DIRECTED EVOLUTION OF AN ORGANOPHOSPHATE HYDROLASE: METHYL PARATHION HYDROLASE

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF THE AUSTRALIAN NATIONAL UNIVERSITY

TEE-KHEANG NG

SUBMITTED FOR EXAMINATION IN MARCH 2014

Declaration by Author

This thesis presents original research that I have performed under the supervision of Professor David Ollis at the Research School of Chemistry, Australian National University. Contributions of individuals to this thesis are listed below:

- Plasmid of wild-type *mpd* in pET47b (including signal sequence) was obtained from the Ollis Group Plasmid Stock Collection.
- Fourier transform ion cyclotron resonance mass spectroscopy (FTICR-MS) measurements and data analysis were performed by Professor Thomas Huber.
- Inductively coupled plasma-optical emission spectroscopy measurements and analysis were performed by Miss Nur Hafizah Azizan.

This work has not been presented, in whole or in part, for any other degree.

This thesis conforms to the Australian National University guidelines and regulations of Higher Degree Research award and thesis production.

Tee-Kheang Ng

Acknowledgements

I would like to express my deepest appreciation to Professor David Ollis for giving me a challenging project and for his patience, support and encouragement throughout my candidature at the Research School of Chemistry. I thank Tracy Murray for her technical help, as well as her commitment in ensuring the smooth running of the lab. I thank Dr. Bradley Stevenson for his invaluable input and assistance in various biochemical techniques and troubleshooting. I am grateful to Dr Christopher Wanty for being my mentor when I first arrived in the group, Professor Thomas Huber for his time and effort in mass spectroscopy measurement and analysis and Miss Nur Hafizah Azizan for her time and effort in ICP-OES measurement and analysis.

I thank all past and present members of the Biological Chemistry Group, especially Jee-Loon, Sylvia, Amalia, Joe, Morgan, David Hou, Jo, Shu-Ann, Jo-Leen, Nurul, Adli, Vy, Tiffany, Ruhu and Karin for their support on a personal and professional level.

Special thanks to Shu-Ann for the long discussions on chemistry, screening designs, statistical theories, as well as the time and effort spent proofreading this thesis.

I am grateful to Research School of Chemistry and the Australian National University for financially supporting my PhD candidature.

DNA sequencing was made possible by the Biomolecular Resource Facility based at the John Curtin School of Medical Research, ANU.

Last but no least, I am grateful to my family for being supportive in my long journey to complete this work.

Abstract

Organophosphates (OPs) are the most common pesticides used in agriculture. Although they can be broken down in nature, OPs pose a severe health hazard to human due to their inhibitory effect on acetylcholinesterase, a major enzyme in nervous transmission. Therefore, detoxification of water and soil contaminated by OPs is important. One way of achieving this is by bioremediation of sites with OP-degrading enzymes. One such enzyme is methyl parathion hydrolase (MPH) isolated from *Pseudomonas* sp. WBC-3.

MPH is a highly efficient enzyme that is capable of hydrolysing methyl parathion at near diffusion-limited rate. While MPH can hydrolyse a wide range of OPs, the substrate specificity of the enzyme was not well characterised. In order to study MPH, the enzyme was expressed and purified. In the process of MPH protein purification, proteolytic degradation was observed. Various methods, including protein engineering and optimising the purification, were employed to investigate and overcome the degradation. A protocol that allowed rapid purification of MPH was developed so that the proteolysis can be minimised. Due to initial suspicion of autoproteolysis, nickel affinity chromatography was also used in further investigations and autoproteolysis was eventually ruled out.

Stability is one of the most important characteristics that define an enzyme's practical use in the industry. For MPH to be an effective bioremediator, it needs to be thermally and chemically stable. Unfortunately, MPH does not have exceptional thermostability and could benefit from extra thermostability. To achieve this, MPH was subjected to directed evolution for enhanced thermostability. In the course of characterising the mutants isolated, it was discovered that MPH expressed in *E. coli* had lower than expected metal content. It was also found that Zn^{2+} supplementation prior to activity and stability assay drastically increased the activity and stability of WT MPH. Since the evolution was performed without metal supplementation and the isolated mutant did not have enhanced stability with Zn^{2+} supplementation, we hypothesised that the mutant isolated was stabilised "metal independently".

Another desired characteristic for a bioremediator is the ability to hydrolyse various OPs efficiently. The substrate profile characterisation of WT MPH revealed that

while MPH is highly efficient towards methyl parathion, its activity towards other OPs varies. To alter and broaden the substrate specificity of MPH, structure-guided site saturation mutagenesis (SSM) on active site lining residues was performed to obtain mutants with enhanced activity towards ethyl paraoxon. Mutants with modest improvements were isolated and two rounds of DNA shuffling were performed to compound the mutations. The best mutant towards ethyl paraxon exhibited 98-fold increase in k_{cat}/K_m. Several other mutants exhibited interesting and respectable changes in their substrate profiles. One mutant with selective activity towards chlorpyrifos class substrates was found. These results highlighted the 'plasticity' of MPH active site that allow efficient hydrolysis of other OPs with only minor changes.

In short, progress had been made in purifying MPH and in evolving it to be more stable – although further work is required in this area. Considerable progress had been made in identifying mutations that alter the substrate specificity of MPH.

Table of Contents

List of Figures

List of Tables

Abbreviations

In this thesis, amino acid residues are referred to by the standard one or three letter codes. For example, Phe119 and F119 refer to phenylalanine at position 119. Mutations are indicated in the format of X#Y, where X is the original amino acid residue, # is the location of the residue and Y is the new amino acid residue.

Abbreviation	**Full name**
2-PAM	pralindoxime
ACh	acetylcholine
AChE	acetylcholinesterase
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AIM-1	B3 metallo-β-lactamase from *Pseudomonas aeruginosa*
AMPP	aminopeptidase P
APS	ammonium persulfate
BcII	B1 metallo-β-lactamase from *Bacillus cereus*
bpNPP	bis-(*p*-nitrophenyl) phosphate
cfu	colony-forming unit
CIP	calf intestinal alkaline phosphatase
CphA	B2 metallo-β-lactamase from *Aeromonas hydrophila*
CV	coefficient of variation
DFP	diisopropyl fluorophosphate
DFPase	diisopropyl fluorophosphatase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	2'-deoxyribonucleotide triphosphate
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
ECO	(ethyl) chlorpyrios oxon
ECS	(ethyl) chlorpyrifos
EDTA	ethylenediaminetetraacetic acid
endA	gene coding for EndA, a non-specific endonuclease
ePCR	error-prone PCR
EPO	(ethyl) paraoxon
EPS	(ethyl) parathion
EtOH	ethanol
FACS	fluorescence-activated cell sorting
FPLC	fast protein liquid chromatography
FTICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
GFP	green fluorescent protein
GloB	glyoxalase II from *Salmonella typhimurium*
Glx2	glyoxlase II from *Homo sapiens*
GpdQ	glycerophosphodiester phosphodiesterase from *Enterobacter aerogenes*
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HTS	high throughput screening
ICP-OES	inductively coupled plasma - optical emission spectroscopy
IMAC	immobilised metal-affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
IVC	*in vitro* compartmentalisation
L1	B3 metallo-β-lactamase from *Stenotrophomonas maltophilia*
LB	Lysogeny broth media
LBA	Lysogeny broth media supplemented with ampicillin
LBK	Lysogeny broth media supplemented with kanamycin
MBL	metallo-β-lactamase
MCO	methyl chlorpyrifos oxon
MCS	multiple-cloning site
MEGAWHOP	megaprimer PCR for whole plasmid
MeOH	methanol
mpd	gene coding for methyl parathion hydrolase
MPH	methyl parathion hydrolase
MPO	methyl paraoxon

MPS	methyl parathion
OP	organophosphate
OPAA	organophosphorus acid anhydrolase
opd	gene coding for OPH
OPDA	organophosphate-degrading enzyme from *Agrobacterium radiobacter* P230
opdA	gene coding for OPDA
OPH	organophosphate-degrading enzyme from *Flavobacterium* sp. ATCC 27551
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
pNDPP	*p*-nitrophenyl diphenylphosphate
pNP	*p*-nitrophenolate
pNPA	*p*-nitrophenyl acetate
pNPP	*p*-nitrophenyl phosphate
PON	paraoxonase
PTE	phosphotriesterase
QCM	Stratagene's Quikchange Method
RBS	ribosomal binding site
RMSD	root mean square deviation
RNA	ribonucleic acid
SAP	shrimp alkaline phosphatase
SDM	site directed mutagenesis
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SOE	sequence overlap extension
SSM	site saturation mutagenesis
StEP	staggered extension process
StEP-ePCR	staggered extension process error-prone PCR
TCPy	3,5,6-trichloro-2-pyridinol
TEMED	*N,N,N'N'*-tetrmethylethylenediamine
TIM	triose phosphate isomerase
tnp	gene coding for transposase
Tris	tris(hydroxymethyl)aminomethane
YENB	yeast extract nutrient broth
ZiPD	zinc phosphodiesterase from *E. coli*

INTRODUCTION

1. Introduction

1.1. Preamble

This thesis is concerned with a phosphotriesterase (PTE) known as methyl parathion hydrolase (MPH) that was identified in China. The enzyme, and its gene, was isolated from the bacterium *Pseudomonas* sp. WBC-3 and was found to be able to fully mineralise methyl parathion (MPS). Other bacteria were later found to express MPH: some with identical sequences and activities while some had similar sequences and activities. These proteins have a number of properties that distinguish them from other organophosphate (OP) degrading enzymes and that may make them useful in the bioremediation of pesticide-contaminated water and other applications. However, MPH is not well understood and some of its characteristics are not ideal for practical applications. Therefore, protein-engineering studies directed towards improving its stability and understanding its kinetic properties, as well as altering its substrate specificity, have been undertaken.

A brief description of OP pesticides will be given along with an overview of MPH publications. A brief review of protein-engineering methods used to modify and improve enzyme will also be given. In the final section of the introduction, an outline of the specific objectives of this study will be given.

1.2. Organophosphate pesticides

Organophosphate (OP) is a term used to describe esters, amides or thiol derivatives of phosphoric acid and other sulphur-containing analogues. OPs form a large family of ~50,000 chemical agents with biological properties[1]. OPs can be further classified based on the extent of esterification (**Figure 1.1**). There are formal naming conventions for these compounds, but they are usually referred to by their common names in the literature and in the present work. Standard reference text can be consulted for a detailed coverage of phosphorus chemistry and only a brief description of relevant compounds will be given here. The chemical structures of OP pesticides and nerve gases have been compiled and those referred to in this thesis are given in **Appendix D**.

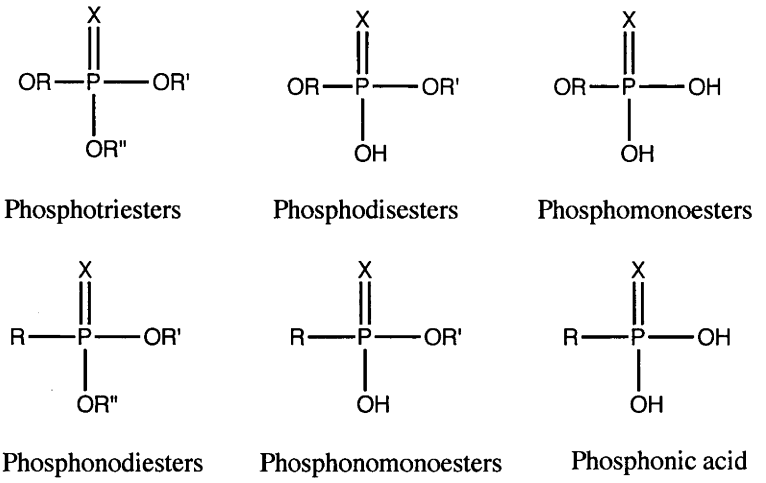
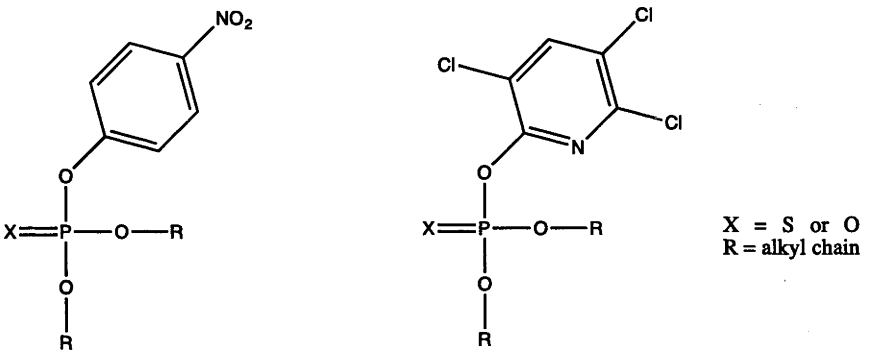


Figure 1.1: Some structures of different classes of organophosphates.
X can be O or S; R, R', R" can be alkyl, ethyl or phenyl groups.

