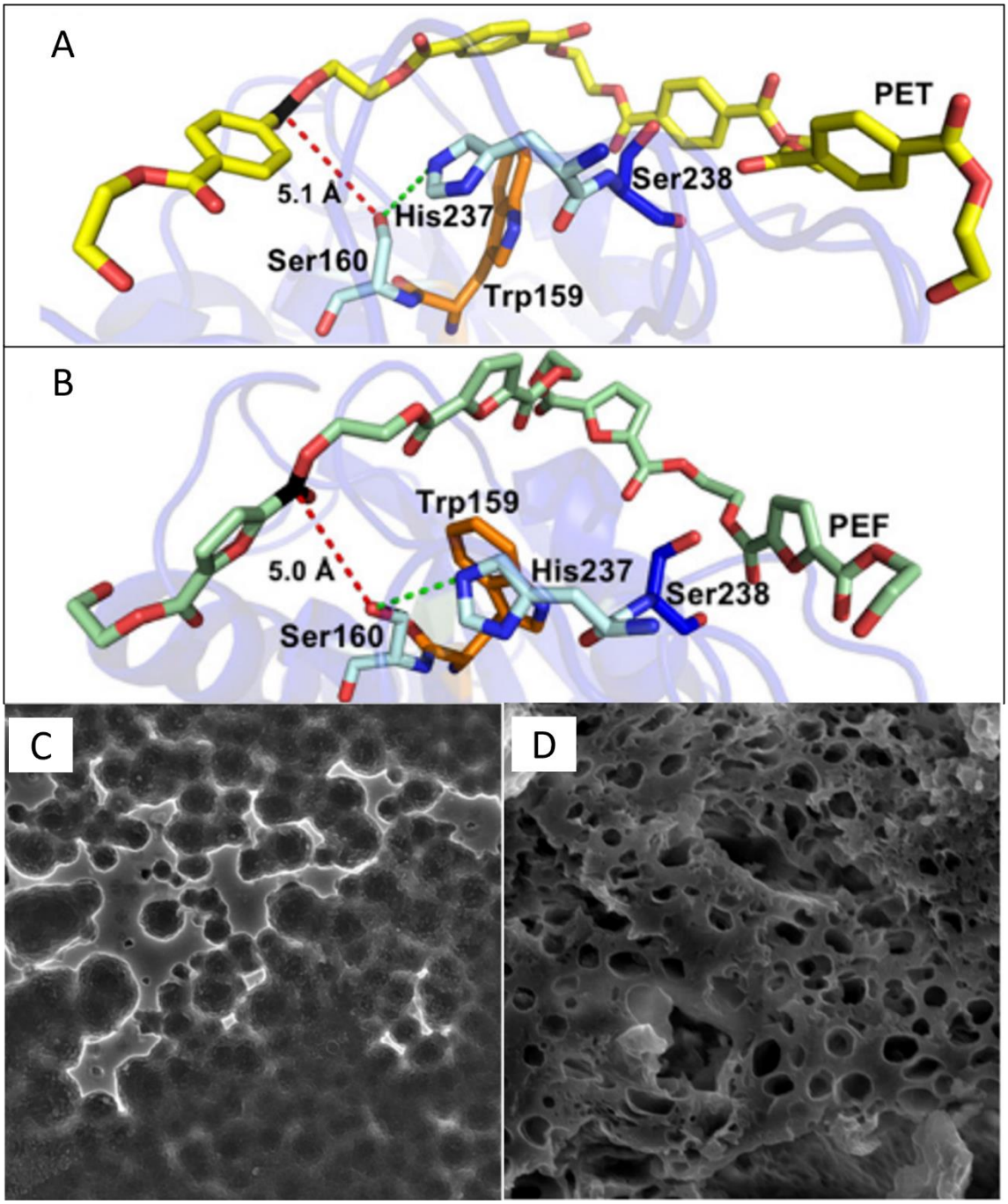


Figure 87: Molecular docking poses for tetrameric PET (A) and tetrameric PEF (B) into the active site of Is-PETase. SEM imaging of a PET coupon (C) and a PEF coupon (D) after incubation with Is-PETase. All images taken from Austin et al.[18].

In its original host bacterium *Idonella sakaiensis*, the enzyme *Is*-PETase works synergistically with a partner enzyme MHETase. *Is*-PETase is responsible for degrading polymeric PET into various products (including bis(2-hydroxyethyl) terephthalate (BHET) and mono(2-hydroxyethyl) terephthalate (MHET)); it then is capable of further degrading MHET and BHET into TA. MHETase is solely responsible for degrading MHET into TA. Since *Is*-PETase must be able to degrade polymeric PET, it has evolved a surface-exposed and rather open active site, while MHETase has evolved a

more specialised, buried active site. This results in MHETase having a far higher catalytic turnover with MHET as a substrate than *Is*-PETase. Together *Is*-PETase and MHETase can degrade polymeric PET faster than *Is*-PETase alone because MHETase can degrade the MHET product quickly, allowing the slower *Is*-PETase enzyme to catalyse the crucial degradation of polymeric PET instead. Interestingly, it has been shown that MHETase is incapable of degrading mono(2-hydroxyethyl) furanoate (MHEF)[141]. This is most likely due to the highly specialised binding site of MHETase being incapable of binding MHEF in a catalytic orientation. In order to create a highly efficient PEF-degrading biocatalytic system, it would be prudent to create an enzyme capable of degrading MHEF at rates similar to (or ideally above) that of MHETase. The obvious route to this potential "MHEFase" would be to modify the active site of MHETase to better accommodate MHEF as a substrate.

5.1.5 The Environmental impact of the Introduction of Poly(ethylene) Furanoate into Commercial Products

There are several advantages to the replacement of PET products with ones made of PEF. The primary advantage being the reduced environmental cost of PEF over PET. This environmental cost can be split up into the energy usage required to produce the material, the sustainability of the feedstock used to produce the material, the greenhouse gasses (GHGs) or other pollutants associated with the material's manufacture, and the environmental impact associated with the end-of-life of the products made from the material.

The full industrial process of the production of PEF from a corn feedstock has been modelled using the ASPEN Plus simulation modelling software package[139]. Using this simulation, it is estimated that replacing PET bottles with PEF bottles would result in a reduction of between 45% and 55% in GHGs produced throughout their manufacture. With ca. 15 million metric tonnes of PET bottles produced per year, this translates to between 20 to 35 Mt of CO_2 equivalents per year. Similarly, the energy cost of producing PEF bottles instead of using PET was calculated to be 40% to 50% less, which would translate to between 440 PJ to 520 PJ (petajoules) of non-renewable energy usage saved per year. To put those numbers in perspective, 520 PJ is roughly equivalent to one tenth of the total energy usage of the UK in 2020[142]. This study focuses on the impact of replacing the material for plastic bottles but notes that even larger savings in energy and GHGs could be achieved through the replacement of PET with PEF in textiles and films.

The replacement of non-renewable feedstocks such as oil with renewable feedstocks such as biomass is increasingly considered to be of great importance. As the industrial production of PET uses non-renewable petrochemicals as its feedstock, the replacement of this material with one that can be produced from renewable material is desirable. As previously discussed, PEF can be produced

from sugars derived from biomass, with corn being particularly popular due to its high fructose content. While production of plastics from plant-derived feedstocks is certainly attractive, valid concerns over the land-area required to achieve this exist. These concerns state that changing the function of farmland from food production to the production of feedstocks for the plastics industry could weaken our global food chain. The alternative to this would be to expand the land-area that is farmed upon, which would require converting uncultivated land into farmland, which raises concerns over ecological damage. It is also possible to generate glucose from less valuable biomass such as cellulose[143]. Cellulose can be sourced from waste products of several industries such as farming or the paper industry; as such it represents a greener source than glucose as a feedstock for PEF production.

As PEF products look and feel very similar to ones made out of PET it is inevitable that they will be disposed of in the same manner. This presents a problem of contamination during the recycling PET and PEF processes. Whilst it has been shown that low levels of PEF contamination in PET waste stream have little to no impact on the physical properties of the recycled PET, with an increased percentage of PEF in the waste stream this could present a significant issue. Currently separation techniques used in the mechanical recycling methods for PET waste such as float separation are incapable of distinguishing between PET and PEF. It has been proposed that near-infrared (NIR) sorting equipment be introduced to PET recycling plants as this technology can identify the differences between the two materials. Ultimately, the total replacement of PET with PEF and other bio-derived polymers will solve the problem of co-contamination, but this problem is important to consider while both PET and PEF present in the products we use.

5.1.6 Comparison of Physical Properties of Poly(ethylene) Terephthalate and Poly(ethylene) Furanoate

Another clear advantage of PEF over PET is that it outperforms its petrochemically-derived competitor in several important physical properties (see Table 4).

Table 4: Comparison of physical and barrier properties of PET and PEF polymers.

Property	**PET**	**PEF**
Glass Transition Temperature (T_{ag})	75°C[144]	75-80°C[132]
Melting Point (T_m)	245-251°C[144]	210-215°C [132]
Young's Modulus	1.3 GPa[132]	2.0 GPa[132]
Yield Stress	64.6 MPa[145]	122.6 MPa[145]
CO_2 Permittivity	0.49[146]	0.026[146]
O_2 Permittivity	0.114[147]	0.0107[147]

The food packaging industry creates the largest demand for PET products globally. PET is clear, cheap and has favourable gas permittivity properties. For the packaging of carbonated drinks, the carbon dioxide permittivity of the material used to make the bottle is important. The lower the carbon dioxide permittivity of the bottle, the longer the drink stays carbonated. PEF makes an attractive drop-in replacement for PET bottles as it has an approximately 20-fold lower carbon dioxide permittivity value when compared to PET[146] (see Table 4). For the packaging of carbonated alcoholic beverages such as beer, the oxygen permittivity of the material becomes important. This is because exposure to oxygen can lead to the ethanol oxidising to acetic acid which results in a sour and unpleasant taste. PEF has an approximately 11-fold lower oxygen permittivity value than PET[147], making it an attractive alternative. The lager manufacturer Carlsberg, in partnership with the plastics manufacturer Avantium, has developed a new bottle consisting of PEF and wood fibres called the "Green Fibre Bottle", which may reach commercial markets in the next few years. Carlsberg claims that the production of a "Green Fibre Bottle" produces up to 80% less emissions than the production of a single-use glass bottle. If this is true (some reasonable doubt is probably appropriate here due to the 'up to' qualification), the replacement of glass bottles with PEF-based ones would result in a significant reduction in GHGs produced[148]. The "Green Fibre Bottle" is opaque and brown (most likely due to the wood fibre component). Whilst this is not objectively a negative characteristic, consumers are used to transparent bottles for beer. This may present a challenge for the marketing of the "Green Fibre Bottle" as a replacement to glass bottles.

Despite its favourable gas permittivity properties, PEF has a tendency to be more brittle than PET under similar conditions. This is shown by its higher Young's modulus[132] and yield stress[145], where the higher the Young's modulus, the higher stress the material can endure before deforming but the lower strain it can endure before shattering. As brittleness is an undesirable property in the majority

of products (including food and drink packaging) this problem needs to be addressed. It is not stated in the article, but it is possible that the introduction of wood fibres into the “Green Fibre Bottle” reduces the brittleness of the final product.

Regarding the thermal properties of PEF, it has a suitably high melting point and suitably low glass transition temperature to allow manufacturing processes currently used for the production of PET products to work for PET.

Studies into the mechanical properties of PEF and PET have concluded that the main factor contributing to their differences in physical properties is the chain conformations favoured by the two polymers. PET forms fairly linear chains, directed by the linearity of its TA monomer[132] with an angle between its C1-C2 and C5-C6 bonds of 180°. More complex behaviour is observed in PEF chains. This is due to the non-linearity of its FDCA monomer with an angle between its C1-C2 and C5-C6 bonds of 129°. Additionally, the lower aromaticity of PEF’s furan ring compared to PET’s phenyl ring lowers the barrier for rotation of FDCA’s C1-C2 and C5-C6 bonds. These two effects combined result in the energetically favoured coiled-helix (with $anti^{FDCA}$- $gauche^{EG}$ domains) being favoured for PEF chains. This coiled-helix chain structure is thought to be the dominating factor in driving PEF’s differing physical properties when compared to PET[132].

5.2 Methodology

5.2.1 Methods used for the exploration of the reaction profile of *Is*-PETase degrading a PEF analogue

The methodology used in this section is described in full in the methodology section of Chapter 2. Our methods were not altered as it was our intention to compare the results from this section with our results obtained in Chapter 2.

5.3 Results and Discussion

5.3.1 Atom Naming Conventions

In the following sections, key atoms in the QM region will be referenced repeatedly. For the sake of conciseness, the following naming convention will be used to reference these key atoms.