Many biological molecules contain phosphate and addition, subtraction or modification of the phosphate group is often involved in the regulation of various metabolic processes[2]. DNA and RNA, the genetic material in organisms, contain phosphodiester backbone that connects individual nucleotides to form long DNA and RNA chains. Phospholipid, the main component of cellular membranes, has a phosphodiester hydrophilic head that gives rise to the amphibatic character of cellular membranes. Glucose-6-phosphate, a monophosphoester from the phosphorylation of glucose, has prominent roles in the cells, as it is the substrate for glycolysis and the pentose phosphate pathway. Phosphate esters are usually stable and are broken down enzymatically when required in cellular metabolism. The stability of phosphate esters makes them of considerable interest in the mechanisms of enzymatic modification and hydrolysis[2]. While phosphomonoesters and phosphodiesters are abundant naturally, phosphotriesters do not occur naturally[2]. OP pesticides are triesters – these are unusual.

OP pesticides were first synthesised in the 19th century as a chemical curiosity. In the early 20th century, German chemist Gerhard Schrader noticed the insecticidal properties of these compounds and began research into the development of more superior insecticides and subsequently military nerve agents[3]. OP pesticides and nerve agents are similar; they differ in their toxicity. Nerve agents are more toxic to humans while the less toxic OPs are used as insecticides[4]. However, OP pesticides actions are non-specific and can be lethal to human if exposed to in large quantities.

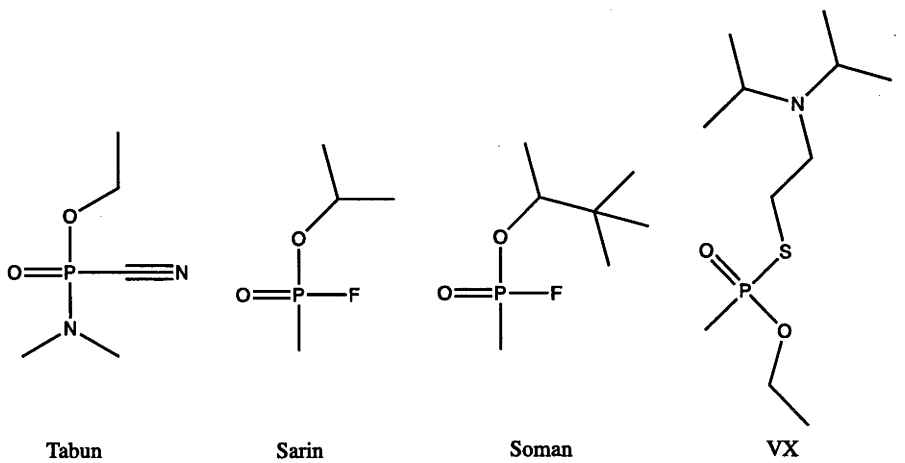
The structures of a few common pesticides are given in the figure below to explain the abbreviation convention used in this thesis. Some of these pesticides are substrates of MPH and the abbreviations used to denote the compounds are given in **Figure 1.2.** The structures of some nerve agents are also given in **Figure 1.3.** The OP pesticides used in the work for this dissertation were mainly parathion and chlorpyrifos derivatives. The parathion derivatives have a *p*-nitrophenolate (pNP) leaving group whereas chlorpyrifos derivatives have a 2,5,6-trichloro-2-pyridinol (TCPy) leaving group. In all the pesticide abbreviations used in this thesis, a three-letter format was adopted. The abbreviation for methyl parathion, for example, is MPS, where M indicates it has a dimethyl substitute; P indicates it has a *p*-nitrophenolate leaving group and S indicates that it is a thiophosphate.



Parathion derivative
Methyl parathion (MPS): R = 1C; X = S
Methyl paraxon (MPO): R = 1C; X = O
Ethyl parathion (EPS): R = 2C; X = S
Ethyl paraoxon (EPO): R = 2C; X = O

Chlorpyrifos derivative
Methyl chlorpyrifos (MCS): R = 1C; X = S
Methyl chlorpyrifos oxon (MCO): R = 1C; X = O
Ethyl chlorpyrifos (ECS): R = 2C; X = S
Ethyl chlorpyrifos oxon (ECO): R = 2C; X = O

Figure 1.2: The general structures of parathion and chlorpyrifos derivatives and the abbreviations.



Tabun Sarin Soman VX

Figure 1.3: Structures of some nerve agents.

Both OP pesticides and nerve agents act on acetylcholinesterase (AChE). AChE is a serine hydrolase that catalyses the breakdown of acetylcholine (ACh) at cholinergic synapses and the catalytic mechanism is shown in **Figure 1.4a**[5,6]. OP pesticides and nerve agents function by covalently modifying the catalytic serine of AChE (**Figure 1.4b**). As can be seen in the figure, the serine displaces an alcohol that is generally a good leaving group like *p*-nitrophenolate (pNP). The specificity of these agents is determined by a number of factors – the alkyl chains being the most important. OP pesticides are usually triesters with two short chain alcohol such as diethyl and dimethyl substituents. The result of the inhibition of AChE by OPs, is the covalent modification of the nucleophile, forming an extremely stable and slow hydrolysing adduct that mimics the intermediate in ACh hydrolysis[5,6]. Dealkylation of the adduct to a phosphodiester, also known as aging, leads to irreversible AChE inhibition and causes hyper stimulation of nerve cells and death. The symptoms of OPs poisoning can be found in a review by Bajgar[7].

Currently, treatment for OP poisoning relies on chemical antidotes such as cholinolytics, oximes, atropine and anticonvulsants to minimise toxic manifestations.[8] Classical studies done by Wilson and his colleagues showed that strong nucleophiles such as oximes can reactivate phosphorylated AChE.[9,10] Examples of commercially available oximes can be found in **Figure 1.5a**. Oximes reactivate the phosphorylated AChE by attacking the phosphate centre of the adduct to regenerate the free catalytic serine (**Figure 1.5b**).[11,12] Oximes such as pralidoxime (2-PAM) and HI-6 are often used with atropine, a muscarinic acetylcholine receptor antagonist, as antidotes in acute OP poisoning therapy.[13] However, they have to be administered quickly after exposure as they are impotent against aged phosphorylated AChE. Furthermore, the effectiveness of oxime reactivation differs among OP compounds.[14] OP-degrading enzymes like MPH and OPDA can be used to remove excess OPs from the bloodstream, allowing reactivation to occur without re-inhibition.

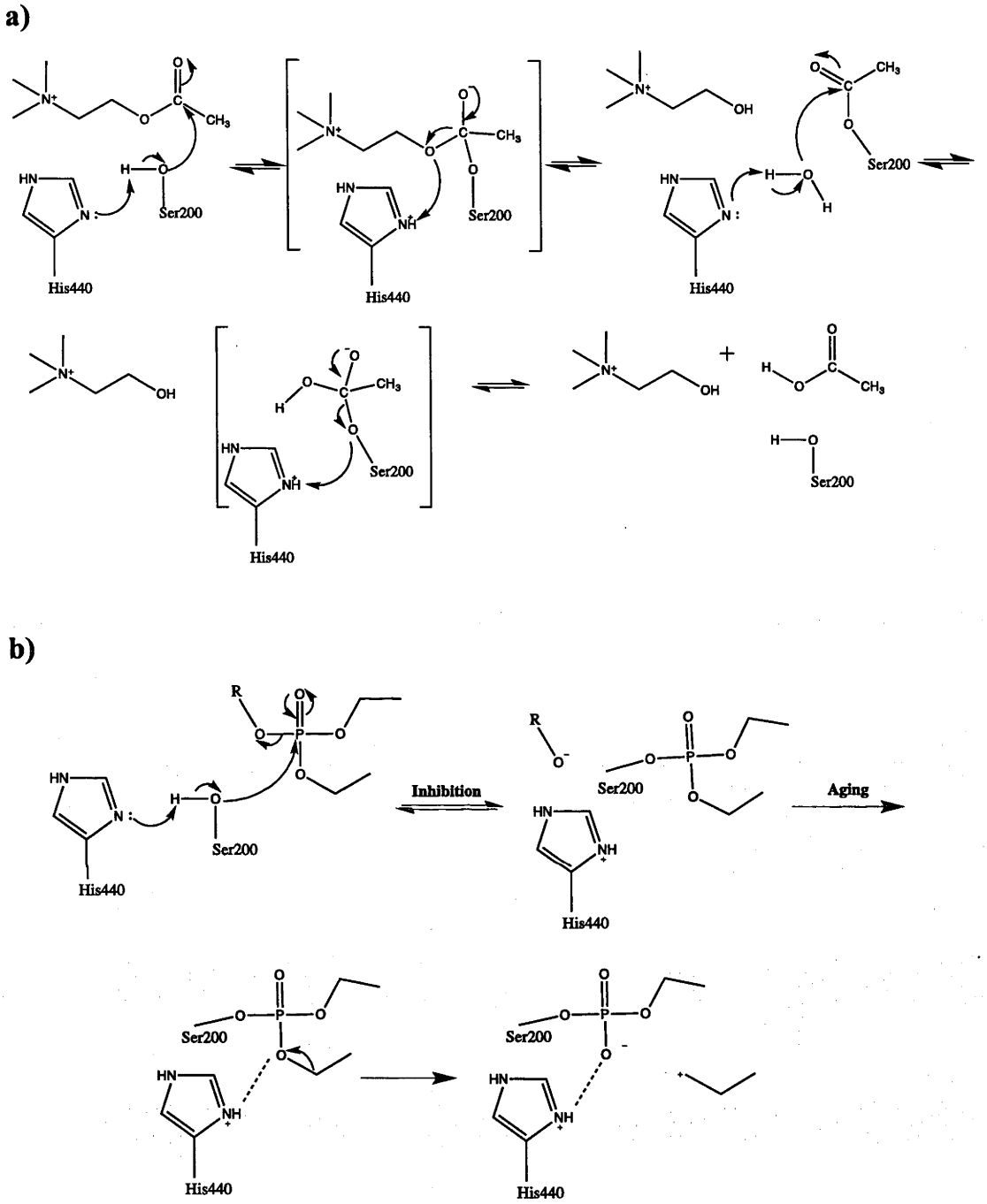
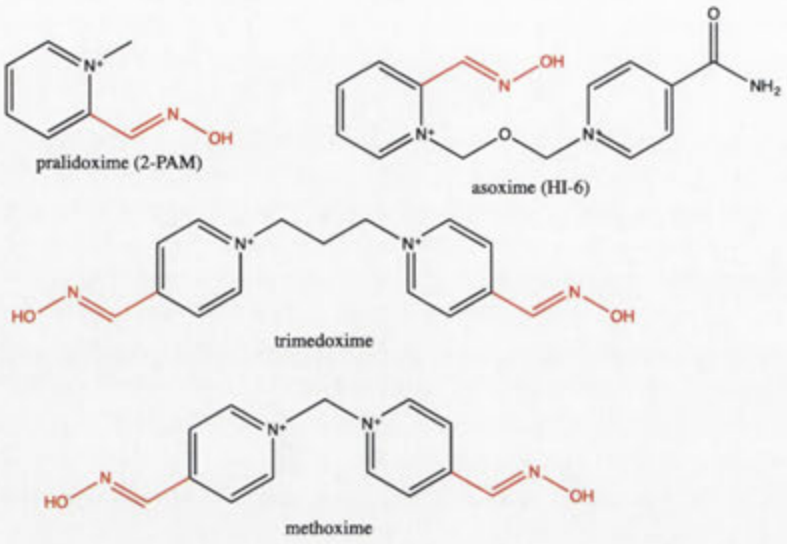


Figure 1.4: The mechanism of acetylcholinesterase and its inhibition by organophosphate.
The reaction mechanism of AChE with its natural substrate, ACh is shown in (a). His440 acts as a general base, abstracting the proton from Ser200. Transesterification of ACh at Ser200 of AChE via an unstable tetrahedral intermediate liberates choline. His440 then again acts as general base, abstracting a proton from a water molecule, leading to the hydrolysis of the acetylated enzyme and regenerates the active enzyme. However, as shown in (b), when exposed to OPs, the compound binds to the active site and phosphorylates Ser200. The resulting adduct is a stable structure that resembles the transition state in ACh hydrolysis that hydrolyses extremely slowly[5,6]. Aging of the adduct through dealkylation leads to irreversible inhibition of AChE. R is the leaving group of the OP.

a)



b)

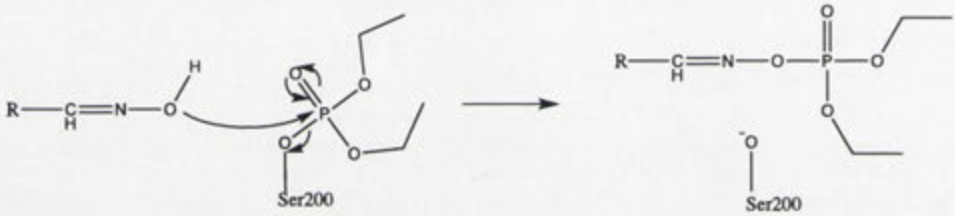


Figure 1.5:Examples of AChE reactivators for OP poisoning and its reactivation mechanism.
a) Examples of commercially available oxime antidotes. The oxime groups are highlighted in red. (b) Oximes are strong nucleophiles and can reactivate phosphorylated AChE through nucleophilic attack on the organophosphate moiety. The oxime is phosphorylated, liberating the Ser200.

1.3. Organophosphate-degrading enzymes

1.3.1. Enzymes other than MPH

Organophosphate-degrading enzymes were thought to have evolved in recent times, as OPs were not present in the environment until after the Second World War.[15] OPs are xenobiotics and are not naturally found in the biosphere. Hence, the enzymes that degrade these compounds have either evolved over a span of less than 100 years, or they are enzymes that have an alternative primary function and fortuitously break down OPs. Enzymes with both evolutionary origins appear to have been identified.[16,17]

In 1946, Abraham Mazur found an enzyme in human and rabbit tissues that was capable of hydrolysing diisopropyl fluorophosphate (DFP), a dialkyl fluorophosphate, and other OP compounds. He found that rabbit and human liver extract contains high amounts of enzymes and noted that liver plays a significant role in DFP detoxification.[18] In 1950s, Aldrige studied the hydrolysis of various carboxylesters and paraoxon using sera from human and various mammals and proposed that serum esterases can be classified into A and B-esterases. He found that the A-esterases (subsequently named paraoxonase) were capable of hydrolysing organophosphates and aromatic carboxyl esters and the B-esterases (serine carboxylesterases and acetylcholinesterases) were inhibited by paraoxon.[16,19,20] The name was eventually shortened to PON and various homologs were identified. The paraoxonase family consists of three members- PON1, PON2 and PON3 that were named in the order of their discovery. These different homologs share about 65% sequence similarities at the amino acid level and have different cellular localisations.[21] PON1 has been the most studied paraoxonase and was found to require two calcium ions for activity and stability.[22] The structure of a PON1 mutant variant was solved and was found to be a six bladed β-propeller as shown in **Figure 1.6**.[23] PON1 is unstable and aggregates in the absence of detergents, as it is a lipoprotein-anchored enzyme.[24]

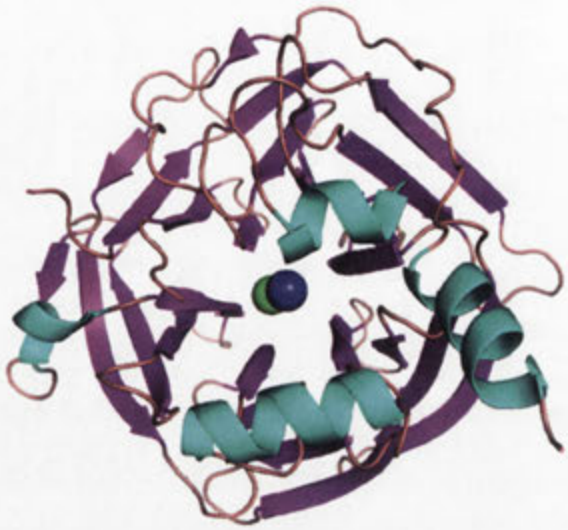


Figure 1.6: Structure of PON1 mutant variant (PDB ID: 1V04).
Figure shows top view of β-propeller fold. The catalytic calcium is shown in blue and the structural calcium is shown in green.

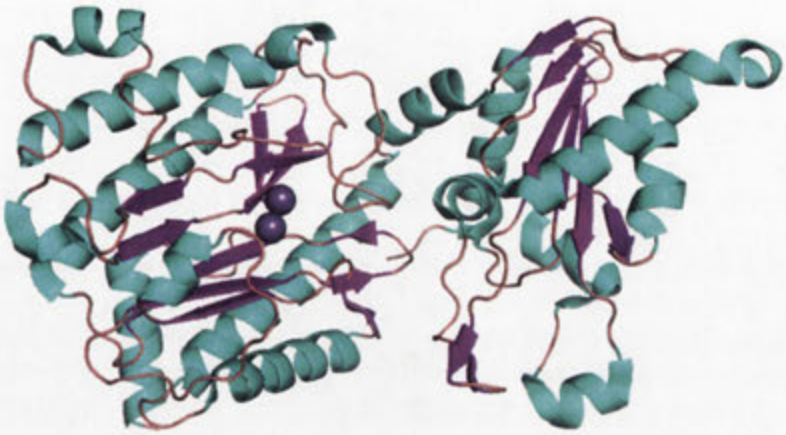
In addition to a variety of OPs, PON1 can degrade aromatic esters as well as aliphatic and aromatic lactones. The latter compounds are probably its physiological targets as OPs are not naturally occurring and its ability to degrade these compounds was found to be relatively modest. Several reviews can be consulted for more information on this enzyme.[21,25]

In 1982, Hoskins and Roush identified a new enzyme that, like PON1, was capable of degrading DFP and OPs and the enzyme was named diisopropyl fluorophosphatase (DFPase).[26] The enzyme was isolated from squid axons and the crystal structure was eventually solved.[27,28] It was found that DFPase is a calcium-dependent hydrolase and is structurally related to PON1 (**Figure 1.7**). Both PON1 and DFPase exhibit the six bladed β-propeller fold.

Figure 1.7: Structure of DFPase (PDB ID: 2GVV).
Figure shows top view of β-propeller fold. The catalytic calcium is shown in blue and the structural calcium is shown in green.

Another class of promiscuous OP-degrading enzymes that received attention is organophosphorus acid anhydrolase (OPAA). While DFPase and OPAAs share the capability of hydrolysing DFP and can be grouped together as OPAAs based on the activities they catalyse, it was eventually shown that DFPase does not share the same protein folding as the other OPAA members. To avoid confusion, these two enzymes would be classified differently in this discussion, based on the classification used by Singh[15] and Bigley *et al.*[29] in their respective reviews. OPAA was isolated from *Alteromonas udina* and *Alteromonas haloplanktis.*[30,31] Similar to DFPase, OPAA was found to be highly efficient against DFP and various G-type nerve agents.[31] While both DFPase and OPAA are highly efficient against DFP, OPAA crystal structure reveals that OPAA and DFPase are structurally unrelated. OPAA is found to be related to aminopeptidase P (AMPP) as both enzymes show a "pita bread" fold (**Figure 1.8**).[32,33] AMPP catalyses the cleavage of amino-terminal X-Pro peptide bonds. OPAA belongs to this dipeptidase family and does not share sequence or structural similarity with other OP-degrading enzymes, suggesting that the OP degrading capability of OPAA might have evolved from a different progenitor.[15] Jao and his colleagues demonstrated that AMPP is capable of hydrolysing a wide range of OP triesters[34], suggesting the possibility of other enzymes in the superfamily to also have OP-degrading activity.

(a)



(b)

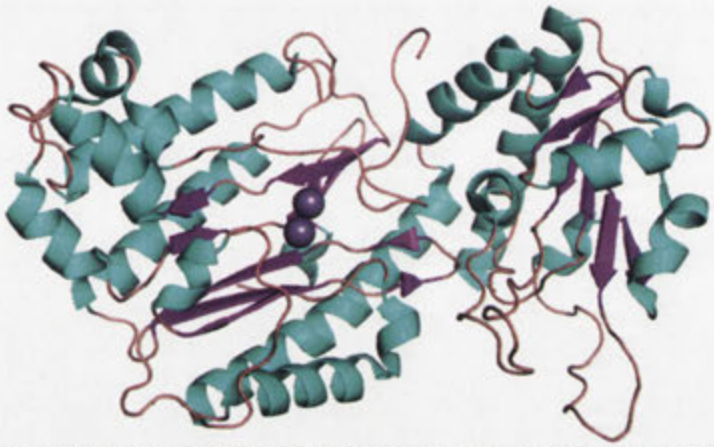


Figure 1.8: Monomeric structures of (a) OPAA (PDB ID: 3L7G) and (b) AMPP (PDB ID: 1WL9). The manganese ions for both enzymes are shown as spheres.

Insects are the main targets of OPs and it is not surprising that resistance has emerged in them. Mutations that confer this resistance have been found in a family of esterase genes. Newcomb and colleagues found a single amino acid substitution in blowfly carboxylesterase that converts it to an organophosphorus hydrolase.[35] This mutation is thought to introduce a general base, which is capable of breaking down the covalent product formed by the OP, into the active site. Although the OP-degrading activity of the mutant is low, the naturally occurring mutant illustrates evolution in action. It should be noted that these insect esterases belong to the α/β hydrolase fold and function without metals.[36] In this respect, they are structurally different from the other OP-degrading enzymes (**Figure 1.9**).

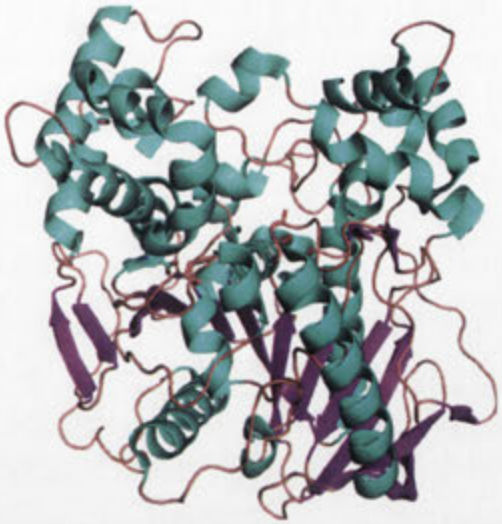


Figure 1.9: Structure of αEsterease7 (PDB ID: 4FNG).

MPH was not the first OP-degrading enzyme that was isolated from bacteria found in soil contaminated with pesticides. Various OP-degrading bacteria were isolated and the genes and the enzymes responsible for the OP-degrading capability were isolated from them. In 1973, Sethunathan and Yoshida reported the first OP-degrading bacteria isolated from the environment. The bacterium, *Flavobacterium* sp. ATCC 27551, was found in a Philippines paddy field that had been treated with diazinon for years[37]. Phosphotriesterases (PTEs), the enzymes eventually isolated and characterised, are a group of enzymes that can degrade OP compounds, and they can be found in various organisms. Singh, in his review, stated that there are currently three different types of well-characterised bacterial PTEs: *opd* type hydrolase, OPAA and MPH[15].