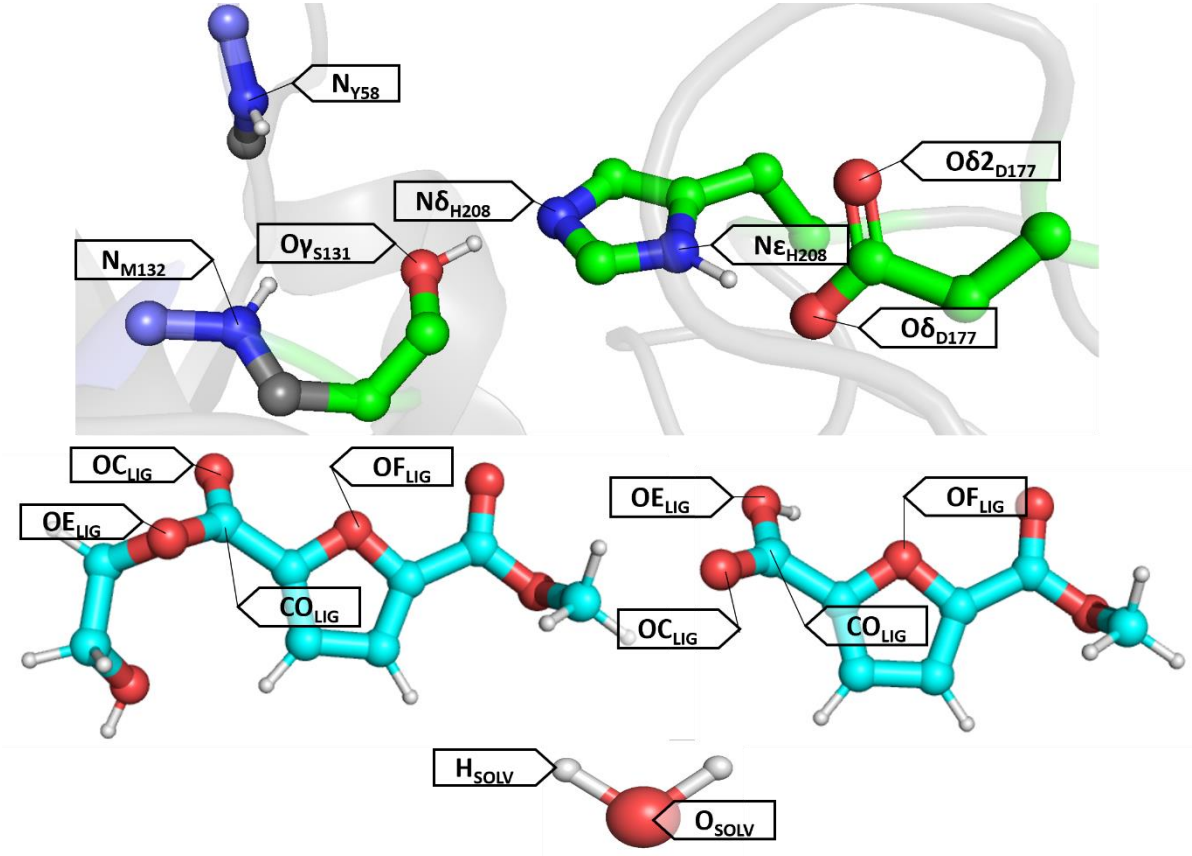


Figure 88: Atom naming conventions used in this chapter. Key atoms in Is-PETase's active site (above), in the substrate mimic HEMF (middle left), the product HMF (middle right) and nucleophilic water molecule in NS2 (below) have been assigned names.

5.3.2 Rationale for a Mechanistic Exploration of Is-PETase degrading PEF

In our work discussed in Chapter 3: we have extensively explored the reaction mechanism of wild-type *Is*-PETase degrading its 'natural' substrate PET. *Is*-PETase has been reported to have promiscuous activity towards the bio-derived polymer PEF. Since PEF is expected to be employed as a drop-in replacement for PET in the coming years, enzymes capable of its biodegradation will be of great importance to this process. As such, a deep understanding of the chemical mechanism of *Is*-PETase degrading PEF would be valuable. Using this rationale, we have applied the same methodology used in Chapter 2 to explore the reaction mechanism of *Is*-PETase catalysing the hydrolysis of a PEF substrate mimic.

5.3.3 Selection of Appropriate Substrate Mimic for PEF

Our calculations of the degradation mechanism of *Is*-PETase upon PEF are to be compared to our previously calculated mechanism of *Is*-PETase upon PET (discusses in Chapter 2). For this reason, it is important to select a substrate mimic for PEF that is as similar as possible to the substrate mimic chosen for PET, 2-hydroxyethyl methylterephthalate (HEMT). By replacing the phenyl ring in HEMT with a furanoate ring, the substrate mimic for PEF 2-hydroxyethyl methylfuranoate (HEMF) was produced (see Figure 89).

The reduced symmetry of the furan ring in HEMF compared to the phenyl ring in HEMT introduced an additional complication: the orientation of the two ester bonds in HEMF relative to the furan ring's oxygen. DFT studies have shown that PEF favours the *syn-syn* conformation (see Figure 89) for its ester bonds relative to oxygens in its furanoate rings[149]. Our docking procedure scans through all possible ligand conformations (ignoring bonds to protons, which are removed). The assumption at this point was made that this conformational search during the docking procedure would select the energetically favoured conformation during the docking procedure.

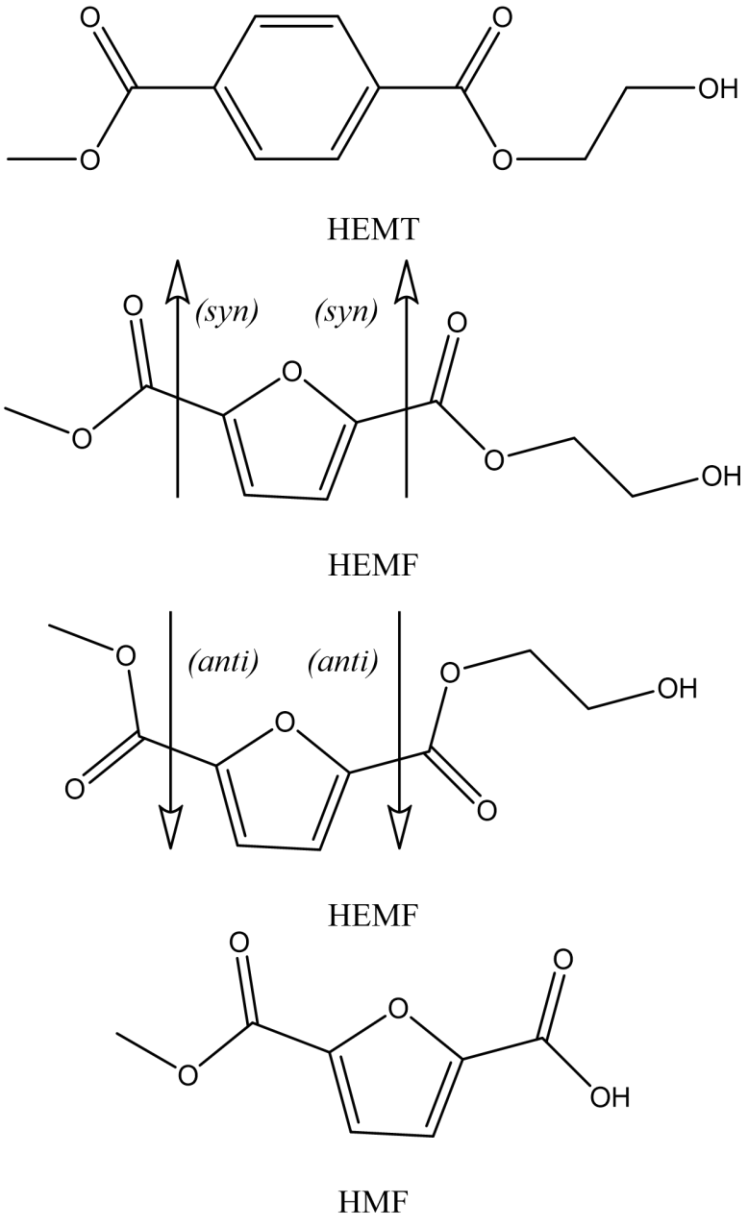


Figure 89: The chemical structure of the substrate mimics HEMT (top) and HEMF (second and third). The syn-syn (second) and anti-anti (bottom) conformations for the ester groups relative to the furanoate oxygen in HEMF are shown.

We will be generating reaction intermediates and transition states for the NS2 nucleophilic reaction starting from its product state (the rationale for which is described in section 3.3.3). For these calculations, the product molecule 2-hydroxyethyl furanoic acid (HMF) was used instead of HEMF.

5.3.4 Molecular docking simulations produce binding modes for HEMF and HMF in Is-PETase active site

Molecular docking simulations were performed for the small molecules HEMF and HMF into the active site of *Is*-PETase using methods described in section 3.2.1.2. For both small molecules, the main criterion for selecting binding poses was the proximity of the atom OC_{LIG} to the oxyanion hole

(composed of N_{Y58} and N_{M132}). Using this criterion, two binding modes per small molecule were presented as suitable candidates for selection.

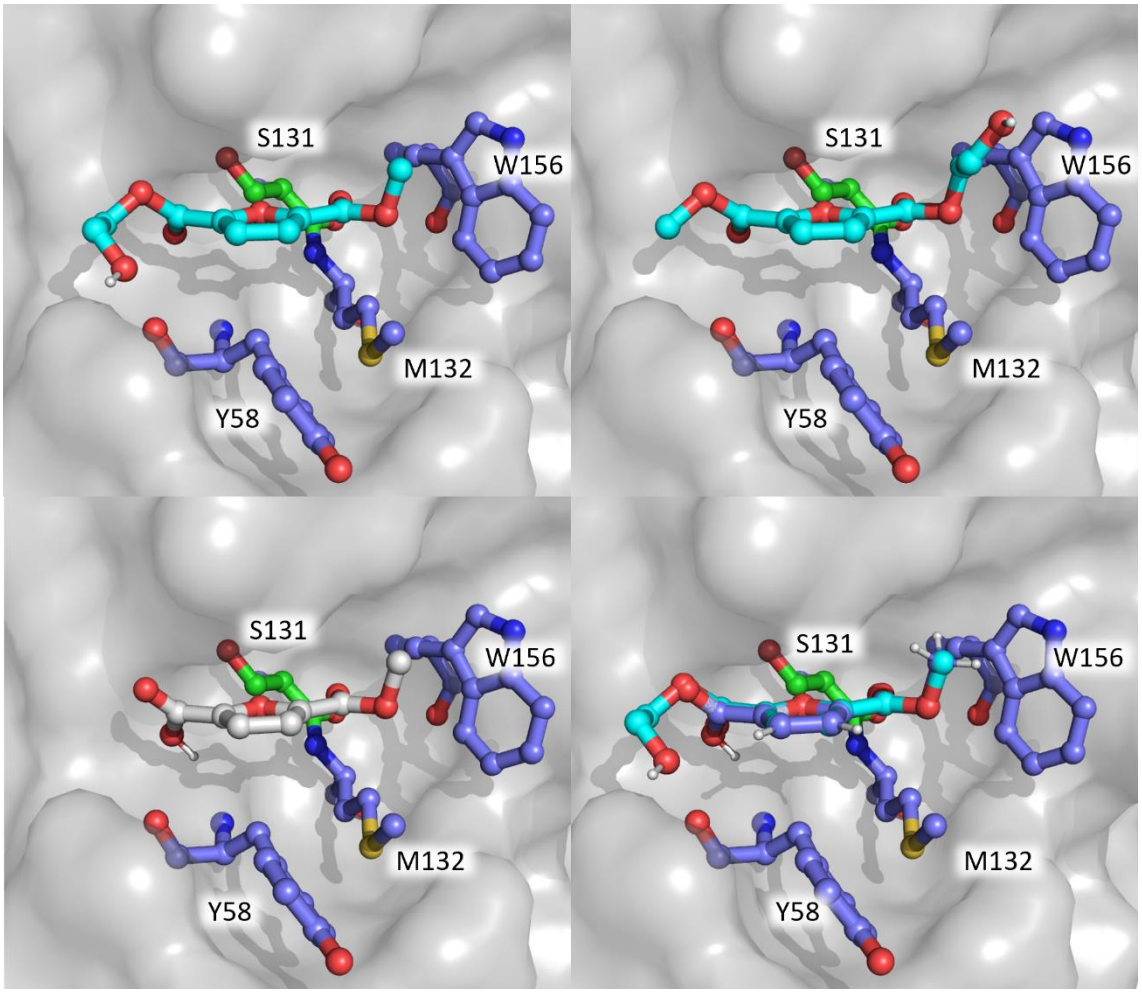


Figure 90: Binding poses for $BM1_{HEMF}$ (top left) and $BM2_{HEMF}$ (top right). Binding pose for $BM1_{HMF}$ (bottom left). An overlay of $BM1_{HEMF}$ and $BM1_{HMF}$ (bottom right).

The binding mode with the lowest energy for the ligand HEMF ($BM1_{HEMF}$) (see Figure 90) had longer than ideal oxyanion hole distances, with OC_{LIG}-N_{Y58} and OC_{LIG}-N_{M132} distances of 3.4 Å and 5.5 Å respectively. The ideal distance for these interactions would be between 2.6 Å and 3.0 Å, representing a strong hydrogen bonding interaction. The binding mode with the second lowest energy for HEMF ($BM2_{HEMF}$) had similarly long oxyanion hole distances, with OC_{LIG}-N_{Y58} and OC_{LIG}-N_{M132} distances of 5.3 Å and 3.4 Å respectively. In $BM1_{HEMF}$ the atom OC_{LIG} is closer to N_{Y58}, while in BM2 the atom OC_{LIG} is closer to N_{M132}. This change in preference between oxyanion hole nitrogen atoms is due to an approximately 180° rotation in the HEMF molecule from $BM1_{HEMF}$ to $BM2_{HEMF}$. To decide which binding pose is more appropriate to select for further study, we compared it to the binding pose previously selected for HEMT (the selection process of which is detailed in section 3.2.1.2). In the binding pose selected for HEMT (BM_{HEMT}), the phenyl ring of the substrate is located

to the side of the active site containing the residues Tyr58 and Trp156. It is thought that these aromatic rings in these residues form strong π-π interactions with the phenyl rings of PET and contribute greatly to its binding. In BM_{HEMT}, the HEMT molecule is oriented such that CO_{LIG} is pointing deeper into the binding pocket than the phenyl ring. In $BM1_{HEMF}$ the atom CO_{LIG} of HEMF is pointing deeper into the binding pocket than the molecule's furan ring, while in $BM2_{HEMF}$ the furan ring is pointing deeper into the binding pocket than the atom CO_{LIG}. To aid comparison with our study of the degradation of HEMT via *Is*-PETase, the binding mode $BM1_{HEMF}$ was selected for further study.

For the docking of the product-mimic HMF to the active site of *Is*-PETase, the binding mode with the lowest energy ($BM1_{HMF}$) was selected for further study (see Figure 90). $BM1_{HMF}$ was selected using the same criteria as for the substrate-mimic HEMF. The HMF molecule in binding mode $BM1_{HMf}$ overlays almost perfectly with the HEMT molecule in $BM1_{HEMF}$, with the obvious exception that it has no EG unit. One key difference between $BM1_{HMF}$ and $BM1_{HEMF}$ is that the OE_{LIG} oxygen is oriented towards the oxyanion hole instead of OC_{LIG}, resulting in an *anti*-conformation of HMF's carboxylic acid group relative to its furan ring (see Figure 89).