Three PTEs from the *opd* type hydrolase, OPH from *Flavobacterium* sp. and *Pseudomonas diminuta* and OPDA from *Agrobacterium radiobacter,* share about 90% sequence similarity and similar activity profiles.[38] Although no physiological substrates are known for OPH and OPDA, structural and sequence similarities suggest that they have evolved from the lactonases that function in quorum sensing.[17] These PTEs were found to have a very high activity towards OPs and degrade some substrates like paraoxon at diffusion-limited rates.[38] They are metal dependent and function with a variety of divalent metals – Zn^{2+} being reported as the preferred metal for OPH while OPDA has been observed to have Fe^{2+}/Zn^{2+} centre.[39] PTEs are also active with binuclear Co^{2+}, Ni^{2+}, Cd^{2+}or Mn^{2+} centres.[40]

Structural studies have been carried out with OPH and OPDA. As expected, both have very similar structures and are placed in the amidohydrolase fold superfamily.[41,42] Both enzymes are homodimers with a TIM barrel fold, consisting of a distorted α/β barrel formed by eight parallel β-sheets and surrounded by 14 α-helices on the outer surface, as is typical of amidohydrolases superfamily enzymes (**Figures 1.10 and 1.11**).[38,43] Reviews on the members of this fold have been published and discussed in detail.[41,44] The metal binding sites of OPH and OPDA are essentially the same and are located at the C-terminal end of the β-strands of the TIM-barrel[43] and features a binuclear metal centres – the metals being designated α and β. His55, His57 and Asp301 coordinate the more buried α-metal, while His201 and His301 coordinate the more solvent exposed β-metal. A bridging hydroxide and a bridging carbamylated Lys169 coordinate both metals (**Figure 1.12**). The α-metal is pentacoordinated, assuming a trigonal bipyrimidal geometry, while the β-metal has a distorted octahedral geometry with two terminally bound water when the metal ion is Co^{2+}, Cd^{2+} and Mn^{2+}.[43,45,46] However, when the β-metal is Zn^{2+}, the metal ion is pentacoordinated with one of the terminal water molecules moved away from Zn^{2+} into a non-coordinating position. The metal assumes a trigonal bipyrimidal coordination geometry.[45,46]

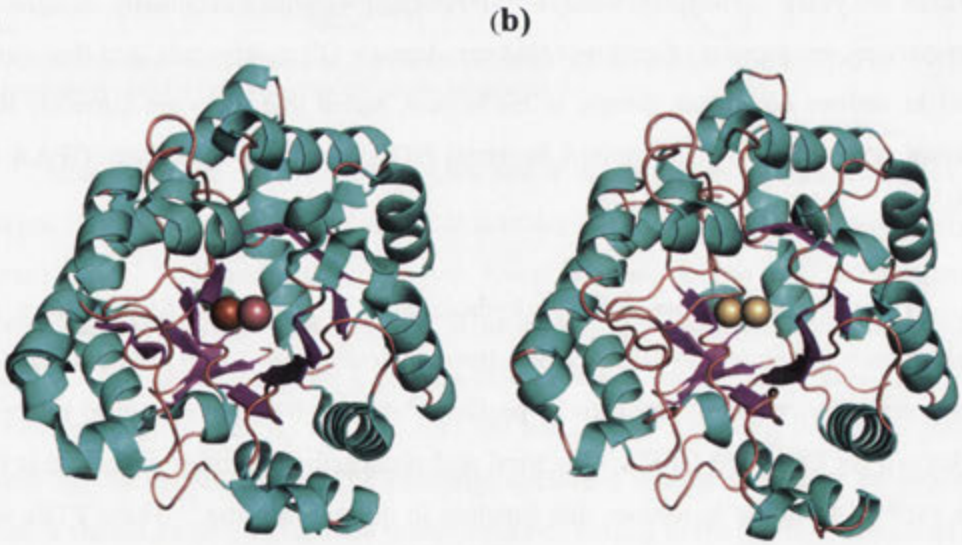


Figure 1.10: Structures of (a) OPDA (PDB ID: 2R1N) and (b) OPH (PDB ID: 1PSC).
(a) and (b) shows the top view of OPDA and OPH respectively. Orange and pink spheres represent the Fe^{2+} and Co^{2+} in OPDA respectively while wheat spheres represent the Cd^{2+} metals in OPH. The OPDA and OPH structures are very similar.

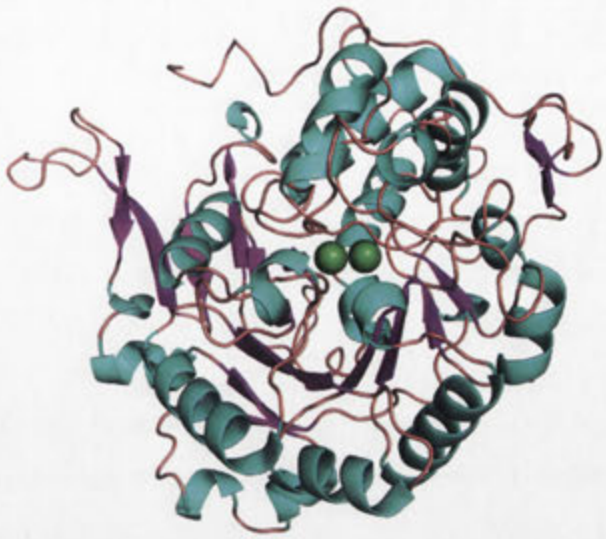


Figure 1.11: Structure of *Klebsiella aerogenes* urease, an amidohydrolase superfamily enzyme (PDB ID: 2KAU).
Figure shows only the TIM-barrel domain of urease. The green spheres represent the catalytic Ni^{2+} metals.

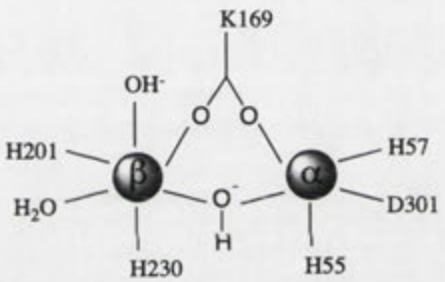


Figure 1.12: Metal coordination centre of OPDA and OPH active site.

1.3.2. Methyl parathion hydrolase (MPH)

1.3.2.1. History and general information

MPH was discovered due to a search for OP-degrading microbes in soil contaminated with pesticides. Two enzymes with identical protein sequences and near-identical DNA sequences were isolated from *Plesiomonas* sp. strain M6 (M6-MPH) and *Pseudomonas* sp. WBC-3 (WBC3-MPH).[47,48] These enzymes exhibit a very high activity towards methyl parathion. Subsequently, many other MPH variants were identified and were found to have high levels of sequence identity to one another.[49–54] *Pseudomonas* sp. WBC-3, from which WBC3-MPH was isolated, was found to fully mineralised methyl parathion as the sole nitrogen and carbon source. The aligned sequences for some of the known MPH variants are shown in **Appendix E.** These sequences all show low levels of similarity to other proteins that are members of the metallo-β-lactamase (MBL) superfamily. When these amino acid sequences are aligned, the conserved putative metal binding residues can be observed. The conserved metal binding motif in MPH is HXHXDH; where X is any amino acid, H is histidine and D is aspartate.

The MBL superfamily was first identified by Neuwald and colleagues while demonstrating the capability of a bioinformatics program to construct multiple alignment models for protein superfamily identification.[55] Daiyasu and colleagues then expanded the zinc superfamily by introducing further classification based on the conservation of metal-binding residues and biological functions.[56] Thus the classification proposed by Daiyasu *et al.* does not always reflect phylogenetic relationship. The MPH variants did not show any sequence similarity to the *opd* type hydrolases such as OPH and OPDA. Subsequent structural analysis, described below, confirmed that MPH is not related to OPH and OPDA. No natural substrate is known for MPH but its structure belongs to the metallo-β-lactamase fold that encompasses esterases that degrade a variety of substrates including β-lactams and lactones.[56,57]

Like OPH and OPDA, MPH variants are binuclear metallo-hydrolases. MPH variants are natively zinc-containing enzyme, but preliminary characterisation by Chu *et al.* showed activity enhancement with Cd^{2+} and Fe^{2+}, as well as activity inhibition with Hg^{2+}, Zn^{2+}, Al^{3+} and Bi^{3+}.[58] However, subsequent work done by Dong *et al.*[59] and

our group confirmed that WBC3-MPH has two Zn^{2+} metals and the Zn^{2+} metals have catalytic and structural roles in MPH as discussed in **Chapter 4**. In addition to methyl parathion (MPS), MPH variants are able to hydrolyse other phosphorothioates (P=S moiety) and oxonates (P=O moiety) such as parathion (EPS), paraoxon (EPO), chlropyrifos (ECS) and fenitrothion.[53,60] WBC3-MPH is able to hydrolyse MPS at near diffusion-limited rate, with k_{cat}/K_m ratio of more than 10^6 $M^{-1}s^{-1}$.[58,59,61] However, the published literature does not have a detailed kinetic characterisation of MPH with many substrates, and only limited information is available. To address this, a detailed kinetic characterisation WBC3-MPH with different OP substrates can be found in **Chapter 5** of this thesis. The literature does however, draw attention to the fact that MPH has a pronounced preference for pesticides with dimethyl substituents and only has low levels of activity with pesticides that have diethyl substituent. This will also be discussed in **Chapter 5**. It was also found in this work that WBC3-MPH has a preference for phosphorothioate substrates like MPS and EPS as opposed to oxonates like MPO and EPO.

1.3.2.2. MPH structure and metallo-β-lactamase superfamily

Dong and his colleagues solved the structure of WBC3-MPH in 2005.[59,62] The structure shows a homodimer with subunits that are related by a two-fold symmetry (**Figure 1.13**). Each subunit is a β-lactamase-like domain, with an αβ/βα sandwich consisting of flanking α-helices and two internal mixed β-sheets. The active site is located between the two β-sheets; with the metal binding residues emanating from loops in between the β-sheets. The dimerisation of MPH is facilitated by domain swapping of N-terminal domains of the two subunits. The N-terminal arm of each monomer extends into the adjacent two-fold related subunit, suggesting that the dimer is not a crystallographic artefact. Structurally, WBC3-MPH is similar to B3 MBLs such as L1 from *Stenotrophomonas maltophilia* and AIM-1 from *Pseudomonas aeruginosa,* and MBL superfamily enzymes such as human glyoxalase II and *E. coli* zinc phosphodiesterase (ZiPD). **Figure 1.14** shows examples of MBL superfamily enzymes and the conservation of the αββα fold among the superfamily members.

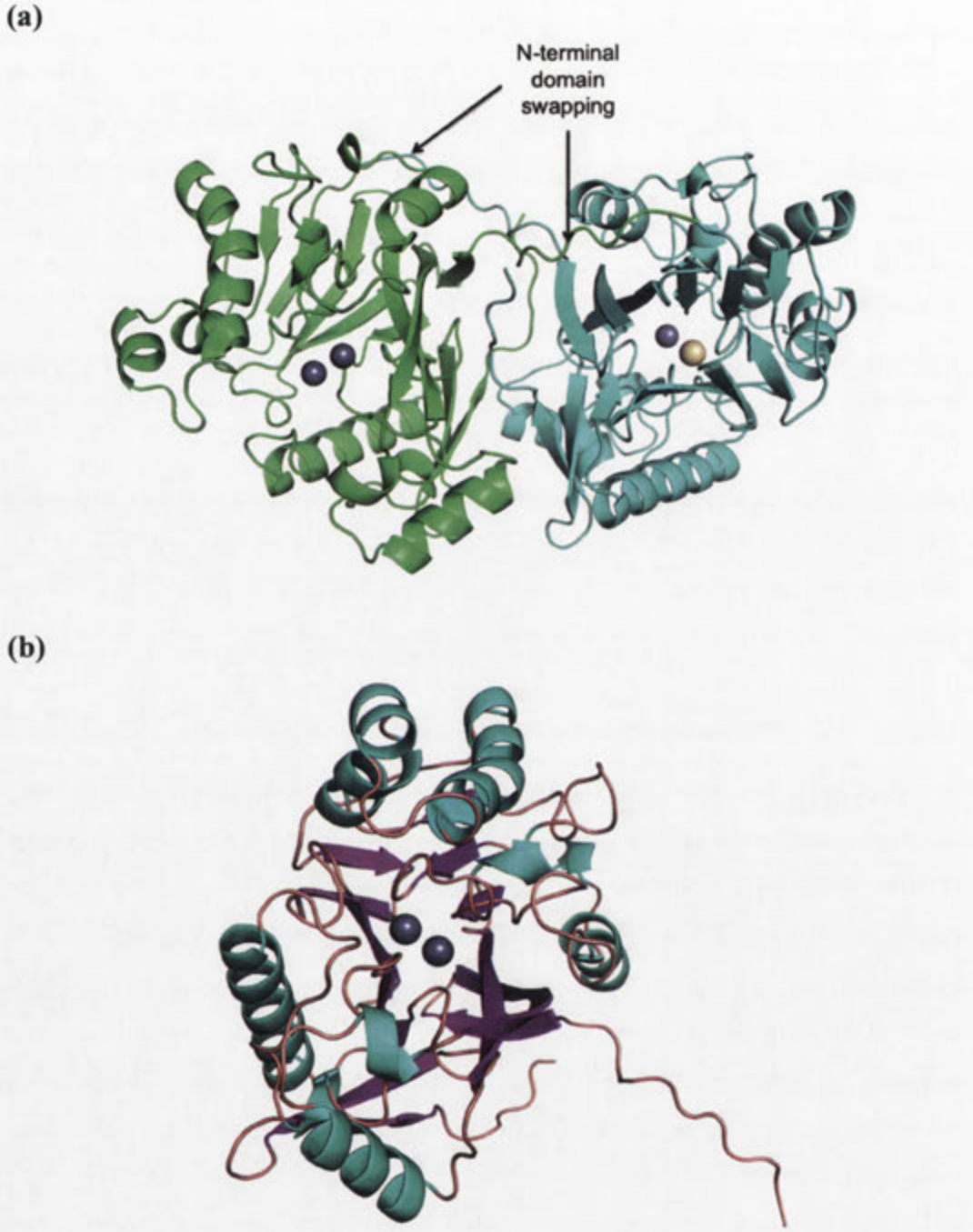


Figure 1.13: Overall structure of MPH (PDB ID: 1P9E).
(A) illustrates the homodimeric structure of MPH. The N-terminal arm of each monomeric unit interacts with the other monomer to form the dimerisation interface. The monomeric units are shown in cyan and green. Zinc (Zn^{2+}) metals are shown in grey and the cadmium metal (Cd^{2+}) is shown in wheat. (B) illustrates the top view of the monomeric MPH. The metals are located in between the parallel β-sheets. The β-sheets are flanked by solvent exposed α-helices.