As discussed above, the docking poses $BM1_{HEMF}$ and $BM1_{HMF}$ selected for further study do not have suitable geometries to act as reactant and product states along the reaction pathway of substitution reactions NS1 and NS2 respectively. At this point in our workflow, this is not a problem as the proceeding molecular dynamics simulations (discussed in section 3.2.2) and restrained QM/MM geometry optimisations (discussed in section 3.2.3) will modify geometry of the enzyme-substrate complex significantly. The goal of molecular docking simulations in this case was to simply create a "good-enough" enzyme-substrate complex as a starting geometry for further calculations.

5.3.5 Molecular dynamics simulations on HEMF/Is-PETase and HMF/Is-PETase enzyme-substrate complexes

Starting with the enzyme-substrate complexes generated via molecular docking simulations for HEMF and HMF ligands (discussed previously in section 5.3.3), MD simulations were performed. The methodology for these simulations was identical to the MD simulations run previously for HEMT and HMT ligands (discussed in section 5.3.5). This resulted in the generation of one 1000 ps MD simulation per ligand. Snapshots were taken from these two MD simulations at 100 ps intervals. These snapshots were then used as starting geometries for our QM/MM calculations to generate reaction profiles for NS1 and NS2 reactions for the degradation of HEMF via *Is*-PETase.

For the MD simulation run on the HEMF/*Is*-PETase enzyme-substrate complex (used to generate starting geometries for the NS1 reaction), the snapshot taken at 100 ps did not generate an ideal geometry to be used to generate intermediate RS (see Figure 91). The carbonyl bound to the methyl

group of HEMF was situated in the oxyanion hole instead of the desired carbonyl bound to the EG subunit of HEMF. Whilst this binding mode had reasonable hydrogen bonding distances between the undesired carbonyl and the oxyanion hole nitrogen (2.9 Å and 3.7 Å for N_{Y58} and N_{M132} respectively), the desired carbonyl (containing CO_{LIG} and OC_{LIG}) would have to be moved into the oxyanion hole with the use of restraints in later QM/MM geometry optimisations.

For the MD simulation run on the HMF/*Is*-PETase enzyme-substrate complex (used to generate starting geometries for the NS2 reaction), the snapshot taken at 100 ps did not share the same problem as described above (see Figure 91). The oxyanion hole distances for HMF were longer than ideal, but this could be modified using subsequent restrained QM/MM geometry optimisations. As discussed on page 212, the docking of HMF into *Is*-PETase resulted in an *anti* conformation of the carboxylic acid group containing CO_{LIG} and OC_{LIG}. In this snapshot, a rotation in this group has resulted in the more energetically stable *syn* conformation (see Figure 89) of the carboxyl group.

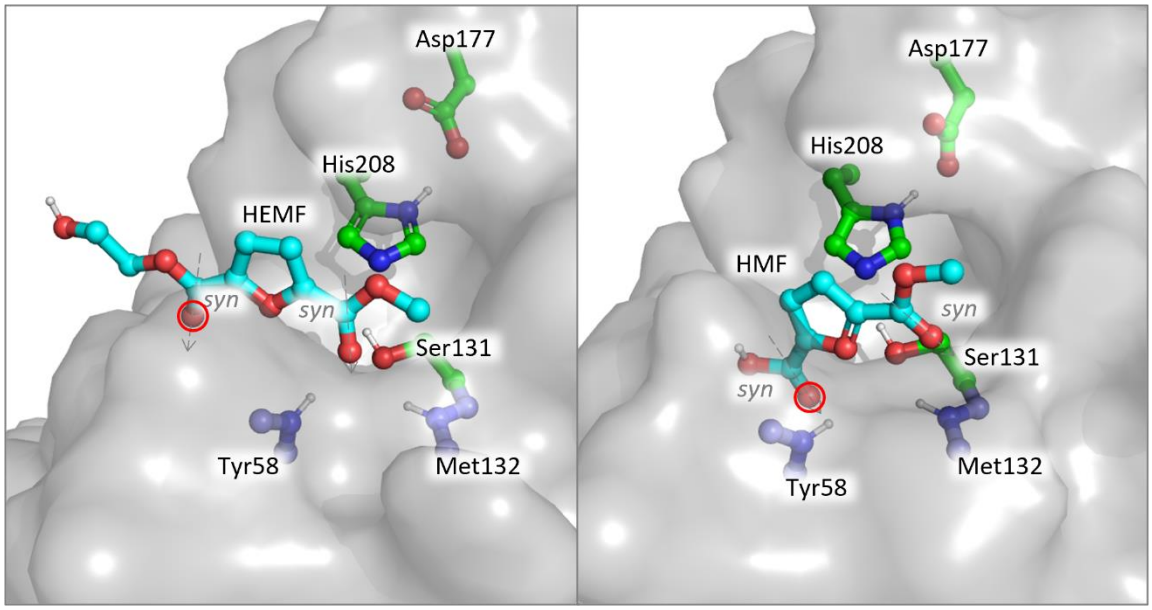


Figure 91: The enzyme-ligand complexes of HEMF (left) and HMF (right) at 100 ps snapshots taken from MD simulations. The location of atom OC_{LIG} is circled in red in both images. The orientation (syn in each case) of each carbonyl has been annotated.

5.3.6 Generation of Intermediates and Transition States

5.3.6.1 Generation of RS intermediate

From the snapshot taken from the MD simulation of HEMF bound to *Is*-PETase, the next step was to generate intermediates along the reaction pathway of NS1. The reaction intermediate RS was generated first: Starting from the snapshot geometry a restrained QM/MM geometry optimisation was performed. This optimisation used a pair of restraints with a r_0 value of 2.9 Å and a force constant of 0.1 between OC_{LIG}-N_{Y58} and between OC_{LIG}-N_{M132} in order to pull the desired carbonyl

into the oxyanion hole. Usually in our restrained geometry optimisations, a force constant of 3.0 is used for all restraints as a default. In this case a far lower force constant was required as the initial distances were far larger than the desired r_0 values. A higher value force constant would result in undesired bond breaking events due to the high energy introduced by the restraints. An additional restraint was placed between CO_{LIG} and OC_{LIG} with a r_0 value of 1.3 Å and a force constant of 3.0; this was added to prevent the CO_{LIG}-OC_{LIG} bond breaking due to the two other restraints. The resultant geometry from this restrained geometry optimisation created a large solvent cavity above the HEMF ligand. To re-fill this cavity with solvent molecules, a further MD simulation was performed using position restraints upon all non-solvent atoms. Following this 'frozen' MD simulation, a further unrestrained QM/MM geometry optimisation was performed to produce the geometry for the intermediate RS.

5.3.6.2 Generation of I1 intermediate

Starting from the geometry previously generated for intermediate RS, a restrained geometry optimisation was performed with a restraint placed between OG_{S131} and CO_{LIG} with a r_0 value of 1.3 Å and a force constant of 1.0. This restraint served to create a new bond between OG_{S131} and CO_{LIG}, mimicking the nucleophilic attack of Ser131. An additional restraint was applied between $HG1_{S131}$ and $NE2_{H208}$ with a r_0 value of 1 Å and a force constant of 1.0. This restraint assisted the proton transfer of $HG1_{S131}$ to $NE2_{H208}$. This was followed by an unrestrained geometry optimisation to produce the geometry for intermediate I1.

5.3.6.3 Generation of I2 intermediate

Starting with the geometry previously generated for intermediate I1, a restrained geometry optimisation was performed with a restraint placed between $HG1_{S131}$ (which is bound to NE2H208 in intermediate I1) and OE_{LIG} with a r_0 value of 1 Å and a force constant of 1.0. This restraint has the effect of protonating OE_{LIG}, causing the reformation of the double bond between CO_{LIG} and OC_{LIG} and the scission of the bond between CO_{LIG} and OE_{LIG}. This was followed by an unrestrained geometry optimisation to produce the geometry for intermediate I2.

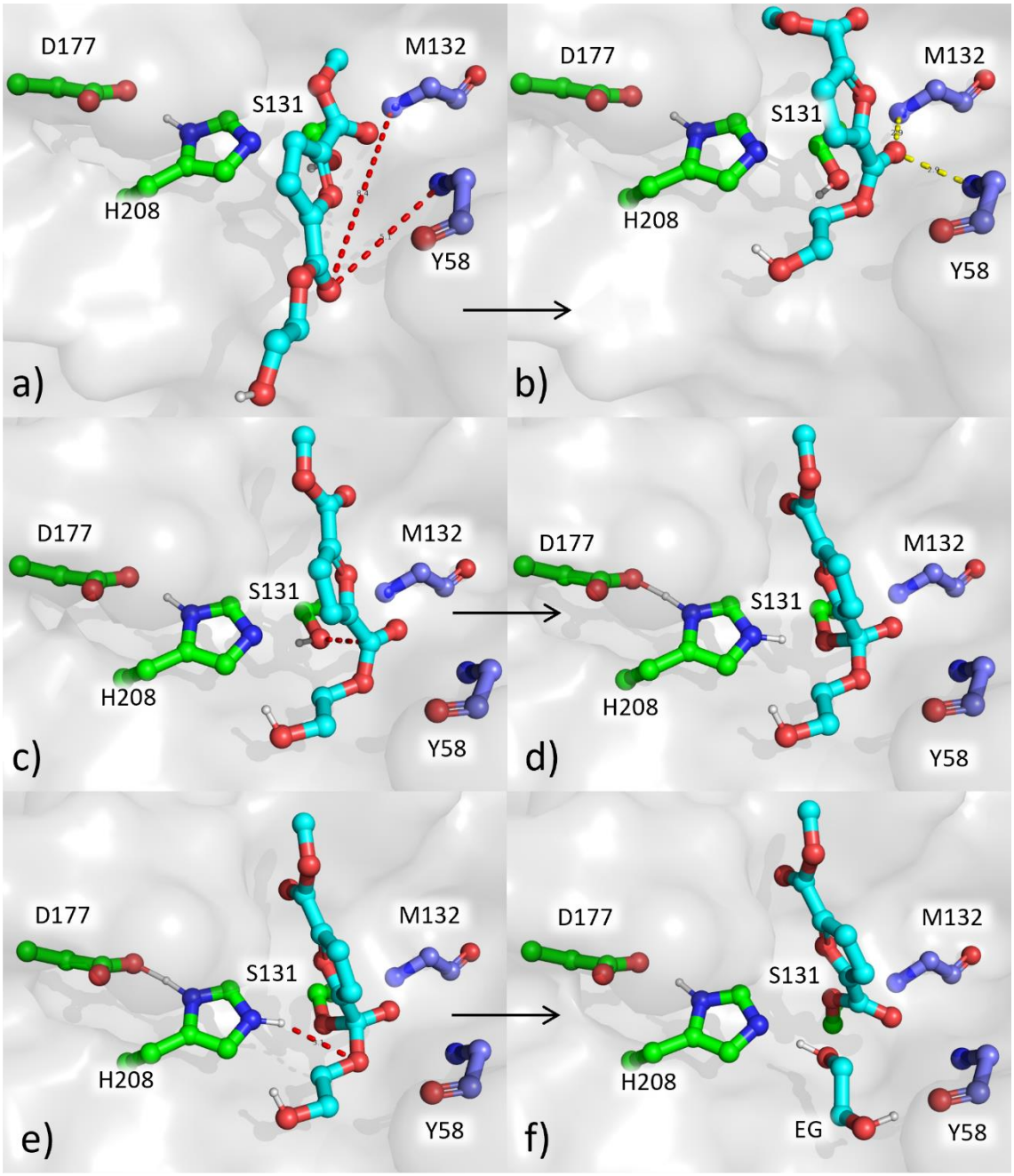


Figure 92: Scheme showing production of intermediates for the degradation of PEF via Is-PETase. Restraints in images a, c and e are shown as red dotted lines. Intermediate RS (b) is produced from the output of a 'frozen' MD simulation (a). Intermediate I1 (d) is produced from RS (c). Intermediate I2 (f) is produced from I1 (e). In this scheme, only key atoms are shown for each residue.

5.3.6.4 Generation of PS intermediate

Starting from the snapshot taken from the MD simulation of HMF bound to *Is*-PETase (discussed previously on page 214) a restrained QM/MM geometry optimisation was performed. A pair of restraints were applied between the pairs of atoms OC_{LIG}-N_{Y58} and the pair of atoms OC_{LIG}-N_{M132} with r_0 values of 2.7 Å and force constants of 0.01. These restraints pulled the desired carbonyl (consisting of CO_{LIG} and OC_{LIG}) further into the oxyanion hole, establishing stronger hydrogen bonding

interactions. Although these restraints did not produce a significant solvent hole above the HMF ligand, following our now established protocol the system was re-solvated using a 'frozen' MD simulation (for further details see page 215). A further unrestrained QM/MM geometry optimisation was then performed upon the resulting geometry from this MD simulation to produce the geometry for the intermediate PS.

5.3.6.5 Generation of I4 intermediate

Starting from the geometry previously generated for PS, a restrained geometry optimisation was performed with a restraint between OG_{S131} and CO_{LIG} with a r_0 distance of 1.3 Å and a force constant of 1.0. This restraint had the effect of creating a new covalent bond between OG_{S131} and CO_{LIG}. Along the forwards trajectory of the reaction profile for NS2 (i.e., from I4 to PS), this bond will be broken in concert with the reformation of the double bond between CO_{LIG} and OC_{LIG}. A further unrestrained geometry optimisation was performed to generate the geometry for intermediate I4.

5.3.6.6 Generation of I3 intermediate

Starting from the geometry previously generated for I4, a restrained geometry optimisation was performed with a restraint between $HG1_{S131}$ (which is bound to $NE2_{H208}$ in I4) and OE_{LIG} with a r_0 distance of 1 Å and a force constant of 1.0. This restraint had the effect of triggering the cleavage of the OE_{LIG}-OC_{LIG} bond, resulting a loss of a water molecule as a leaving group. Along the forwards directory of the reaction profile of NS2 (i.e., from I3 to I4), the water molecule will be deprotonated by His208 and nucleophilically attack CO_{LIG}. A further unrestrained geometry optimisation was performed to generate the geometry for intermediate I3.

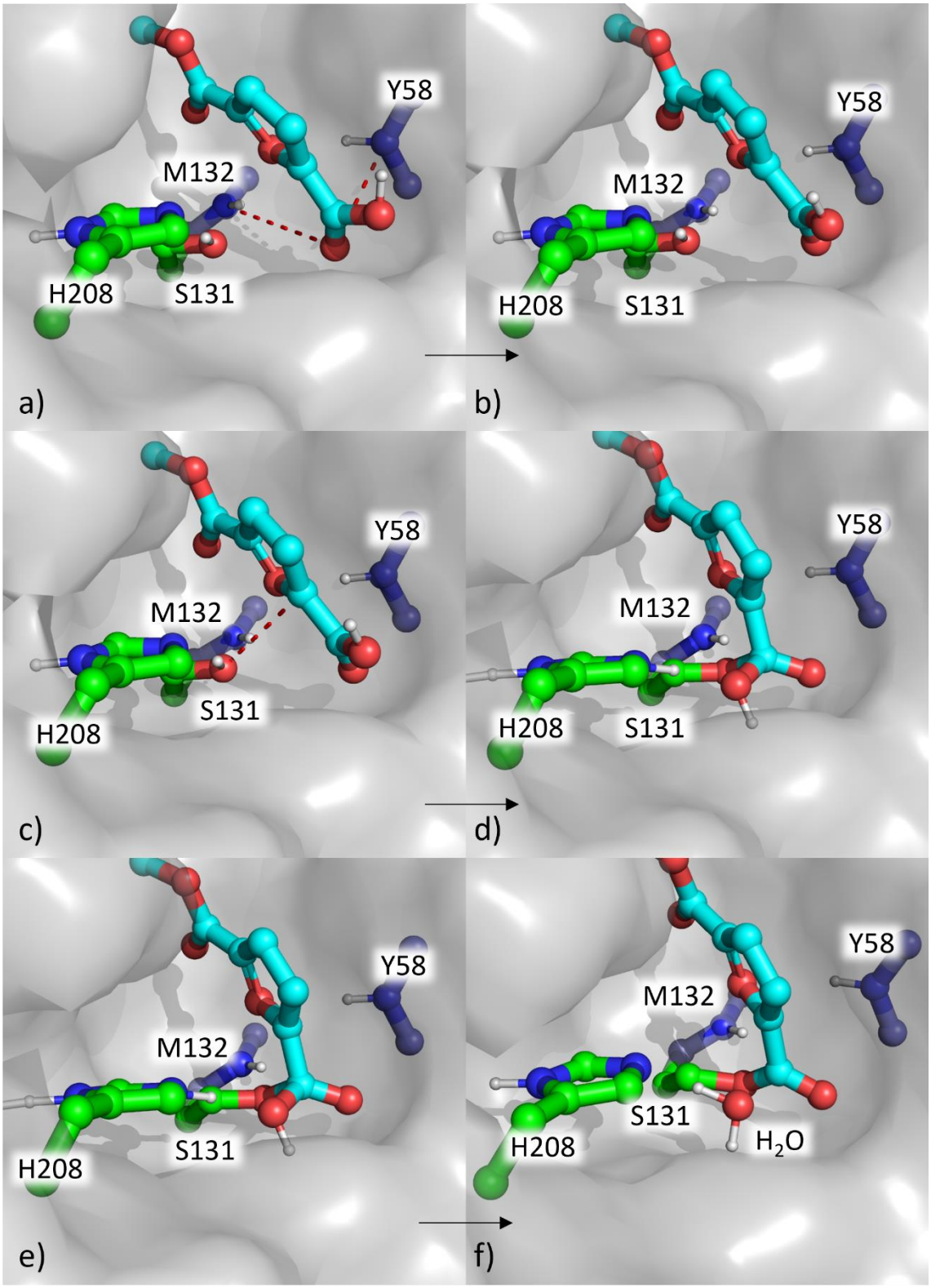


Figure 93: Scheme showing production of intermediates for the degradation of PEF via Is-PETase. Restraints in images a, c and e are shown as red dotted lines. Intermediate PS(b) is produced from the output of a 'frozen' MD simulation (a). Intermediate I4 (d) is produced from PS (c). Intermediate I3 (f) is produced from I4 (e).

5.3.6.7 Generation of transition states

Transition states along the reaction profiles NS1 and NS2 for the degradation of HEMF via *Is*-PETase were generated using the Nudged-Elastic Band method (NEB) (detailed in section 3.2.3.4). The transition state Nuc1 was generated using a NEB calculation with intermediates RS and I1 as start and end points respectively. The transition state LgR1 was generated using a NEB calculation with intermediates I1 and I2 as start and end points respectively. The transition state Nuc2 was generated using a NEB calculation with intermediates I3 and I4 as start and end points respectively. The transition state LgR2 was generated using a NEB calculation with intermediates I4 and PS as start and end points respectively. Since an energy minimum associated with the product state PS was found, the presence of the aPS state (discussed in section 3.3.6) was not investigated.

5.3.7 Evaluation of the energy profile of Is-PETase degrading HEMF substrate

Single-point energy calculations were performed using the geometries for intermediates and transition states generated (discussed in the previous section 5.3.6). The DFT method (DSD-PBEP86), basis set (md-def2-TZVP) and MD forcefield (CHARMM27) used for these energy calculations were retained from our single-point calculations for the degradation of HEMT via *Is*-PETase.

Following the same methods we used in our exploration of WT-PETase degrading HEMT (a PET substrate mimic), for the reaction profile of WT-PETase degrading HEMF we calculated energies of reaction intermediates relative to the first intermediate in each reaction. For the reaction NS1, energies were calculated relative to the RS geometry at the start of each reaction profile. For the reaction NS2, energies were calculated relative to the I3 geometry at the start of each reaction profile.

As we have only generated one reaction profile for the degradation of HEMF via WT-PETase, the discussion in section 3.3.4 regarding what approach (“lowest energy profile”, arithmetic averaging or Boltzmann-weighted averaging) was a more appropriate method for selecting the rate-determining step becomes moot. We only have one reaction profile per reaction, so we can only select the rate-determining steps from these.

For the NS1 reaction, an unexpectedly high energy reaction profile was generated (see Figure 94). Energies for intermediates I1 and I2 and transition states Nuc1 and LgR1 have relative energy values in excess of 75 kJ mol^{-1}. Across this single energy profile, the rate-determining step is the LgR1 reaction step with an activation energy of 111.3 kJ mol^{-1}. This activation barrier exceeds all activation barriers calculated for the degradation of HEMT via *Is*-PETase.

For the reaction NS2, a similarly high energy barrier was obtained (see Figure 94). The highest activation barrier along this NS2 reaction profile belongs to the LgR2 reaction step, with an activation energy of 88.2 kJ mol^{-1}. Again, this activation barrier exceeds all activation barriers calculated for the degradation of HEMT via *Is*-PETase.

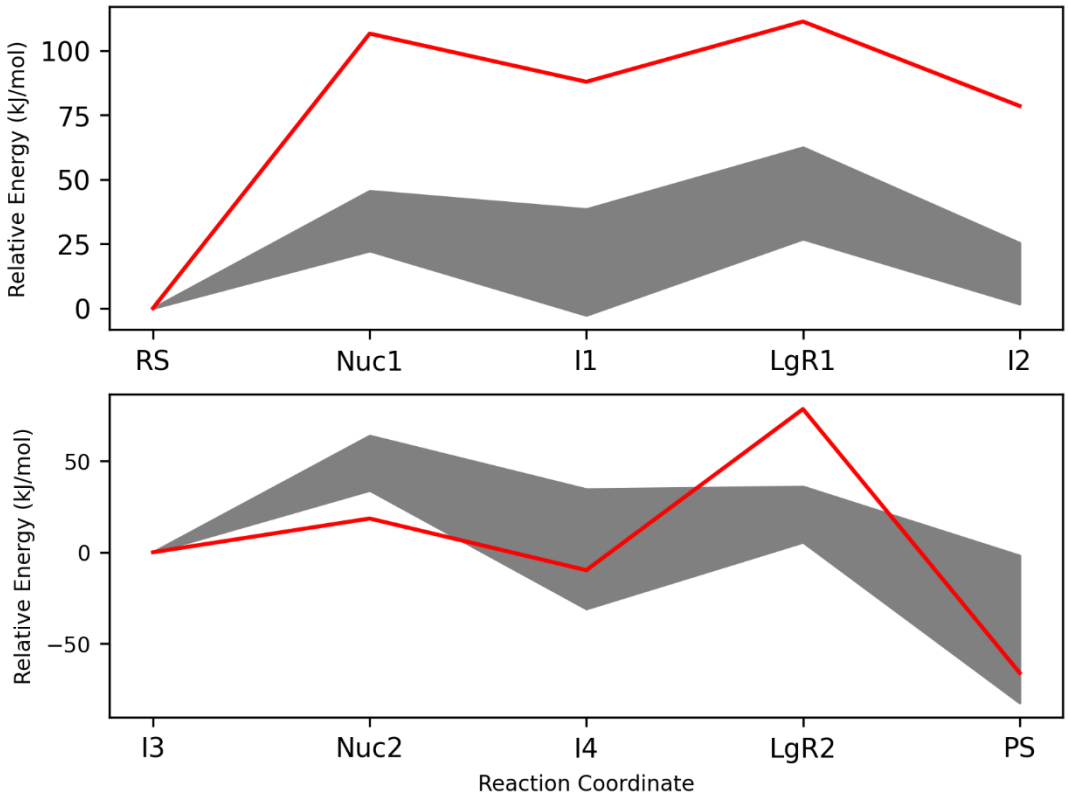


Figure 94: Energy profiles for the degradation of a PEF substrate mimic (shown as red lines) compared to the range in energies calculated for the degradation of an equivalent PET mimic (grey fill). The energy profile for the NS1 reaction (top) and the NS2 reaction (bottom) are shown separately

It is our opinion that the high energy reaction profile we have generated for the reaction NS1 does not reflect a true low energy reaction profile for the degradation of HEMF via *Is*-PETase. Our NS2 reaction profile shares more similarities with the NS2 energy profiles we previously generated for the degradation of HEMT via *Is*-PETase. Despite this, its highest activation barrier is far higher than we would have expected.

As it has been reported that *Is*-PETase is capable of degrading PEF at similar rates to PET, we would expect similar energy barriers for each process's rate-determining step. As we discussed in section 3.4, our studies of the degradation of PET via *Is*-PETase led us to the conclusion that the rate-determining step for this process does not occur during the chemical steps of the reaction. Instead, we suggest that the rate-determining step could be the physical product-release step, based on the

formation of a highly stable state formed via proton transfer between the product and Ser131 (discussed in section 3.3.6). As such, we cannot use a comparison with our reaction profiles generated for the degradation of PET via *Is*-PETase to validate our reaction profiles generated for the degradation of PEF via *Is*-PETase. We can however state that the activation barrier of 111.1 kJ mol^{-1} we have calculated for this reaction is far higher than we would expect for an enzymatic reaction that proceeds at a moderate rate at room temperature[150].

Ideally, we would have generated several reaction profiles for the degradation of PEF via *Is*-PETase. This would allow us to test whether the high energy reaction profile we generated for reactions NS1 and NS2 were anomalously high. We expect that lower energy reaction profiles would be generated, resulting in greatly reduced energy barriers for both reactions.

Using the two reaction profiles we have generated we could conclude that the rate-determining step for the degradation of PEF via *Is*-PETase is the LgR1 step with an activation energy of 111.3 kJ mol^{-1}. It is our opinion however that the true activation energy for this process is far lower than this, and that the rate-determining step is not necessarily LgR1. To verify this opinion, several more reaction profiles for NS1 and NS2 must be generated.

5.3.8 Geometry analysis for the reaction profile generated for NS1

In this section, we will discuss geometrical features of our NS1 reaction profile. We will pay particular attention to features that may have caused the high energy barriers associated with this reaction profile.

In this (and the following) section, we will refer to features associated with reaction profiles for the degradation of both PEF and PET via wild-type *Is*-PETase. For brevity and clarity, we will use superscript PEF or PET to identify which reaction is being referred to (e.g., reaction step $Nuc1^{PEF}$ or $O\gamma_{S131}{}^{PET}$)

5.3.8.1 Geometry analysis for reaction step $Nuc1^{PEF}$

In the energy profile we have generated for the reaction $NS1^{PEF}$, the energy of intermediate $I1^{PEF}$ is 87.9 kJ mol^{-1} relative to the preceding intermediate RS^{PEF}. As discussed previously, this energy difference is far larger than any $I1^{PET}$ energy we have generated. It is likely that the high energy of $I1^{PEF}$ is responsible for the high energy barrier for the preceding reaction step $Nuc1^{PEF}$.

During the analysis we performed for $Nuc1^{PET}$ (see section 3.3.5.1.1), we found that the most important factor to afford low energy barriers was the extent to which OC_{LIG} formed hydrogen bonding interactions with the oxyanion hole (expressed as measurements between OC_{LIG}-N_{Y58} and OC_{LIG}-N_{M132}) in RS^{PET}. We found that shorter oxyanion hole distances (particularly the OC_{LIG}-N_{M132}

distance) resulted in Nuc1^{PET} transition states with lower activation energies. If we apply this relationship to our Nuc1^{PEF} reaction step, we should expect that a low energy barrier would be calculated for Nuc1^{PEF}. This is because strong hydrogen bonds are made between OC_{LIG} and the oxyanion hole in the geometry RS^{PEF} (with OC_{LIG}-N_{Y58} and OC_{LIG}-N_{M132} distances of 2.7 Å and 3.1 Å respectively).