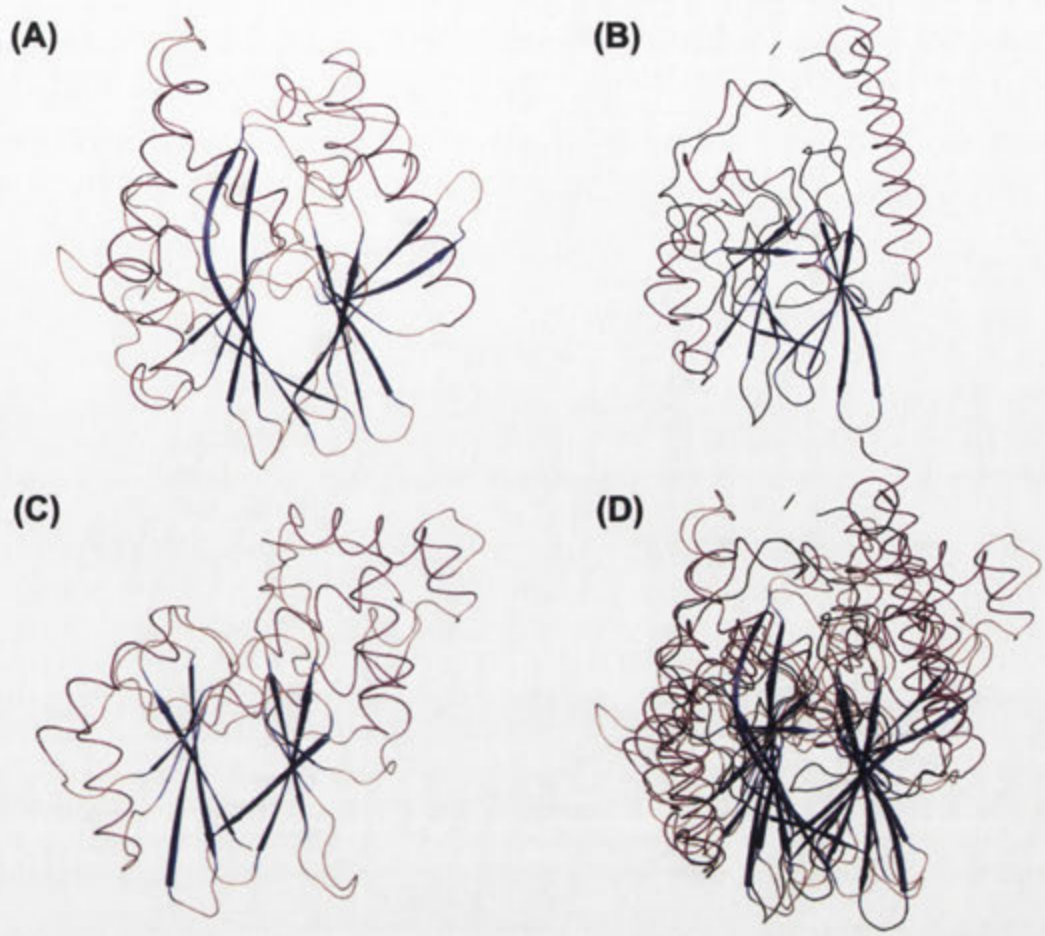
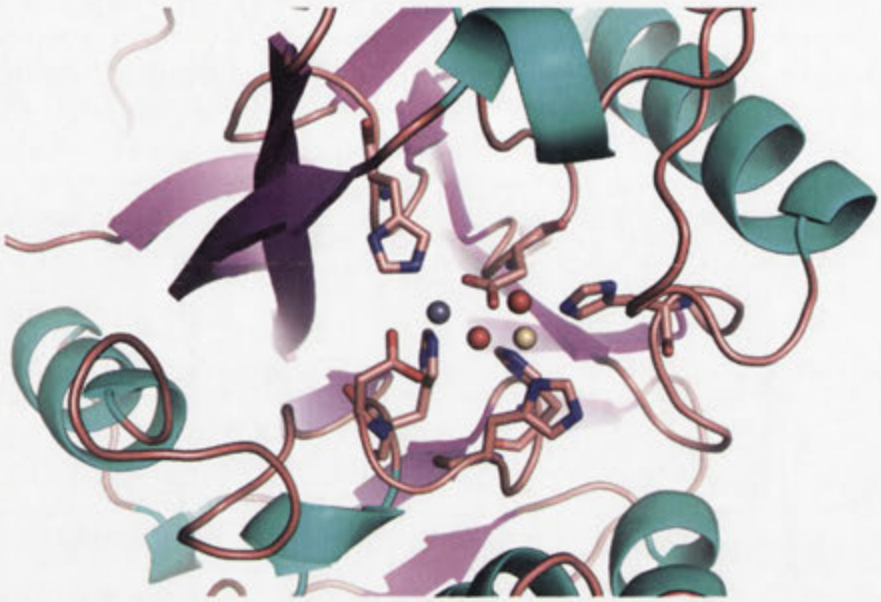


Figure 1.14: Overall structure of MBL superfamily enzymes.
(A) illustrates the structure of MPH (PDB ID: 1P9E). (B) illustrates the structure of AIM-1 lactamase (PDB ID: 4AWY). (C) shows the structure of human glyoxalase II (PDB ID: 1QH5). (D) shows the structural superimposition of the 3 enzymes. The superimposition reveals the conservation of the αββα fold among the superfamily members.

As noted above, the active site of WBC3-MPH has a binuclear metal centre and is situated between the two central β-sheets. The metals have been identified spectroscopically to be zinc but the crystallographic data revealed that they have two different metals.[59] Dong *et al.* reported each MPH subunit has a hybrid Cd^{2+}/Zn^{2+} centre. Since, MPH could only be crystallised in the presence of Cd^{2+}, it was concluded that Cd^{2+} displaced the more solvent exposed β-Zn^{2+} during crystallisation.[59] However, upon closer inspection of the coordinates obtained from the protein databank, the structure reveals a Zn^{2+} binuclear centre in one subunit and a hybrid Cd^{2+}/Zn^{2+}centre in another subunit. As shown in **Figure 1.15**, a water molecule and Asp255 bridge the two metals. Asp151, His152 and His302 coordinate the α-metal while His147, His149, His234 and a terminally bound water coordinate the β-metal. The more buried metal (α-metal) has a trigonal bipyrimidal coordination, while the more solvent exposed metal

(β-metal) is coordinated with an octahedral geometry. The putative substrate-binding pocket is hydrophobic and is formed by Leu65, Leu67, Phe119, Trp179, Phe196, L258 and Leu273. Phe119, Trp179 and Phe196 were suggested as possible substrate-binding residues. The side chain of these residues may form π-stacking interactions with the aromatic *p*-nitrophenolate leaving group of MPS.[59]

(A)



(B)

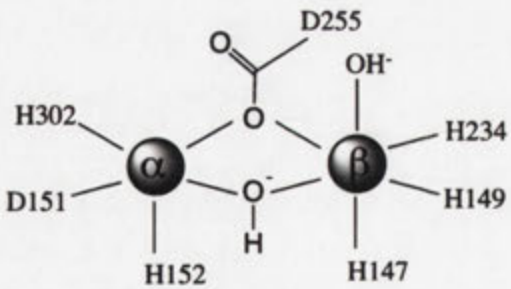


Figure 1.15: **Metal coordination centre of MPH active site.**
(A) shows the location of the metal centre in MPH. The metal centre is located between the β-sheets and the metal coordinating residues emanating from the loops in between β-sheets. The Zn^{2+} is shown in grey, Cd^{2+} in wheat and water in red spheres. (B) shows the schematic of the MPH metal centre.

It is worth noting that, despite the lack of sequence and structural homologies between MPH and *opd* type hydrolases, there are similarities between the binuclear metal centre in WBC3-MPH and those found in OPH and OPDA. The α-metals are coordinated by an aspartate that also forms a hydrogen bond with the bridging

hydroxide in the active site. In addition, several histidine residues coordinate the metals in OPDA and OPH, as is the case for MPH. The metals are also bridged by a carboxylated lysine and a hydroxide. In MPH, OPH and OPDA, hydrophobic residues line the active site pockets.

1.3.2.3. MPH mechanism

Since the MPH and OPH share a number of common features in the active site metal centre, as well as the ability to hydrolyse OPs at near diffusion limited rates, it is generally assumed that the two enzymes also share mechanistic similarity. On this basis, Dong *et al.* proposed a catalytic mechanism that is based on the mechanism proposed by Benning *et al.* for OPH with paraoxon.[45,59] Detailed mechanism for OPH, OPAA, PON1 and DFPase can be found in a review by Bigley and Raushel.[29] It should be noted that the choice of nucleophile has been a subject of speculation. Work done by Jackson *et al.* provided some evidence to the alternate view that a terminally bound water to the α-metal is the nucleophile, and that the bridging hydroxide acts as a base to activate it for attack in OPDA.[63,64] The two possible mechanisms for MPH are summarised in **Figure 1.16**.