We must therefore look to other geometrical differences between RS^{PEF} and I1^{PEF} to rationalise the high energy barrier for Nuc1^{PEF}. An overlay of these two geometries (see Figure 95) reveals that a significant rotation in the aromatic ring of HEMF has occurred during the Nuc1^{PEF} reaction step. In RS^{PEF} the reactive carbonyl and the aromatic ring are effectively coplanar, with an angle between them of only 14.0°. By contrast, this angle increases dramatically to 118.7° in I1^{PEF}. In RS^{PEF}, this coplanarity will allow for delocalisation of electrons between the furan ring and the reactive carbonyl, which may be providing a stabilising effect to RS^{PEF}. This delocalisation must be broken to form I1^{PEF}, which may contribute to the high energy barrier associated with Nuc1^{PEF}. In RS^{PET} and I1^{PET} geometries along our lowest energy profile, this angle between the reactive carbonyl and the adjacent phenyl ring does not change as dramatically (the angle is 55.9° in RS^{PET} and 93.4° in I1^{PET}). It is worth noting that in this RS^{PET} geometry, the reactive carbonyl and the phenyl ring are not coplanar. This means that the delocalisation between the carbonyl and the aromatic ring does not need to be broken during the Nuc1 reaction step. We therefore tentatively propose that this breaking of the delocalisation between the reactive a carbonyl and the adjacent furan ring in HEMF is responsible for our high activation energy calculated for Nuc1^{PEF}. Generation of further reaction profiles could be used to confirm or deny this theory. For the generation of further reaction profiles for NS1^{PEF}, it may be necessary to identify a stable intermediate in between RS^{PEF} and I1^{PEF} in which the delocalisation has already been broken, but the new covalent bond between $O\gamma_{S131}$ and CO_{LIG} is not formed.

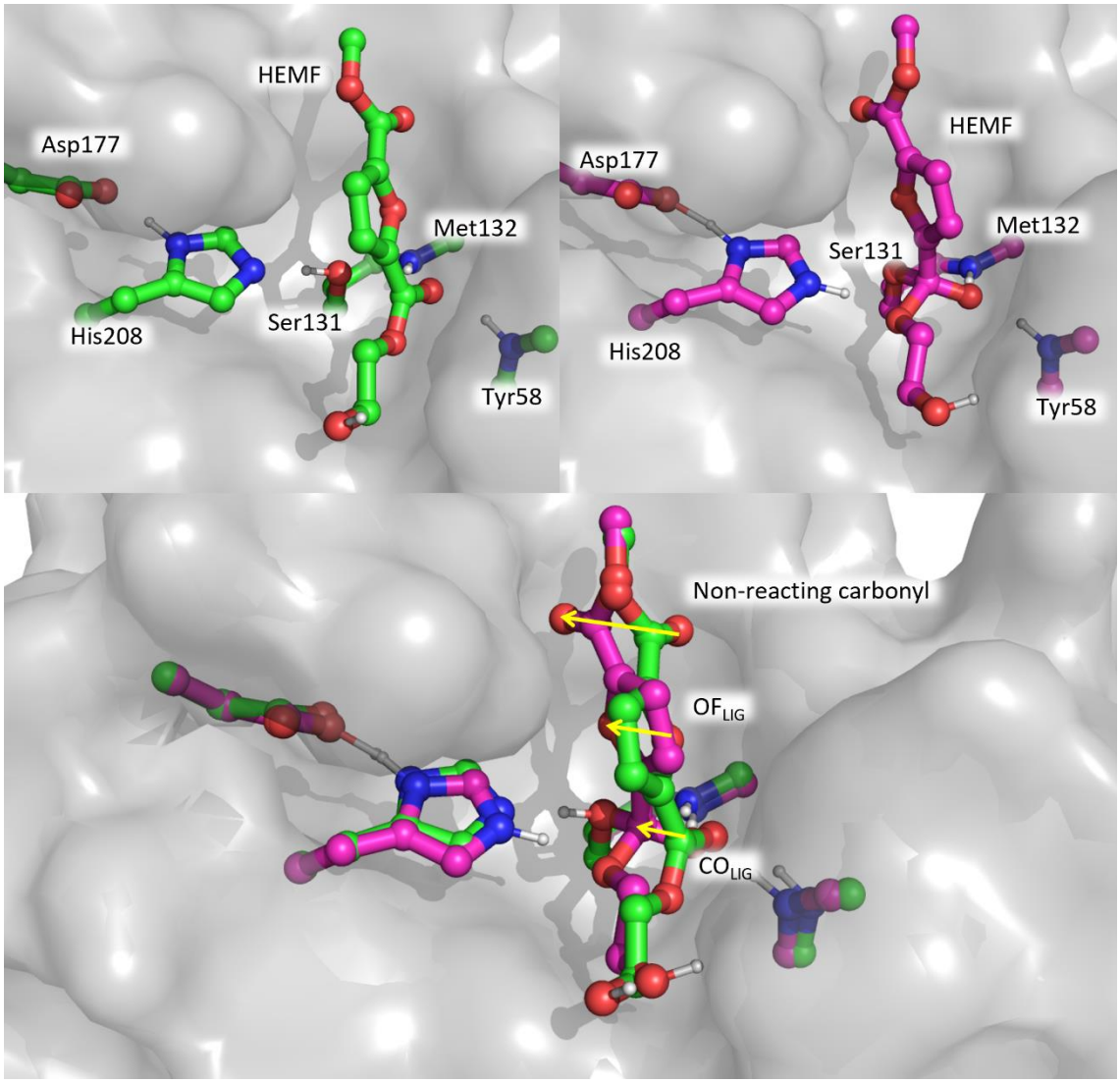


Figure 95: Geometries for RS^{PEF} (top left) and I1^{PEF} (top right). An overlay of the two geometries is shown below, with RS^{PEF} shown in magenta and I1^{PEF} shown in green. Yellow arrows are used to show regions where the two geometries differ greatly.

5.3.8.2 Geometry analysis for reaction step LgR1^{PEF}

The activation energy for the reaction step LgR1^{PEF} was calculated to be 111.3 kJ mol^{-1}. This high energy barrier was calculated as the difference between the energies of LgR1^{PEF} and RS^{PEF}. If we consider the LgR1 reaction step in isolation (i.e., we only consider the energies of I1^{PEF}, LgR1^{PEF}, and I2^{PEF}), the reaction step's energy profile becomes more similar to the energy profiles we generated for LgR1^{PET}. In fact, the energy difference between LgR1^{PEF} and I1^{PEF} is 23.5 kJ mol^{-1}, which is slightly lower in energy than the lowest energy barrier we found for LgR1^{PET} (28.9 kJ mol^{-1}).

5.3.9 Geometry analysis for the reaction profile generated for NS2

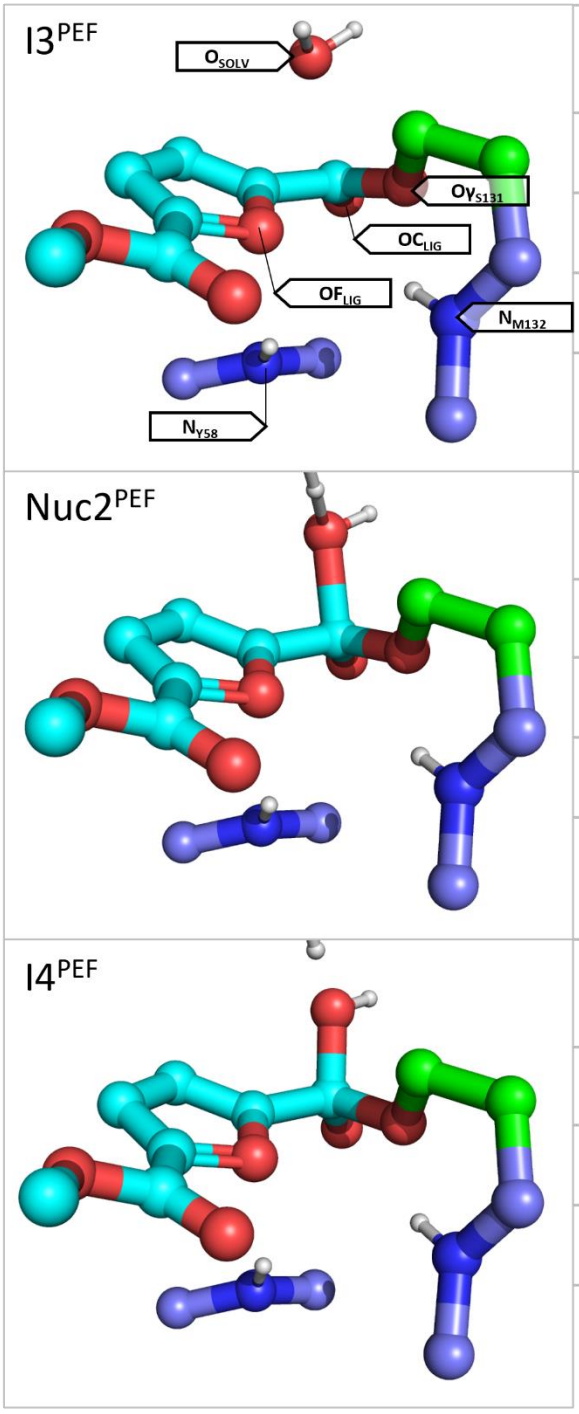
In this section, we will discuss the interesting features of our NS2 reaction profile for the degradation of PEF via *Is*-PETase. We will compare key measurements within the active site in this reaction profile with observations we made in our previous study of the degradation of PET via *Is*-PETase.

5.3.9.1 Geometry analysis for reaction step $Nuc2^{PEF}$

A brief glance at the geometry of $I3^{PEF}$ reveals a remarkable difference to the geometries generated for $I3^{PET}$. In our set of $I3^{PET}$ geometries, the atom $OC_{LIG}{}^{PET}$ forms hydrogen bonds with the atoms in the oxyanion hole $N_{Y58}{}^{PET}$ and $N_{M132}{}^{PET}$. For $I3^{PEF}$ however, $OC_{LIG}{}^{PEF}$ is proximal, but oriented away from the oxyanion hole, forming no hydrogen bonds with $N_{Y58}{}^{PEF}$ or $N_{M132}{}^{PEF}$.

We would expect from our previous analysis of geometries for $Nuc2^{PET}$ that the activation energy for $Nuc2^{PEF}$ would be extremely high. This is because for $Nuc2^{PET}$, the binding of $OC_{LIG}{}^{PET}$ to the oxyanion hole in the preceding $I3^{PET}$ geometry was found to be the most important factor in generating low energy barriers for $Nuc2^{PET}$. It is therefore surprising that the energy barrier for $Nuc2^{PEF}$ is very low (18.5 kJ mol^{-1}). To rationalise this low energy barrier, we searched the active site in $I3^{PEF}$, $Nuc2^{PEF}$, and $I4^{PEF}$ for stabilising interactions that were not present in their PET counterparts.

In geometries for $I3^{PEF}$, $Nuc2^{PEF}$, and $I4^{PEF}$, $N_{Y58}{}^{PEF}$ seems to be forming a weak hydrogen bond with $OF_{LIG}{}^{PEF}$, rather than with $OC_{LIG}{}^{PEF}$. Although N_{Y58} is closer to $OC_{LIG}{}^{PEF}$ than $OF_{LIG}{}^{PEF}$ (with distances of 3.3 Å and 3.5 Å respectively), the hydrogen bond angle between N_{Y58}, $OC_{LIG}{}^{PEF}$ and the intervening proton is too acute to facilitate hydrogen bonding (see Figure 96). For these geometries, $N_{M132}{}^{PEF}$ seems to be forming a strong hydrogen bond with $O\gamma_{S131}{}^{PEF}$ with hydrogen bonding distances of 2.7 Å to 2.9 Å. It is worth noting that the distances and angles measured between $N_{M132}{}^{PEF}$ and $OF_{LIG}{}^{PEF}$ may indicate some hydrogen bonding between these two atoms (see Figure 96).



	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.2 122.3	2.9 137.4
OF_{LIG}	3.5 149.7	3.3 130.7
OC_{LIG}	3.3 86.9	4.6 163

	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.2 127.4	2.8 140.3
OF_{LIG}	3.5 155.4	3.4 127.8
OC_{LIG}	3.3 91.7	4.6 167

	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.2 132.4	2.7 143.3
OF_{LIG}	3.5 160	3.5 126.8
OC_{LIG}	3.3 95.4	4.6 169.8

Figure 96: Geometries for $I3^{PEF}$ (top), $Nuc2^{PEF}$ (middle), and $I4^{PEF}$ (bottom) with associated measurements concerning hydrogen bonding with the oxyanion hole. Note that angles have been measured between the two atoms specified and the intervening proton.

The differences in oxyanion hole bonding in geometries involved in the $Nuc2^{PEF}$ step with those in the $Nuc2^{PET}$ step indicates that the chemical structure of the PEF substrate is having an effect on the overall mechanism of *Is*-PETase. By forming hydrogen bonds with the oxyanion hole, $OF_{LIG}{}^{PEF}$ displaces OC_{LIG} from the oxyanion hole.

A broader search of the active site in geometries $I3^{PEF}$, $Nuc2^{PEF}$, and $I4^{PEF}$ revealed that $OC_{LIG}{}^{PEF}$ forms a novel hydrogen bond with the backbone nitrogen atom of $Glu57^{PEF}$ (with a hydrogen bonding distance of 2.9 Å) (see Figure 97). $Glu57^{PEF}$ however is not included in our QM region, so this hydrogen bond will not be treated at the QM level. It would be interesting to generate reaction profiles for the degradation of PEF via *Is*-PETase these geometries with the inclusion of this residue in the QM region. In our brief investigation into the effects of including or excluding hydrogen bond donors to $OC_{LIG}{}^{PET}$ from the QM region (see section 3.2.3.2), we found that the inclusion of these atoms resulted in increased stabilisation of the tetrahedral intermediates $I1^{PET}$ and $I4^{PET}$ as well as a reduction in all energy barriers. We therefore expect that the inclusion of the backbone of Glu57 into our QM region would reduce the barrier height of $Nuc2^{PEF}$.