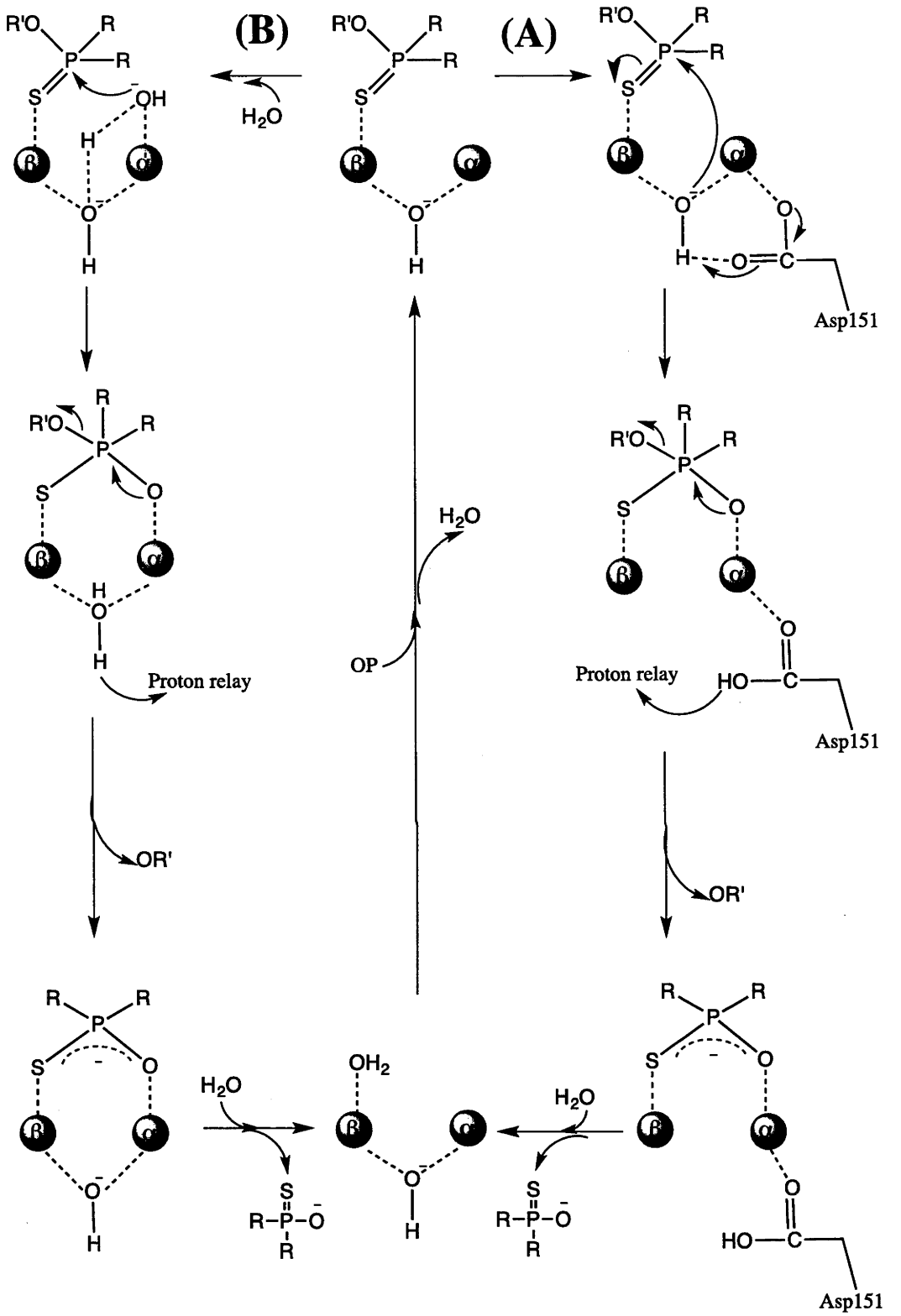


Figure 1.16: Proposed mechanisms for OPH, OPDA and MPH.
Two mechanisms have been proposed for the catalytic hydrolysis of organophosphates (OP) by OPH, OPDA and MPH. Mechanism (A) utilises the bridging hydroxide as the nucleophile while mechanism (B) utilises a terminally bound hydroxide as the nucleophile.

1.3.2.4. Genetic organisation of MPH

The genes coding for the various MPHs are all designated *mpd,* and they were all isolated in China, with the exception of *Burkholderia cepacia mpd* that was isolated in Thailand. [47,48,51–53] Most of the *mpd* genes isolated are found in the bacterial genome, with the exception of *mpd* isolated from *Pseudomonas* sp. WBC-3. This *mpd* gene is located in a 70 kb plasmid[65] and the gene product, WBC3-MPH, is 98% and 100% identical to two other MPH variants, where the *mpd* genes are located in the genome.[48,49] This observation suggests that the widespread distribution of *mpd* genes is possibly due to horizontal gene transfer aided by mobile genetic elements. Liu and co-workers investigated the MPS and *p*-nitrophenol hydrolysing capabilities of *Pseudomonas* sp. WBC-3 and found that the plasmid-cured bacteria could not catabolise MPS and *p*-nitrophenol, confirming the location of the *mpd* gene. They also identified the plasmid harbouring *mpd* to be a class I transposon, as the *mpd* gene is flanked by two copies of insertional sequence (**Figure 1.17a**).[65] Wei *et al.* demonstrated the transposibility of the plasmid, supporting the hypothesis that horizontal genetic transfer was responsible for the distribution of *mpd* genes.[66] In the search for evidence that horizontal gene transfer had taken place, Zhang *et al.* found that the *mpd* gene clusters from the genome of seven bacterial strains isolated from MPS contaminated soil were conserved.[67,68] These bacteria belong to the genera of *Pseudaminobacter* sp., *Achromobacter* sp., *Brucella* sp. and *Ochrobactrum* sp. Computer analysis revealed that the region contains five open reading frames (ORFs); two on one strand and three on the complementary strand. The cluster contains *mpd,* a *tnp* and three other ORFs that highly resemble permease component of ABC-type transport system (**Figure 1.17b**). The distribution of highly similar *mpd* genes in both bacterial genomic DNA and plasmid DNA is similar to that observed for *opd* and *opdA,* the genes coding for OPH and OPDA respectively.[69–71] The gene clusters of *Flavobacterium* sp. *opd* and *Agrobacterium radiobacter opdA* were analysed and transposon elements were found.[72,73] The genetic organisation of *opd, opdA* and *mpd* are shown in **Figure 1.17.**

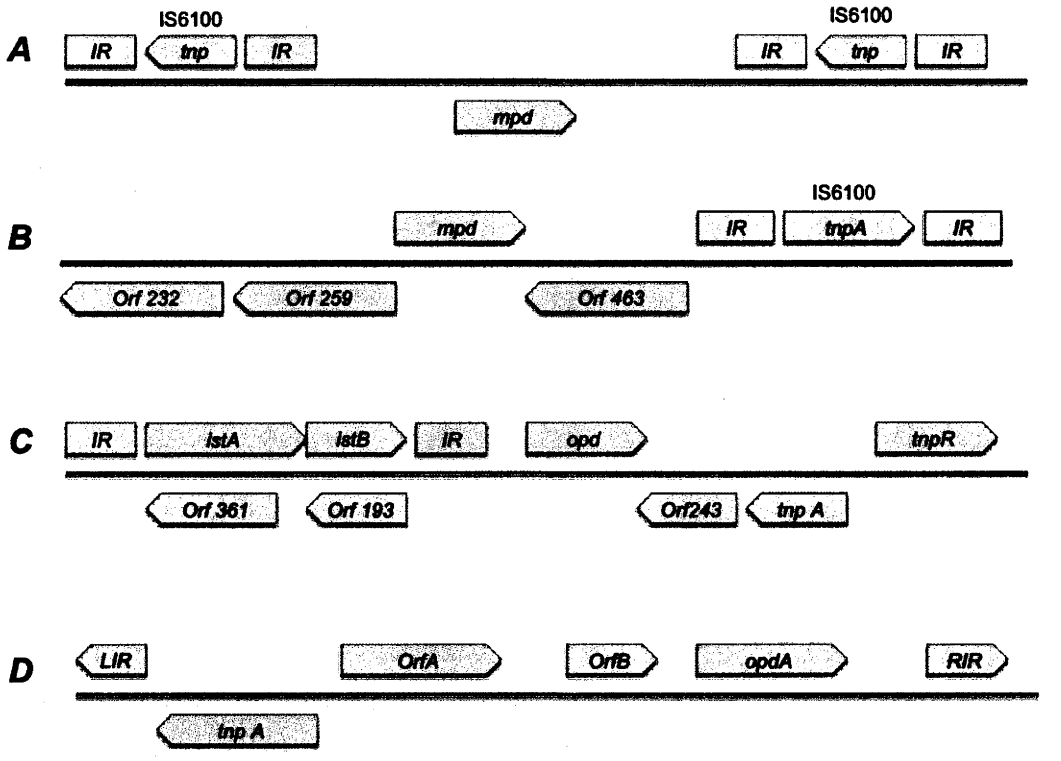


Figure 1.17: Genetic organisation of known organophophorus-degrading genes.
The arrowheads indicate the location and the direction of transcription. (A) Gene clusters of *mpd* located in a catabolic plasmid isolated from *Pseudomonas* sp. WBC-3. The *mpd* gene is flanked by two IS6100 elements, each containing a gene encoding transposase (*tnp*) that is flanked by inverted repeats.[65] (B) Gene clusters of *mpd* found in the genome of *Ochrobactrum* sp. The cluster includes *mpd,* inverted repeats flanking a transposase encoding gene (*tnpA*) and three ORFs of unknown function.[67] (C) Gene clusters of *opd* found in a plasmid isolated from *Flavobacterium* sp. The cluster includes a complete *istAB* operon, *opd, tnpA* and *tnpR* that codes for transposase and resolvase, o*rf243* that codes for metabolite utilization and 2 other *orfs* of unknown function.[73] (D) Gene clusters of *opdA* found in the genome of *Agrobacterium radiobacter* P230. The cluster includes *tnpA, opdA* and two *orfs* of unknown function.[72]

1.4. Protein engineering

Enzymes have been studied for decades, and rather glowing descriptions of their properties have been given in many basic biochemistry texts. However, these descriptions often present a rather incomplete picture and leave the reader wondering why these marvellous catalysts are not used more frequently in the industry. For this reason, it is worth spending a little time looking at how enzymes evolve in nature, and how this process place limitations on their usefulness in practical applications.

Enzymes have evolved over millions of years with mutation and genetic recombination, providing sizeable gene pools from which exceptional characteristics are selected for propagation. This process is repeated over many generations so that highly optimised enzymes would be produced – enzymes ideally suited for cellular functions. However, it should be noted that natural evolution occurs at the level of the organism and that, unless an enzyme gives the organism a selective advantage, it will only evolve to the point where it satisfies the need of the cell. Thus, it may fall short of being suitable for industrial applications. In short, an enzyme evolves to the point where it satisfies the need of the cell and no more. For example, a highly regulated enzyme in a pathway may only function when required – under a specific set of conditions, and these conditions may not be appropriate for practical applications outside the cell. Increased activity can be generated in several ways. Rather than increasing the catalytic efficiency of the enzyme, increasing the amount of enzyme produced in the cell will give the desired outcome. Alternatively, increasing the stability will effectively increase the activity displayed by the cell. Suppose that an enzyme is not crucial for the survival of an organism and that relatively low levels of activity are required for survival. Over the course of evolutionary time, the properties may alter – the expression may drop, solubility might decrease, but as long as the needs of the organism are satisfied, then the properties of the enzyme can drift in directions that may be detrimental for industrial applications. In short, enzymes are often not suitable or optimal for industrial applications due to limitations such as instability in reaction conditions required in large-scale applications. Furthermore, despite the fact that there are an enormous number of enzymes, there are many reactions for which an enzyme cannot be identified.

The idea of improving or engineering an enzyme is not a novel idea, but the technology to alter an enzyme has only been available for a relatively short period of time. The advent of molecular biology and DNA recombinant technology offer avenues to modify proteins in a specific way. Initially these developments led to studies that were aimed at understanding protein structure function. Residues were mutated site specifically and changes in activity were monitored. Eventually, the idea of modifying protein in a useful way emerged, and the term "protein engineering" was coined.[74,75] Early studies were directed at improving the stability of enzymes used in laundry detergent, but as the tools improved, the objectives became more ambitious. However, it was not until mid 90s that the tools necessary for artificial evolution became available. It should be clear, from this brief history of protein engineering, that the field is relatively new, and that it should not be surprising that detailed protocols for improving specific aspects of an enzyme function are not available. There are three general approaches to engineering proteins: the rational approach, the semi rational approach and the evolutionary approach.

1.4.1. Rational and semi-rational approaches

The rational approach relies upon computational calculations such as molecular dynamics[76] that enable predictions of molecular behaviour to be made. In principle, these techniques can be applied without the structure known, but are usually applied to a protein structure that has been obtained experimentally. These studies are usually restricted to peptides or small proteins and require significant amount of computing power. These techniques have not been applied in the present study and will not be described further.

The term "semi-rational" is meant to imply the use of approaches that draw upon the significant amount of structural and sequence data that has been accumulated for proteins. Analysis of sequence and structural databases is generally referred to as "bioinformatics" – this type of study can involve very sophisticated statistical analysis of very large libraries that is beyond the scope of present work. Hogewed and Hesper first coined the term "bioinformatics" in 1970s, referring to the study of informatic processes in biotic systems.[77] Many of the bioinformatics approaches have been incorporated into softwares and web servers that are readily available and can be of

some value to inexperienced users. Some of these programs have been used in this study and references are made in the next chapter.

1.4.2. Directed evolution

Directed (molecular) evolution is the term used to describe evolutionary approaches to protein engineering – it mimics natural evolution although it is applied to a single molecule, and it can be quite rapid compared to natural evolution. One of the earliest examples of directed evolution can be traced to the *in vitro* evolution of self-replicating nucleic acid by Mills *et al.* in the late 1960s.[78] But it was not until the invention of polymerase chain reaction (PCR) decades later that the concept of directed evolution gained traction, and directed evolution is widely used today to improve proteins. Directed evolution usually involves two processes: *generating diversity* (through mutagenesis and/or DNA recombination) and *screening/selection.* One complete cycle is usually called one round of evolution. The mutants chosen at each round usually display the greatest improvements towards the chosen characteristics and are used as parents for the next round of evolution. These processes are repeated until desirable properties or convergence is obtained (**Figure 1.18**).

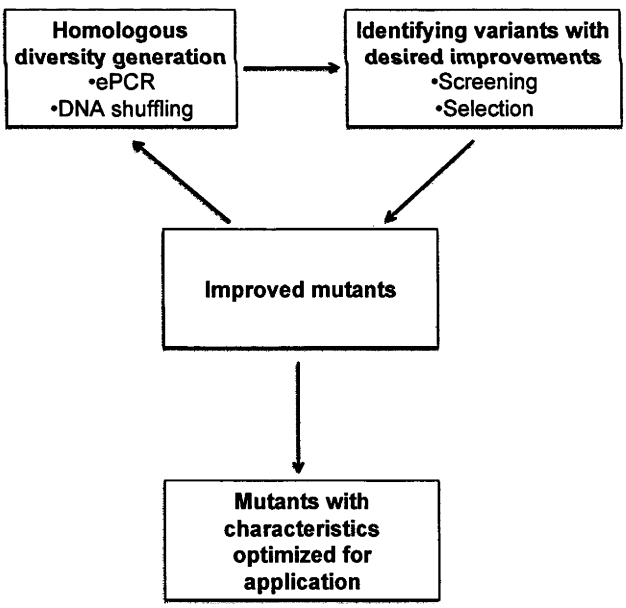


Figure 1.18: The general process of directed evolution.

1.4.2.1. Generating diversity

The polymerase chain reaction (PCR) was invented in 1985 while error-prone PCR (ePCR) was first used about five years later.[79–81] ePCR is an easy to use technique to introduce diversity for library generation – the amplified full-length gene (or part thereof) of an enzyme is subject to random mutagenesis. The general principle of ePCR is shown in **Figure 1.19.** The experimental details of introducing mutations into genes will be discussed in **Chapter 2.** In addition to random mutagenesis, DNA shuffling (*in vitro* sexual recombination) can be used to generate diversity. This technique involves fragmentation of DNA and reassembly of the fragments. This creates new genes with various combinations of mutations, thus accelerating the evolutionary process. The two main techniques in recombining genes are Stemmer's DNA shuffling[82] and Zhao's staggered extension process (StEP)[83]. Stemmer first used shuffling to increase the diversity of a gene library, but this method requires large quantities of DNA and is difficult to implement. For this reason, StEP was used in this work – the experimental details are given in **Chapter 2.** To further increase the efficiency of introducing homologous diversity, an elegant one-step approach that integrates ePCR and StEP, which consist of many cycles of denaturation and extremely short extension times, was developed in our group. **Figures 1.20** and **1.21** show the two main methods of DNA shuffling.

The ideal mutation rate for directed evolution experiments is debatable. Choosing an error rate for evolution is a matter of balancing a number of considerations, namely the possible combinations of a library for a given mutation rate, the proportion of inactivated mutants due to accumulation of mutations and the efficiency of introducing homologous diversity. Most single mutations are deleterious to the enzyme's apparent function and high mutation rate would render most of the mutants generated inactive. However, if the desired character is far from the enzymes physiological character, higher mutation rates may be required to bring about activity enhancement. It should be noted that although mutant libraries with high mutation rates can be generated, these libraries would quickly become too large to screen. By keeping the average mutation rates between one and three amino acids per gene, one can be confident of being able to screen a reasonable fraction of the theoretical library size that consists of a reasonable number of active mutants. This requires several evolution

rounds before the desired level of change is identified, and it assumes that mutations are additive and they can act independently of one another.

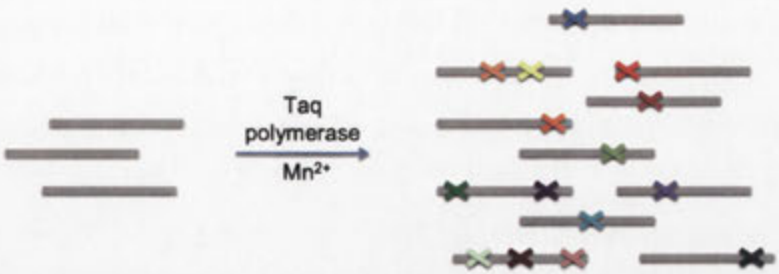


Figure 1.19: Error-prone PCR.
Mutations are introduced to the gene by including Mn^{2+} to a PCR mixture. The inclusion of Mn^{2+} reduces the DNA replication fidelity of Taq polymerase by promoting base pair mismatches.

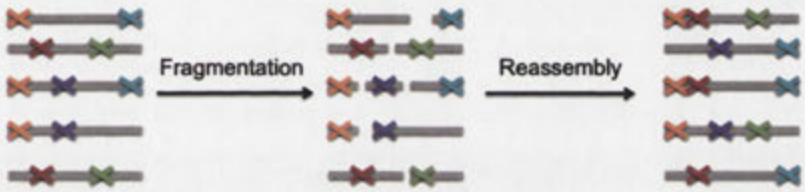


Figure 1.20: DNA shuffling by fragmentation and reassembly.
Homologous genes containing different mutations are first fragmented with DNase I. The fragmented DNA is then reassembled by primerless PCR.

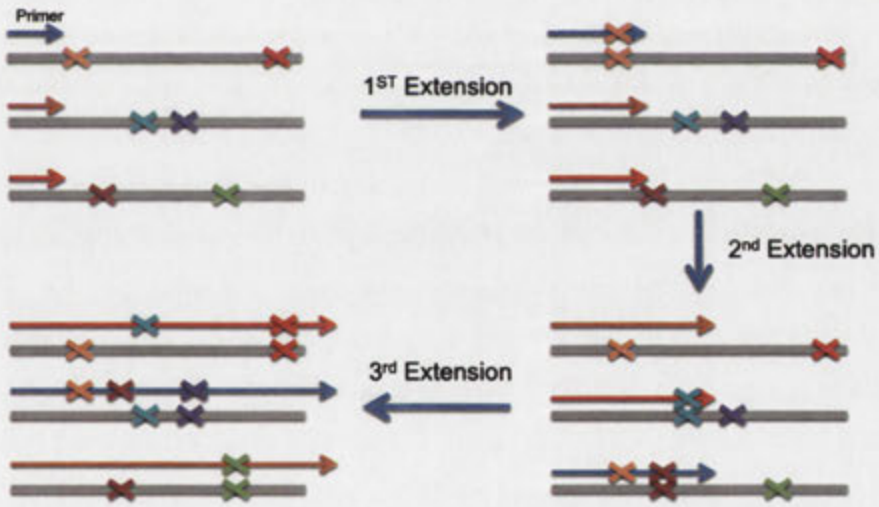


Figure 1.21: DNA shuffling by staggered extension process (StEP).
Homologous genes containing different mutations are amplified using very short cycles of denaturation and extension. This allows partially extended PCR product to switch template and pick up mutations on a different template molecule. The PCR cycles are repeated until the full-length gene product is formed.

1.4.2.2. Screening and selection

The techniques for generating mutant libraries can be transferred from one system to another, although some "tweaking" of the experimental parameters may be necessary. In addition, the methods used to screen libraries or select mutants are very enzyme dependent, and both are often very challenging processes, frequently being the limiting steps in directed evolution. In selection method, the target enzyme is often tied to the survival of the cell so that the mutants carrying the desired function can be readily identified, as is the case of antibiotic resistance. Selection methods are generally less laborious and allow for larger libraries to be evaluated. However, due to the fact that the enzyme activity has to be linked to the host's survival, it is not always possible to employ selection as the evaluation method.

A screen, involving the direct measurement of the enzymatic activity of a protein of interest, is more versatile than selection. The throughput for a screen is usually lower than that for selection. This can be done spectrophotometrically if the product, substrate or co-factors are fluorescent or absorb in the visible or UV range. Coupling with another reagent or enzymatic reaction may be required for reactions that do not show spectrophotometric changes. For example, liberation of thiols can be monitored from the formation of 2-nitro-5-sulfanylbenzoic acid upon reacting the thiol with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB). 2-nitro-5-sulfanylbenzoic acid is a yellow compound and can be observed using a spectrophotometer.[84] The formation of an alcohol can be monitored by coupling the alcohol formation to the depletion of NAD^+ catalysed by alcohol dehydrogenese. There are several techniques for observing spectrophotometric changes during screening. For an enzymatic reaction with coloured product, selection of bacterial colonies can be done directly on agar plate or via top agar plate assay based on the size of the haloes formed as a result of substrate hydrolysis. For changes that cannot be observed visually, colonies can be cultured in 96-well microtitre plates and assayed using a plate reader. This method has been used throughout the present work.

1.4.2.3. Applications of directed evolution

Traditionally, one thinks of reaction rate acceleration when one thinks of an enzyme. In this regard, directed evolution has produced enzymes with increased k_{cat}. In some instances, this has involved lifting the activity of an enzyme towards a promiscuous substrate. However, in addition to the k_{cat} of an enzyme, there are many other enzyme properties that are important for practical applications. For example, the K_m of an enzyme is crucial for its ability to catalyse a reaction at a given substrate concentration. The stability, solubility and ease of expression are clearly important factors for industrial application because they will dramatically affect the cost and convenience of an operation. Again, directed evolution has been successful in improving these enzyme characteristics.[85–89] Similarly, directed evolution has been used to enhance the activity of an enzyme in an organic solvent, as organic solvent is required in some processes. One such example is the lipase-catalysed transesterification in biodiesel production. Lipase catalyses the transesterification reaction instead of the hydrolysis reaction in high methanol environment. Increasing the stability of lipase in high methanol would reduce the cost of using lipase by reducing alcohol-induced denaturation.[90] There are also other applications such as altering or improving the enantioselectivity of an enzyme. May and co-workers demonstrated that directed evolution can be used to invert the enantioselectivity of an enzyme[91] while Liebeton and colleagues improved the enatioselectivity of a bacterial lipase dramatically using a chiral model substrate.[92]

1.4.2.4. Advantages of directed evolution

The main advantage of directed evolution over rational approaches is that it does not require prior knowledge of enzyme structure and mechanism. It can also provide multiple and independent solutions to a problem and can lead to the discovery of beneficial mutations that are unlikely to have been revealed by simple inspection of an enzyme structure or by the use of computational approaches. This is supported by the discovery of beneficial non-active site mutations in various directed evolution experiments to alter enzyme substrate specificity and to increase catalytic activity.[93,94] However, it is also possible that theoretical methods can arrive at possible beneficial mutations that do not arise during the course of artificial evolution. However, such mutations can be made site specifically and shuffled with mutated genes that have been obtained from directed evolution experiments. It is not a matter of choosing between

theoretical methods and directed evolution, but rather of incorporating both approaches so that the benefits of both approaches can be reaped while searching for improved mutants.

1.4.2.5. Directed evolution of phosphotriesterases

OPH, OPDA and more recently MPH have all been the subjects of directed evolution experiments to improve their activity and to alter their substrate specificity. Cell surface display methods have been applied to phosphotriesterases as one of the bioremediator deployment methods and have met with some success.[95–99] Liu and co-workers demonstrated that by displaying OPH and MPH-green fluorescent protein (GFP) fusion on the surface of *E. coli*, the substrate specificity of the whole-cell bioremediator can be broadened, as well as being used as a biosensor for OP detection.[97] In directed evolution, cell surface methods have been applied to OPH to obtain improved activity towards methyl parathion[100] and chlorpyrifos[101] – activity enhancements of 25 and 725 folds towards methyl parathion and chlorpyrifos respectively, were achieved. Directed evolution has also been utilised to improve the activity of OPH towards that of OPDA for coumaphos-o-analogue, methyl paraoxon, methyl parathion and demeton. In the process, a comparison between natural and directed evolution of OPH into OPDA was made.[38] In another directed evolution of OPH, Griffiths and Tawfik isolated a mutant with 63-fold increase in k_{cat} by utilising *in vitro* compartmentalisation (IVC) that allowed the screening of very large libraries. A library of 3.4×10^7 OPH mutants was screened by immobilising the mutant genes to microbeads. The microbead-bound mutant genes were translated *in vitro* in water-in-oil emulsion droplets and a paroxon analogue containing was added into the emulsion. The hydrolysis product formed was then bound to the microbeads and subsequently; a fluorescent-labelled anti-product antibody was added to bind the microbead-bound product and the beads were sorted using flow cytometry.