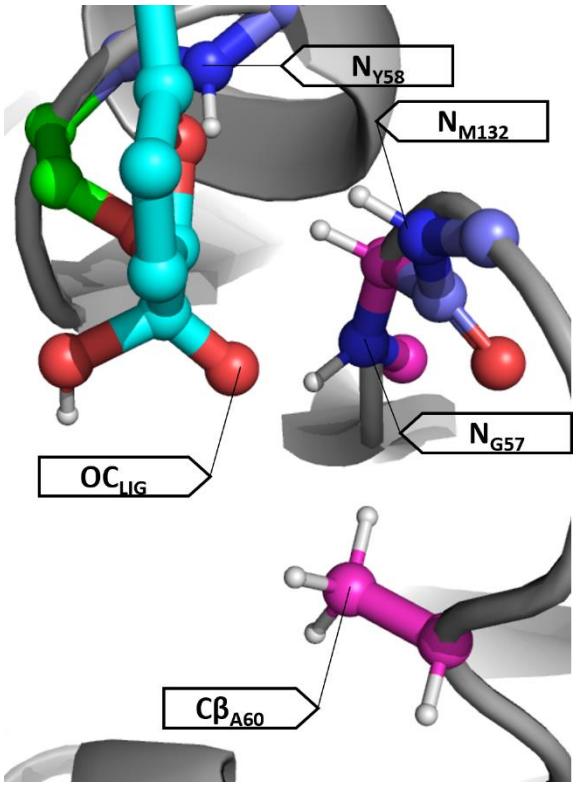


Figure 97: Interactions between $OC_{LIG}{}^{PEF}$ with $N_{G57}{}^{PEF}$ in the $I3^{PEF}$ geometry. The location of the residue Ala60 is also shown, which we propose as a target for mutagenesis.

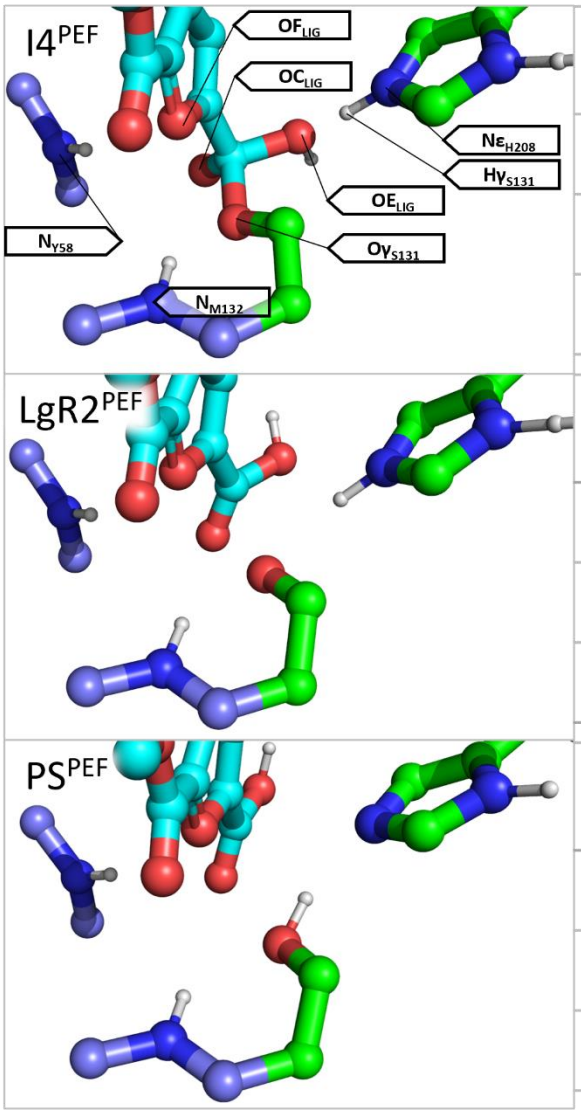
Proximal to $OC_{LIG}{}^{PEF}$ in these geometries, is the sidechain of the residue $Ala60^{PEF}$. This residue is in an ideal location to hydrogen bond with $OC_{LIG}{}^{PEF}$, but due to its non-polar sidechain is unable to form hydrogen bonding interactions with $OC_{LIG}{}^{PEF}$. The mutation A60S has the potential to introduce a new hydrogen bonding interaction with $OC_{LIG}{}^{PEF}$ throughout the Nuc2 step and could further stabilise the negative charge that builds up on $OC_{LIG}{}^{PEF}$ in intermediate $I4^{PEF}$. This mutation has the potential to relocate the oxyanion hole from N_{Y58} and N_{M132} to N_{G57} and $O\gamma_{S60}$.

5.3.9.2 Geometry analysis for the reaction step LgR2^{PEF}

In contrast to the preceding reaction step Nuc2^{PEF}, LgR2^{PEF} has a rather high calculated activation energy. Analysis of the hydrogen bonding within the oxyanion hole in the geometries I4^{PEF}, LgR2^{PEF} and PS^{PEF} revealed a similar pattern to the one discussed previously in section 5.3.9.1: The atom OC_{LIG} is oriented such that it cannot form hydrogen bonding interactions with atoms N_{Y58} or N_{M132} (see Figure 98) . Instead, the oxyanion hole atom $N_{Y58}{}^{PEF}$ seems to be forming a hydrogen bond with $OF_{LIG}{}^{PEF}$. The oxyanion hole atom $N_{M132}{}^{PEF}$ seems to be forming a hydrogen bonding interaction with $O\gamma_{S131}{}^{PEF}$. Again, this is in contrast with our geometries generated for the degradation of HEMT via *Is*-PETase. As discussed in section 5.3.9.1, $OC_{LIG}{}^{PEF}$ seems to be forming a hydrogen bond with the backbone nitrogen atom of the residue Gly57.

The most striking difference between the geometry I4^{PEF} and its corresponding geometries I4^{PET}, is the orientation of $O\gamma_{S131}$ relative to the sidechain of His208 (see Figure 98). In our I4^{PET} geometries, the sidechain of Ser131 is rotated such that there is a short distance between $O\gamma_{S131}{}^{PET}$ and $N\varepsilon_{H208}{}^{PET}$ which facilitates the proton transfer between these atoms during the LgR2 reaction step with a fairly low energy barrier. In I4^{PEF}, the distance between $O\gamma_{S131}{}^{PEF}$ and $N\varepsilon_{H208}{}^{PEF}$ is longer, and the orientation of intervening proton $H\gamma_{S131}{}^{PEF}$ is such that a hydrogen bond is formed between $N\varepsilon_{H208}{}^{PEF}$ and $OE_{LIG}{}^{PEF}$. In the transition state geometry LgR2^{PEF} $H\gamma_{S131}{}^{PEF}$ is still bound to $N\varepsilon_{H208}{}^{PET}$. This is most likely to be the cause of the high energy barrier associated with the LgR2^{PEF} step.

We propose that a lower lying reaction barrier could be generated by finding a stable intermediate with the sidechain of Ser131^{PEF} rotated such that a stronger hydrogen bond between $O\gamma_{S131}{}^{PEF}$ and $N\varepsilon_{H208}{}^{PEF}$ could form. Using this intermediate instead of our I3^{PEF} geometry to generate the transition state LgR2^{PEF} may be able to find a transition state with a lower activation energy.



	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.2 132.4	2.7 143.3
OF_{LIG}	3.5 160	3.5 126.8
OC_{LIG}	3.3 95.4	4.6 169.8

	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.7 143.6	2.8 140.2
OF_{LIG}	3.2 149.4	3.87 142
OC_{LIG}	3.3 99.2	4.52 158.7

	N_{Y58} (Distance Å, Angle °)	N_{M132} (Distance Å, Angle °)
$O\gamma_{S131}$	4.9 149.5	3 133.2
OF_{LIG}	3.2 153.7	4.3 151.3
OC_{LIG}	3.3 100.8	5.3 165

Figure 98: Geometries for $I4^{PEF}$ (top), $LgR2^{PEF}$ (middle), and PS^{PEF} (bottom) with associated measurements concerning hydrogen bonding with the oxyanion hole. Note that angles have been measured between the two atoms specified and the intervening proton.

5.4 Conclusions

In this section, we have detailed our initial attempts to explore the first energy profiles for the reactions associated with the degradation of HEMF (a substrate mimic for PEF) via the enzyme *Is*-PETase. We have applied the same methodology to this problem as we used to successfully generate energy profiles for the degradation of HEMT (a substrate mimic of PET). The energy profiles we generated for the reactions $NS1^{PEF}$ and $NS2^{PEF}$ have energy barriers that far exceed their counterparts $NS1^{PET}$ and $NS2^{PET}$. As discussed in section 5.3.7, we believe that reaction profiles of $NS1^{PEF}$ and $NS2^{PEF}$ that we have generated so far are insufficient to create energy profiles that reflect the true reaction pathways the enzyme *Is*-PETase utilises to degrade PEF. Our results can therefore be seen as preliminary and not yet conclusive. Our work however, particularly our development of our workflow pipeline and scripts, provides a solid foundation for further study.

We believe that our general method for generating intermediates and transition states along reaction profiles is sound. However, we suggest that a different restraint scheme be applied to generate the specific intermediates along the reaction profile for the degradation of PEF via *Is*-PETase. We also suggest that several reaction profiles be generated for both $NS1^{PEF}$ and $NS2^{PEF}$ as our study on the degradation of PET via *Is*-PETase showed the importance of multiple profiles in obtaining low energy rate-determining steps. We also suggest that the size of the QM region be increased to include the backbone nitrogen atom of Gly57, as our studies provide evidence for a hydrogen bonding interaction between this atom and OC_{LIG}^{PEF}.

Conclusions and Outlook

In this thesis, we have presented several computational approaches to explore the catalytic mechanism and the thermostability of the enzyme *Is*-PETase and some of its variants.

Our main project (detailed in Chapter 2: has been to calculate reaction profiles for the degradation of PET via *Is*-PETase using QM/MM methods. We have created these reaction profiles along the reaction coordinate of the canonical α/β hydrolase reaction mechanism (of which *Is*-PETase is a member[17]). For practicality, we have separated this reaction mechanism into two separate, sequential, nucleophilic substitution reactions, NS1 and NS2 (see Figure 99).

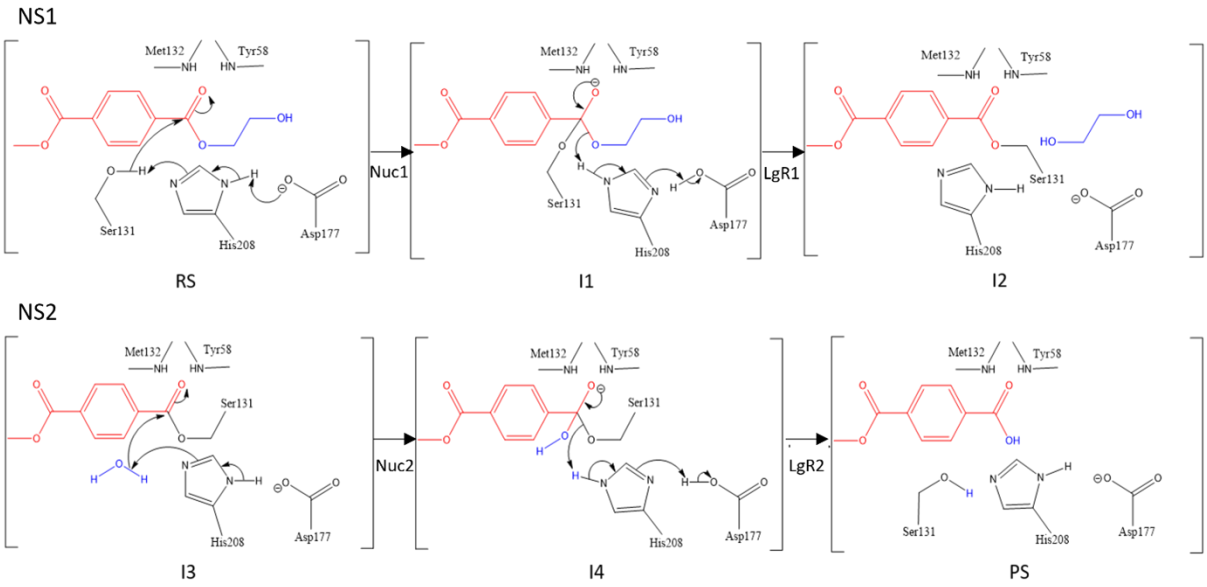


Figure 99: The proposed catalytic mechanism of for the degradation of PET via Is-PETase, separated into two nucleophilic substitution reactions NS1 (above) and NS2 (below).

In this project, we have generated several reaction profiles that provide strong evidence that *Is*-PETase proceeds via the canonical α/β hydrolase reaction mechanism with fairly low activation barriers. Our results have identified the reaction steps Nuc2 and LgR2 (see Figure 99) as dual rate determining steps, with activation barriers of 34.1 kJ mol^{-1} and 35.6 kJ mol^{-1} respectively (see Figure 100). These low activation barriers are not reflected in the slow catalytic rate of *Is*-PETase which suggests that a physical process such as substrate binding or product dissociation is the true rate-limiting step for the overall process. Our results also provide preliminary evidence for product inhibition of *Is*-PETase through the identification of a highly stable after-product state caused by a

proton transfer between the product and the active site of *Is*-PETase.

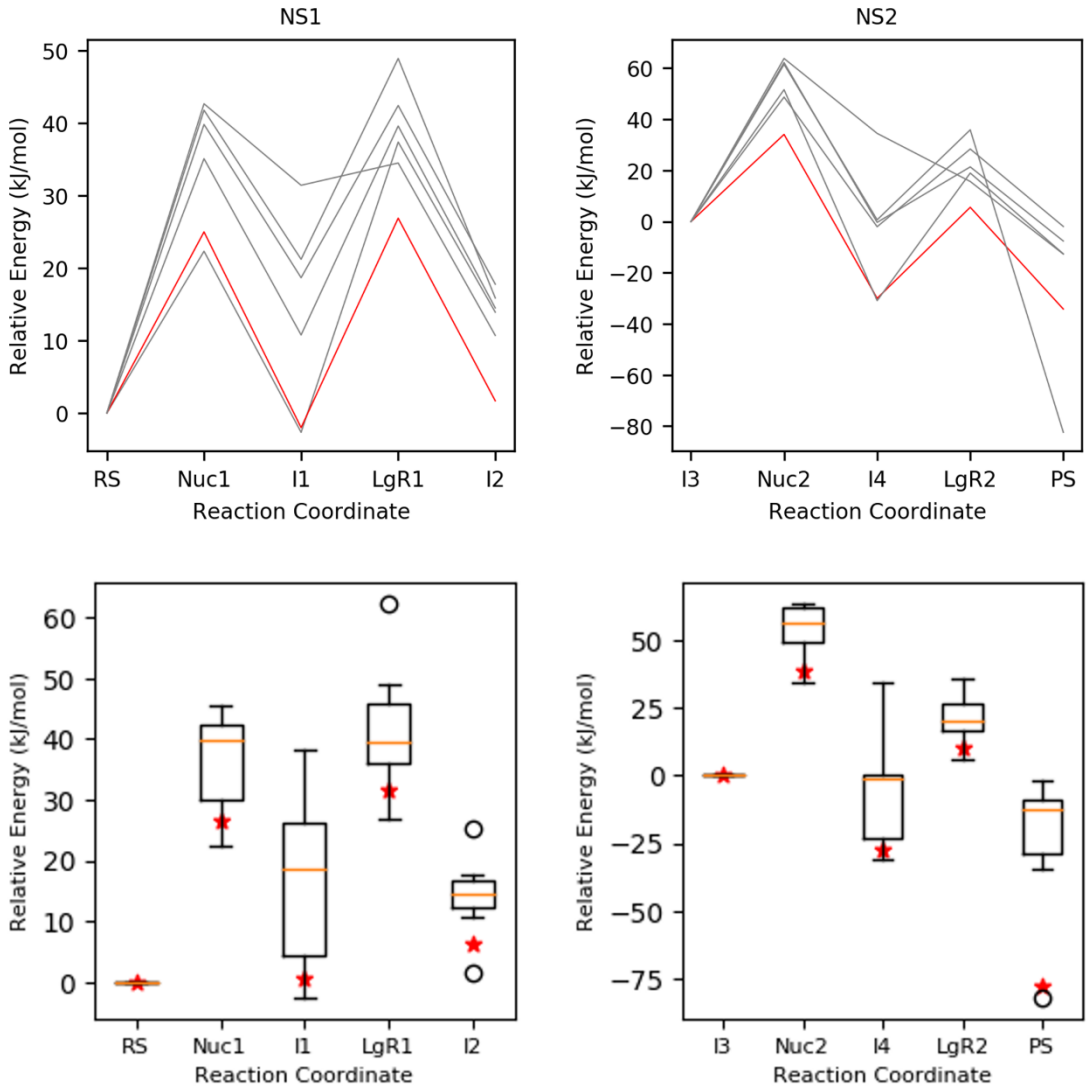


Figure 100: Energies for reaction profiles calculated for NS1 (top left) and NS2 (top right). The lowest energy profiles for NS1 and NS2 have been displayed in red. Box plots for the energies of intermediates and transition states along reaction profiles for NS1 (bottom left) and NS2 (bottom right). The arithmetic average for each geometry is displayed as an orange line, while the Boltzmann-weighted average is displayed as a red star.

Our results can be used to guide protein engineers to improve the catalytic efficiency of *Is*-PETase. Specifically, we advise against making mutations to *Is*-PETase in order to reduce the activation energy of the catalytic reaction (as this is already fairly low). Instead, we recommend that mutations be made to improve the enzyme's thermostability.

In our second project, discussed in Chapter 4:, we applied computational methods to explore the thermostability of some variants of *Is*-PETase. We have created our own variant (dubbed DISU-PETase) with the double mutation S163C S194C and a novel disulfide bond between the sidechains

of the two residues. Based on detailed analysis of MD trajectories, we expect that this novel disulfide bond will contribute to the thermostability of the enzyme.

We generated reaction profiles for the degradation of PET via DISU-PETase using the same QM/MM methods used for wild-type PETase in Chapter 3: Our results found that the reaction profile for DISU-PETase was not significantly different to that of wild-type PETase. As such, we conclude that the introduction of an additional disulfide bond through the S163C S194C mutations has no deleterious effects on the chemical reaction catalysed by *Is*-PETase. We therefore recommend that this mutant be made experimentally and assayed for its activity and thermostability.

In Chapter 4: we also used a series of MD simulations to investigate the roles of mutant residues in increasing the thermostability of a pair of highly thermostable *Is*-PETase variants from the literature. Our analysis of these MD simulations has provided evidence for the existence of multiple interactions present in these mutants that are not present in wild-type *Is*-PETase. We also show that this type of investigation can provide a far more detailed and dynamic picture of intramolecular bonding within a protein structure than X-ray crystallography alone.

In Chapter 5: we discuss our preliminary exploration of the catalytic mechanism of the degradation of an alternative substrate, poly(ethylene) furanoate, via wild-type *Is*-PETase. Our initial reaction profiles have unexpectedly high activation barriers, suggesting that these do not represent the true lowest energy pathway for this reaction mechanism. This work therefore is best understood as a solid foundation for further study.

Bibliography

1. Roman, L., Schuyler, Q., Wilcox, C. & Hardesty, B. D. Plastic pollution is killing marine megafauna, but how do we prioritize policies to reduce mortality? *Conserv. Lett.* **14**, e12781 (2021).
2. Frias, J. P. G. L. & Nash, R. Microplastics: Finding a consensus on the definition. *Mar. Pollut. Bull.* **138**, 145–147 (2019).
3. Prata, J. C., da Costa, J. P., Lopes, I., Duarte, A. C. & Rocha-Santos, T. Environmental exposure to microplastics: An overview on possible human health effects. *Sci. Total Environ.* **702**, 134455 (2020).
4. Jambeck, J. *et al.* the Ocean : the Ocean : *Mar. Pollut.* **347**, 768–771 (2015).
5. Cózar, A. *et al.* Plastic accumulation in the mediterranean sea. *PLoS ONE* **10**, 1–12 (2015).
6. Kershaw, P. *et al.* Plastic Debris in the Ocean. *UNEP Year b. 2011* 20–33 (2011).
7. Zhang, H. & Wen, Z. G. The consumption and recycling collection system of PET bottles: A case study of Beijing, China. *Waste Manag.* **34**, 987–998 (2014).
8. Department, S. R. Annual production of plastics worldwide from 1950 to 2021. *Statistica.com* (2023).
9. Chu, J. *et al.* A life-cycle perspective for analyzing carbon neutrality potential of polyethylene terephthalate (PET) plastics in China. *J. Clean. Prod.* **330**, 129872 (2022).
10. Kleeberg, I., Welzel, K., VandenHeuvel, J., Müller, R. J. & Deckwer, W. D. Characterization of a new extracellular hydrolase from Thermobifida fusca degrading aliphatic-aromatic copolyesters. *Biomacromolecules* **6**, 262–270 (2005).
11. Fich, E. A., Segerson, N. A. & Rose, J. K. C. The Plant Polyester Cutin: Biosynthesis, Structure, and Biological Roles. *Annu. Rev. Plant Biol.* **67**, 207–233 (2016).
12. Araújo, R. *et al.* Tailoring cutinase activity towards polyethylene terephthalate and polyamide 6,6 fibers. *J. Biotechnol.* **128**, 849–857 (2007).

13. Roth, C. *et al.* Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from Thermobifida fusca. *Appl. Microbiol. Biotechnol.* **98**, 7815–7823 (2014).

14. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92 (2002).

15. Sulaiman, S. *et al.* Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Appl. Environ. Microbiol.* **78**, 1556–1562 (2012).

16. Sulaiman, S., You, D. J., Kanaya, E., Koga, Y. & Kanaya, S. Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. *Biochemistry* **53**, 1858–1869 (2014).

17. Yoshida, S. *et al.* A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**, 1196–1199 (2016).

18. Austin, H. P. *et al.* Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc. Natl. Acad. Sci.* **115**, 4350–4357 (2018).

19. David, L. *et al.* The α/β hydrolase fold. *Protein Eng. Des. Sel.* **5**, 197–211 (1992).

20. Liu, B. *et al.* Protein Crystallography and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate) Hydrolase PETase from Ideonella sakaiensis. *ChemBioChem* 1471–1475 (2018) doi:10.1002/cbic.201800097.

21. Fecker, T. *et al.* Active Site Flexibility as a Hallmark for Efficient PET Degradation by I. sakaiensis PETase. *Biophys. J.* **114**, 1302–1312 (2018).

22. Taniguchi, I. *et al.* Biodegradation of PET: Current Status and Application Aspects. *ACS Catal.* **9**, 4089–4105 (2019).

23. Chen, C. C., Han, X., Ko, T. P., Liu, W. & Guo, R. T. Structural studies reveal the molecular mechanism of PETase. *FEBS J.* **285**, 3717–3723 (2018).

24. Matak, M. Y. & Moghaddam, M. E. The role of short-range Cys171-Cys178 disulfide bond in maintaining cutinase active site integrity: A molecular dynamics simulation. *Biochem. Biophys. Res. Commun.* **390**, 201–204 (2009).

25. Han, X. *et al.* Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* **8**, 2106 (2017).

26. Müller, R. J., Schrader, H., Profe, J., Dresler, K. & Deckwer, W. D. Enzymatic degradation of poly(ethylene terephthalate): Rapid hydrolyse using a hydrolase from T. fusca. *Macromol. Rapid Commun.* **26**, 1400–1405 (2005).

27. Son, H. F. *et al.* Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. *ACS Catal.* **9**, 3519–3526 (2019).

28. Kawai, F., Kawabata, T. & Oda, M. Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Appl. Microbiol. Biotechnol.* **103**, 4253–4268 (2019).

29. Boneta, S., Arafet, K. & Moliner, V. QM/MM Study of the Enzymatic Biodegradation Mechanism of Polyethylene Terephthalate. *J. Chem. Inf. Model.* **61**, 3041–3051 (2021).

30. Jerves, C., Neves, R. P. P., Ramos, M. J., Da Silva, S. & Fernandes, P. A. Reaction Mechanism of the PET Degrading Enzyme PETase Studied with DFT/MM Molecular Dynamics Simulations. *ACS Catal.* 11626–11638 (2021) doi:10.1021/acscatal.1c03700.

31. Feng, S. *et al.* IsPETase- andIsMHETase-Catalyzed Cascade Degradation Mechanism toward Polyethylene Terephthalate. *ACS Sustain. Chem. Eng.* **9**, 9823–9832 (2021).

32. Aboelnga, M. M. & Kalyaanamoorthy, S. QM/MM Investigation to Identify the Hallmarks of Superior PET Biodegradation Activity of PETase over Cutinase. *ACS Sustain. Chem. Eng.* (2022) doi:10.1021/acssuschemeng.2c04913.

33. Magalhães, R. P., Fernandes, H. S. & Sousa, S. F. The critical role of Asp206 stabilizing residues on the catalytic mechanism of the Ideonella sakaiensis PETase. *Catal. Sci. Technol.* **12**, 3474–3483 (2022).

34. Zhong-Johnson, E. Z. L., Voigt, C. A. & Sinskey, A. J. An absorbance method for analysis of enzymatic degradation kinetics of poly(ethylene terephthalate) films. *Sci. Rep.* **11**, (2021).

35. Ma, Y. *et al.* Enhanced Poly(ethylene terephthalate) Hydrolase Activity by Protein Engineering. *Engineering* **4**, 888–893 (2018).

36. Joo, S. *et al.* Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat. Commun.* **9**, (2018).

37. Lu, H. *et al.* Machine learning-aided engineering of hydrolases for PET depolymerization. **604**, (2022).

38. Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem. - Int. Ed.* **58**, 14420–14426 (2019).

39. Bell, E. L. *et al.* Directed evolution of an efficient and thermostable PET depolymerase. *Nat. Catal.* **5**, 673–681 (2022).

40. Tournier, V. *et al.* An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* **580**, 216–219 (2020).

41. Organisation for Economic Co-operation and Development. Improving plastics management: trends, policy responses, and the role of international co-operation and trade. *Environ. Policy Pap. No 12* 20 (2018).

42. Shen, L. & Worrell, E. *Plastic Recycling*. *Handbook of Recycling: State-of-the-art for Practitioners, Analysts, and Scientists* (Elsevier Inc., 2014). doi:10.1016/B978-0-12-396459-5.00013-1.

43. Tiso, T. *et al.* Towards bio-upcycling of polyethylene terephthalate. *Metab. Eng.* **66**, 167–178 (2021).

44. Hiraga, K., Taniguchi, I., Yoshida, S., Kimura, Y. & Oda, K. Biodegradation of waste PET . *EMBO Rep.* **20**, 1–5 (2019).

45. Textile Exchange. Preferred Fiber & Materials Market Report 2021. 1–118 (2021).

46. Alisch-Mark, M., Herrmann, A. & Zimmermann, W. Increase of the hydrophilicity of polyethylene terephthalate fibres by hydrolases from Thermomonospora fusca and Fusarium solani f. sp. pisi. *Biotechnol. Lett.* **28**, 681–685 (2006).

47. Getnet, M. & Chavan, R. Catalyzation of alkaline hydrolysis of polyester by oxidizing agents for surface modification. *Int. J. Sci. Basic Appl. Res.* **22**, 232–252 (2015).

48. Shrimpton-Phoenix, E., Mitchell, J. B. O. & Bühl, M. Computational Insights into the Catalytic Mechanism of Is-PETase: An Enzyme Capable of Degrading Poly(ethylene) Terephthalate. *Chem. - Eur. J.* **28**, (2022).

49. Frank Jensen. *Introduction to Computational Chemistry Volume 3*. (2017).

50. Hohenberg, P. & Kohn, W. Inhomogenous Electron Gas. *Phys. Rev.* **136**, 864–871 (1964).

51. Thomas, L. H. The calculation of atomic fields. *Math. Proc. Camb. Philos. Soc.* **23**, 542–548 (1927).

52. Fermi, E. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z Phys.* **48**, 73–79 (1928).

53. Dirac, P. A. M. Quantum Theory of the Electron. *Phys Rev* **77**, 253 (1932).

54. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, 1113–1138 (1965).

55. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

56. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999).

57. Kozuch, S. & Martin, J. M. L. DSD-PBEP86: In search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections. *Phys. Chem. Chem. Phys.* **13**, 20104–20107 (2011).

58. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the Damping Function in Dispersion CorrectedDensity Functional Theory. *J. Comput. Chem.* **32**, 174–182 (2012).

59. Henkelman, G., Uberuaga, B. P. & Jonsson, H. Nudged_Elastic Bands. *J. Chem. Phys.* **113**, 9901–9904 (2006).

60. Henkelman, G., Uberuaga, B. P. & Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).

61. Karplus, M. Development of Multiscale Models for Complex Chemical Systems. *Sci. Backgr. Nobel Prize Chem. 2013* 63–95 (2013) doi:10.1063/1.4732181.

62. Warshel, A. Computer Simulations of Biological Functions: From Enzymes to Molecular Machines. *Nobel Lect.* 159–183 (2013).

63. Jindal, G. & Warshel, A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J. Phys. Chem. B* **120**, 9913–9921 (2016).

64. Naray-Szabo, G., Warshel, A., Sussman, F. & Hwang, J. K. How do serine proteases really work? *Biochemistry* **28**, 3629–3637 (1989).

65. Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. - Int. Ed.* **48**, 1198–1229 (2009).

66. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).

67. Svensson, M. *et al.* ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)3)2 + H2 oxidative addition. *J. Phys. Chem.* **100**, 19357–19363 (1996).

68. Trott, O. & Olson, A. J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

69. Knott, B. C. *et al.* Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25476–25485 (2020).

70. Ventura, O. N., Coitiño, E. L., Lledós, A. & Berteán, J. AM1 study of hydrogen bonded complexes of water. *J. Mol. Struct. THEOCHEM* **187**, 55–68 (1989).

71. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB □ An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. (2019) doi:10.1021/acs.jctc.8b01176.

72. Hanke, F., Dyer, M. S., Björk, J. & Persson, M. Structure and stability of weakly chemisorbed ethene adsorbed on low-index Cu surfaces: Performance of density functionals with van der Waals interactions. *J. Phys. Condens. Matter* **24**, (2012).

73. Goerigk, L. *et al.* A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).

74. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).

75. Dennington, R., Keith, T. A. & Millam, J. M. GaussView, Version 6.1. Preprint at (2016).

76. MORRIS, G. M. *et al.* Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. *J. Comput. Chem.* **32**, 174–182 (2012).

77. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

78. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

79. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).

80. Lemkul, J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, 1–53 (2019).

81. Páll, S. & Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **184**, 2641–2650 (2013).

82. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

83. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

84. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

85. Zheng, J., Xu, X. & Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **128**, 295–305 (2011).

86. Sherwood, P. *et al.* QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. THEOCHEM* **632**, 1–28 (2003).

87. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model1. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).

88. Bickelhaupt, F. M. & Baerends, E. J. Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry. *Rev. Comput. Chem.* **15**, 1–86 (2000).

89. Kästner, J. *et al.* DL-FIND: An open-source geometry optimizer for atomistic simulations*. *J. Phys. Chem. A* **113**, 11856–11865 (2009).

90. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–616 (1998).

91. Cooper, A. M. & Kästner, J. Averaging Techniques for Reaction Barriers in QM/MM Simulations. *ChemPhysChem* **15**, 3264–3269 (2014).

92. Pirillo, V., Pollegioni, L. & Molla, G. Analytical methods for the investigation of enzyme-catalyzed degradation of polyethylene terephthalate. *FEBS J.* **288**, 4730–4745 (2021).

93. Silva, C. *et al.* Engineered Thermobifida fusca cutinase with increased activity on polyester substrates. *Biotechnol. J.* **6**, 1230–1239 (2011).

94. Bürgi, H. B., Dunitz, J. D., Lehn, J. M. & Wipff, G. Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron* **30**, 1563–1572 (1974).

95. R. S. Mulliken. Electronic Population Analysis on LCAO. *J Chem Phys* **23**, 1833 (1955).

96. Philips, J. J., Hudspeth, M. A., Browne, P. M. & Peralta, J. E. Basis set dependence of atomic spin populations. *Chem. Phys. Lett.* **495**, 146–150 (2010).

97. Zong, Z. *et al.* Mechanism and biomass association of glucuronoyl esterase: an α/β hydrolase with potential in biomass conversion. *Nat. Commun.* **13**, 1–10 (2022).

98. Lonsdale, R., Hoyle, S., Grey, D. T., Ridder, L. & Mulholland, A. J. Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling. *Biochemistry* (2012) doi:10.1021/bi201722j.

99. Lodola, A., Mor, M., Sirirak, J. & Mulholland, A. J. Insights into the mechanism and inhibition of fatty acid amide hydrolase from quantum mechanics/molecular mechanics (QM/MM) modelling. *Biochem. Soc. Trans.* **37**, 363–367 (2009).

100. Goldenzweig, A. & Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).

101. Miroshnichenko, M. L. & Bonch-Osmolovskaya, E. A. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. *Extremophiles* **10**, 85–96 (2006).

102. Arey, J. S., Aeberhard, P. C., Lin, I. C. & Rothlisberger, U. Hydrogen bonding described using dispersion-corrected density functional theory. *J. Phys. Chem. B* **113**, 4726–4732 (2009).

103. Hubbard, R. E. & Kamran Haider, M. Hydrogen Bonds in Proteins: Role and Strength. *eLS* (2010) doi:10.1002/9780470015902.a0003011.pub2.

104. Kazlauskas, R. Engineering more stable proteins. *Chem. Soc. Rev.* **47**, 9026–9045 (2018).

105. Levinson, N. M. & Boxer, S. G. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.* **10**, 127–132 (2014).

106. Camilloni, C. *et al.* Towards a structural biology of the hydrophobic effect in protein folding. *Sci. Rep.* **6**, 1–9 (2016).

107. Xiao, Z. *et al.* Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting-temperature- guided sequence alignment. *Appl. Environ. Microbiol.* **74**, 1183–1189 (2008).

108. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. F. & Thomas, S. T. Contribution of surface salt bridges to protein stability: Guidelines for protein engineering. *J. Mol. Biol.* **327**, 1135–1148 (2003).

109. Gromiha, M. M. Protein Stability. *Protein Bioinforma.* 209–245 (2010) doi:10.1016/b978-8-1312-2297-3.50006-0.

110. Acids, A. & Dougherty, D. a. Cation-p Interactions Involving Aromatic. *J. Nutr.* **137**, 1504–1508 (2007).

111. Zhou, Y., Wang, S. & Zhang, Y. Catalytic reaction mechanism of acetylcholinesterase determined by born-oppenheimer Ab initio QM/MM molecular dynamics simulations. *J. Phys. Chem. B* **114**, 8817–8825 (2010).

112. Mitchell, J. B. O., Nandi, C. L., McDonald, I. K., Thornton, J. M. & Price, S. L. Amino/aromatic interactions in proteins: Is the evidence stacked against hydrogen bonding? *Journal of Molecular Biology* vol. 239 315–331 Preprint at https://doi.org/10.1006/jmbi.1994.1370 (1994).

113. Prajapati, R. S., Sirajuddin, M., Durani, V., Sreeramulu, S. & Varadarajan, R. Contribution of cation-π interactions to protein stability. *Biochemistry* **45**, 15000–15010 (2006).

114. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 (2014).

115. Xu, Z., Cen, Y., Zou, S., Xue, Y. & Zheng, Y. Critical Reviews in Biotechnology Recent advances in the improvement of enzyme thermostability by structure modification. *Crit. Rev. Biotechnol.* **40**, 83–98 (2020).

116. Dani, V. S., Ramakrishnan, C. & Varadarajan, R. MODIP revisited: Re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins. *Protein Eng.* **16**, 187–193 (2003).

117. Dombkowski, A. A. Disulfide by Design^{TM}: A computational method for the rational design of disulfide bonds in proteins. *Bioinformatics* **19**, 1852–1853 (2003).

118. Le, Q. A. T., Joo, J. C., Yoo, Y. J. & Kim, Y. H. Development of thermostable Candida antarctica lipase B through novel in silico design of disulfide bridge. *Biotechnol. Bioeng.* **109**, 867–876 (2012).

119. Chopra, N., Kumar, A. & Kaur, J. Structural and functional insights into thermostable and organic solvent stable variant Pro247-Ser of Bacillus lipase. *Int. J. Biol. Macromol.* **108**, 845–852 (2018).

120. Han, N. *et al.* Enhancing thermal tolerance of a fungal GH11 xylanase guided by B-factor analysis and multiple sequence alignment. *Enzyme Microb. Technol.* **131**, (2019).

121. Bonds, D. crossm Enhancing the Thermostability of Rhizomucor miehei Lipase with a Limited Screening Library by Rational-Design Point. **84**, 1–16 (2018).

122. Marx, C. K., Hertel, T. C. & Pietzsch, M. Random mutagenesis of a recombinant microbial transglutaminase for the generation of thermostable and heat-sensitive variants. *J. Biotechnol.* **136**, 156–162 (2008).

123. Buettner, K., Hertel, T. C. & Pietzsch, M. Increased thermostability of microbial transglutaminase by combination of several hot spots evolved by random and saturation mutagenesis. *Amino Acids* **42**, 987–996 (2012).

124. Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Res.* (2006) doi:10.1093/nar/gkl163.

125. Wang, X. *et al.* Significantly Improving the Thermostability and Catalytic Efficiency of Streptomyces mobaraenesis Transglutaminase through Combined Rational Design. *J. Agric. Food Chem.* **69**, 15268–15278 (2021).

126. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

127. Khan, S., Farooq, U. & Kurnikova, M. Exploring protein stability by comparative molecular dynamics simulations of homologous hyperthermophilic, mesophilic, and psychrophilic proteins. *J. Chem. Inf. Model.* **56**, 2129–2139 (2016).

128. Berendsen, H. J. C., Van Der Spoel, D. & Van Drunen, R. *GROMACS: A message-passing parallel molecular dynamics implementation PROGRAM SUMMARY Title of program: GROMACS version 1.0*. *Computer Physics Communications* vol. 91 (1–3) (1995).

129. Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes Downloaded from. *Science* **319**, 1387–1391 (2018).

130. Drewitt, J. G. N. & Lincocoln, J. Improvements in polymers. (1946).

131. Gandini, A. The irruption of polymers from renewable resources on the scene of macromolecular science and technology. *Green Chem.* **13**, 1061–1083 (2011).

132. Loos, K. *et al.* A Perspective on PEF Synthesis, Properties, and End-Life. *Front. Chem.* **8**, 1–18 (2020).

133. Rosenboom, J. G., Hohl, D. K., Fleckenstein, P., Storti, G. & Morbidelli, M. Bottle-grade polyethylene furanoate from ring-opening polymerisation of cyclic oligomers. *Nat. Commun.* **9**, 1–7 (2018).

134. Carlos Morales-Huerta, J., Martínez De Ilarduya, A. & Muñoz-Guerra, S. Poly(alkylene 2,5-furandicarboxylate)s (PEF and PBF) by ring opening polymerization. *Polymer* **87**, 148–158 (2016).

135. Qu, X. ling *et al.* A Brønsted Acidic Ionic Liquid as an Efficient and Selective Catalyst System for Bioderived High Molecular Weight Poly(ethylene 2,5-furandicarboxylate). *ChemSusChem* **12**, 4927–4935 (2019).

136. Jiang, Y., Woortman, A. J. J., Alberda Van Ekenstein, G. O. R. & Loos, K. A biocatalytic approach towards sustainable furanic-aliphatic polyesters. *Polym. Chem.* **6**, 5198–5211 (2015).

137. Chheda, J. N. & Dumesic, J. a. Production of Hydroxymethylfurfural from Fructose. *Science* **312**, 1933 (2006).

138. Gruter, G. J. M. & Dautzenberg, F. Method for the synthesis of 5-hydroxymethylfurfural ethers and their use. **1**, US Patent 8,877,950 B2 (2014).

139. Eerhart, A. J. J. E., Faaij, A. P. C. & Patel, M. K. Replacing fossil based PET with biobased PEF; Process analysis, energy and GHG balance. *Energy Environ. Sci.* **5**, 6407–6422 (2012).

140. Corma Canos, A., Iborra, S. & Velty, A. Chemical routes for the transformation of biomass into chemicals. *Chem. Rev.* **107**, 2411–2502 (2007).

141. Palm, G. J. *et al.* Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. *Nat. Commun.* **10**, 1–10 (2019).

142. Department for Business Energy & Industrial Strategy. Energy Consumption in the UK (ECUK) 1970 to 2020. 1–21 (2021).

143. Huang, Y. B. & Fu, Y. Hydrolysis of cellulose to glucose by solid acid catalysts. *Green Chem.* **15**, 1095–1111 (2013).

144. Wei, R. *et al.* Biocatalytic Degradation Efficiency of Postconsumer Polyethylene Terephthalate Packaging Determined by Their Polymer Microstructures. *Adv. Sci.* **6**, (2019).

145. Van Berkel, J. G., Guigo, N., Visser, H. A. & Sbirrazzuoli, N. Chain Structure and Molecular Weight Dependent Mechanics of Poly(ethylene 2,5-furandicarboxylate) Compared to Poly(ethylene terephthalate). *Macromolecules* **51**, 8539–8549 (2018).

146. Burgess, S. K., Kriegel, R. M. & Koros, W. J. Carbon dioxide sorption and transport in amorphous poly(ethylene furanoate). *Macromolecules* **48**, 2184–2193 (2015).

147. Burgess, S. K., Karvan, O., Johnson, J. R., Kriegel, R. M. & Koros, W. J. Oxygen sorption and transport in amorphous poly(ethylene furanoate). *Polymer* **55**, 4748–4756 (2014).

148. Carlsberg. Carlsberg makes bio-based and fully recyclable bottles available to consumers in its largest ever trial. https://www.carlsberggroup.com/newsroom/carlsberg-makes-bio-based-and-fully-recyclable-bottles-available-to-consumers-in-its-largest-ever-trial/ (2022).

149. Araujo, C. F. *et al.* Inside PEF: Chain Conformation and Dynamics in Crystalline and Amorphous Domains. *Macromolecules* **51**, 3515–3526 (2018).

150. Arcus, V. L. *et al.* On the Temperature Dependence of Enzyme-Catalyzed Rates. *Biochemistry* **55**, 1681–1688 (2016).