In our group, a screening method based on cell growth has been developed for directed evolution of OPDA in *E coli*.[84] This was achieved by constructing a short degradation pathway in *E. coli* through co-expression of both OPDA and GpdQ. *E. coli* is unable to grow on medium with paraoxon as the sole phosphorus due to its inability to metabolise phosphodiester. By co-expressing GpdQ, the phosphodiester is metabolised to phosphomonoester that can then be converted *in vivo* to inorganic phosphate, effectively coupling the cell growth to its ability to scavenge inorganic

phosphate. OPDA variants with improved expression were successfully isolated when this selection method was employed in directed evolution.[89] A recent study by Ghanen *et al.* suggested that cell growth rate using this system is dependent on the expression level of the variant and the rates of hydrolysis of the organophosphate and the phosphodiester products by OPDA and GpdQ respectively.[102]

MPH is the newer phosphotriesterase isolated. Only a few known reports of MPH evolution are found in the literature. Two reports for evolving MPH to have higher activity towards poor substrates were found. Xie and co-workers subjected MPH from *Pseudomonas stutzeri* strain ZK-5 to one round of error-prone PCR and found a mutant with five-fold increase in k_{cat} for chlorpyrifos hydrolysing efficiency.[103] Yeong and colleagues designed a high throughput screening (HTS) system that utilises a phenolic-responsive transcription activator for directed evolution of OP-degrading enzymes.[104] The system is composed of a cytoplasmically expressed OP-degrading enzyme, a phenolic-responsive transcription activator DmpR, and a green fluorescent protein (GFP) placed under the control of P_{dmp} promoter. The hydrolysis of OPs often produces phenolic compounds such as *p*-nitrophenol. DmpR forms a transcription activation complex by binding the phenolic compounds produced and activates the P_{dmp} promoter, leading to the expression of GFP. The cells can then be sorted using fluorescence-activated cell sorting (FACS), allowing large mutant libraries to be screened. In this approach, the hydrolysing capability of the enzyme is coupled to the fluorescence intensity of host cell. Using this screening method for directed evolution, Yeong *et al.* found an MPH mutant with 100-fold improvement in k_{cat}/K_m for *p*-nitrophenyl diphenylphosphate hydrolysing efficiency.[104]

1.5. Applications of OP-degrading enzymes

With the rapid isolation and discovery of OP-degrading microorganisms and the genes responsible for the ability, there has been much interest in tapping the potential of these organisms and genes for industrial applications. One such application is the bioremediation of OP-contamination. Accumulation of OP pesticides in the environment due to excessive use and accidental spillage poses health risk to humans and animals. It was estimated that OP pesticides account for ~38% of total pesticide used globally.[105] One example of high pesticide usage is the Tick Eradication Program ran by the US government to prevent re-introduction of cattle fever through ticks on cattle imported from Mexico. The program employs 42 vats, each containing about 15000 litres of coumaphos at concentration of about 1600 mgL^{-1}.[106] Conventional decontamination of OP waste involves chemicals such as NaOH and $NaBO_3$ that are corrosive and will in turn generate substantial residual waste.[107] Use of these chemicals will also change the salinity and pH of the soil and may not be very suitable for foodstuff.

Bioremediation presents a cheap and environmental friendly alternative for removing OP compounds from the environment. Currently, there are two main methods of deploying bioremediation – live cell and purified enzyme, with both methods having their own advantages and disadvantages. Mulbry and coworkers demonstrated one successful example of using live cells for bioremediation. They developed a biofilter containing a consortium of microorganisms that can mineralise coumaphos.[108] The biofilter, capable of treating 15,000 litres of OP-containing dip waste in a single batch, was subsequently used in a bioreactor.[106] Other forms of live cell approach include cell surface display of OP degrading enzyme[99], localisation of OP degrading enzyme to periplasm[98,109] and bioengineer of *E. coli* to fully mineralise OP compounds by establishing a short OP degradation pathway[84].

However, the use of live cells in bioremediation has inherent practical issues such as the release of microorganisms into the environment and nutritional requirements to keep the cells alive for a reasonable period of time. As a result, there is a growing interest in using purified enzyme for bioremediation. A commercial product containing OPDA formulation was marketed under the brand name Landguard™ for decontamination of sheep-dip waste.[110] However, large-scale usage of OP-degrading

enzymes in bioremediation is still limited due to their short shelf lives, activity of environmental proteases and trapping of enzyme to other substrates in the soil.[15] To overcome this, numerous molecular evolutions were undertaken to produce more stable enzymes.[111–114] Enzyme immobilisation was also used to increase enzyme stability, as well as to facilitate enzyme recovery and reuse. These approaches drove down the overall cost of using purified enzyme for bioremediation. In conventional enzyme immobilisation, the enzyme was often immobilised to silica[115], agarose[116] and other gel-based carriers[117]. Recent advances in nanotechnology open new avenues in enzyme immobilisation. Nanoparticles offer several advantages including minimum diffusional limitation, maximum surface area per unit mass and high enzyme load.[118] The usefulness of this approach has been established by various works done on immobilising OP-degrading enzyme to nanoparticles.[119,120] Daumann and co-workers demonstrated that glycerophosphodiesterase from *Enterobacter aerogenes* (GpdQ) immobilised to magnetite nanoparticles suffered no activity loss after 120 days and was active after multiple usage cycles.

Other than usage for bioremediation, OP-degrading enzymes have also been used as biosensors for OP contamination and therapies for OP poisoning. Potentiometric[121] approach to measure local pH change, amperometric[122] measurement of electroactive product and optical[123] measurement of chromogenic product were the main approaches used in OP biosensor development. These biosensors have broad linear range and low detection limit. For example, the amperometric biosensor developed by Mulchandani *et al.* has a linear range up to 140 μM and a lower detection limit of 20 nM for methyl parathion. OP poisoning, caused by unintentional exposure, occupational exposure or suicide attempts, has resulted in approximately three million cases of poisoning and 220,000 deaths worldwide annually.[124] As mentioned earlier, current treatment of OP poisoning relies on chemical antidotes such as oximes and atropine, and the effectiveness of these antidotes differs from one OP to another OP compound. Enzymatic therapies were therefore explored to develop efficient therapy for patients with severe OP exposure. OP-degrading enzyme like MPH can be used to clear excess OPs from bloodstream, allowing AChE reactivation to take place. Limited success was achieved by using OPH, OPDA and PON1 to treat OP poisoning in animal models.[125–129] Bacterial PTEs are good candidates for use in clinical therapy due to their capability of hydrolysing several OPs at near diffusion limit. However, to effectively utilise bacterial PTEs in human therapy, immunological reaction due to foreign protein

injection and activity of proteases in blood plasma have to be addressed.[15] On this front, OP-degrading enzymes can be immobilised in a carrier to protect the enzyme against protease activity as well as avoiding immunological reactions.[129] Liposome[130], erythrocyte encapsulations[131] and hollow fibre reactor[128] were used in developing carriers for OP poisoning therapy.

1.6. Objectives and overview of this thesis

OP pesticides are toxic materials that can find their ways into the environment and pose health risks to humans and other higher organisms. There are numerous reviews covering the dangers of OP compounds and the interested reader can consult these for more information. In short, pesticides are found at low levels in the environment and as such present a danger to wildlife and members of the agricultural community. OPs pesticides are also used in some communities as poison for committing suicide. OP chemical weapons are also classified as weapons of mass destruction due to the devastating toxicity they have on humans.

It should be clear from the discussion above that phosphotriesterases are enzymes of great interest, in large part due to their practical application. However, to be able to fully manipulate the enzyme, the enzyme's characteristics and behaviours have to be known. The initial aim of this thesis was to express and biochemically characterise MPH. The gene for the protein had been synthesised so that the protein could be expressed in a functional form. However, during the course of MPH purification, it was found that the protein was degraded and the purified product was not stable for very long. The purification of MPH was optimised and the cause of the degradation was investigated in **Chapter 3**.

The first aspect of MPH that was subjected to directed evolution was its thermostability (**Chapter 4**). While MPH is not particularly unstable, better thermostability is beneficial to an enzyme that is to be considered for deployment in bioremediation and OP poisoning therapies. During the course of the work, it was found that activity and thermostability of the enzyme increased with Zn^{2+} supplementation, leading to the investigation of Zn^{2+} content in purified enzyme, as well as the effect of Zn^{2+} on MPH stability. These results are discussed in **Chapter 4.**

The long-term aim of this work is to evolve the protein to have broader substrate specificity or to produce mutants that have altered substrate specificity. In order to achieve this, better understanding of how MPH functions, particularly the key residues in substrate binding, have to be identified. While there are numerous reports on OPDA and OPH with enhanced activity and altered substrate specificity through directed evolution, directed evolution reports on MPH are scarce. The substrate specificity and substrate binding mechanism are not well understood. The substrate profile of MPH on various OP pesticides was investigated and directed evolution experiments aimed at altering the substrate specificity, as well as enhancing the activity against a poor substrate, were conducted. The substrate profiles of the mutants isolated from the evolution were characterised and the results are discussed in **Chapter 5**.

1.7. References

1. Kamanyire, R. & Karalliedde, L. Organophosphate toxicity and occupational exposure. *Occupational Medicine* **54,** 69–75 (2004).
2. Cleland, W. W. & Hengge, A. C. Enzymatic mechanisms of phosphate and sulfate transfer. *Chemical Reviews* **106,** 3252–78 (2006).
3. Davies, R., Ahmed, G. & Freer, T. Chronic exposure to organophosphates: background and clinical picture. *Advances in Psychiatric Treatment* **6,** 187–192 (2000).
4. Upadhyay, S. *et al.* in *Pesticides-Strategies for Pesticides Analysis* (Stoytcheva, M.) 316–332 (UnTech, 2011).
5. Millard, C. B. *et al.* Crystal structures of aged phosphonylated acetylcholinesterase: nerve agent reaction products at the atomic level. *Biochemistry* **38,** 7032–7039 (1999).
6. Froede, H. C. & Wilson, I. B. Direct determination of acetyl-enzyme intermediate in the acetylcholinesterase-catalyzed hydrolysis of acetylcholine and acetylthiocholine. *J. Biol. Chem.* **259,** 11010–11013 (1984).
7. Bajgar, J. Organophosphates/nerve agent poisoning: mechanism of action, diagnosis, prophylaxis, and treatment. *Advances in Clinical Chemistry* **38,** 151–216 (2004).
8. Weinbroum, A. A. Pathophysiological and clinical aspects of combat anticholinesterase poisoning. *Br. Med. Bull.* **72,** 119–133 (2004).
9. Wilson, I. B. & Ginsburg, S. A powerful reactivator of alkylphosphate-inhibited acetylcholinesterase. *Biochim. Biophys. Acta* **18,** 168–170 (1955).
10. Wilson, I. B. The inhibition and reactivation of acetylcholinesterase. *Ann. N. Y. Acad. Sci.* **135,** 177–183 (1966).
11. Wong, L. *et al.* Mechanism of oxime reactivation of acetylcholinesterase analyzed by chirality and mutagenesis. *Biochemistry* **39,** 5750–5757 (2000).
12. Patočka, J., Cabal, J., Kuča, K. & Jun, D. Oxime reactivation of acetylcholinesterase inhibited by toxic phosphorus esters: in vitro kinetics and thermodynamics. *Journal of Applied Biomedicine* **3,** 91–99 (2005).
13. Radić, Z., Kalisiak, J., Fokin, V. V, Sharpless, K. B. & Taylor, P. Interaction kinetics of oximes with native, phosphylated and aged human acetylcholinesterase. *Chem.-Biol. Interact.* **187,** 163–166 (2010).
14. Worek, F., Thiermann, H., Szinicz, L. & Eyer, P. Kinetic analysis of interactions between human acetylcholinesterase, structurally different organophosphorus compounds and oximes. *Biochem. Pharmacol.* **68,** 2237–2248 (2004).
15. Singh, B. K. Organophosphorus-degrading bacteria: ecology and industrial applications. *Nat. Rev. Microbiol.* **7,** 156–164 (2009).
16. Aldridge, W. N. Serum esterases 1. Two types of esterase (A and B) hydrolysing *p*-nitrophenyl acetate, propionate and butyrate, and a method for their determination. *Biochem. J.* **51,** 110–117 (1953).
17. Afriat, L., Roodveldt, C., Manco, G. & Tawfik, D. S. The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry* **45,** 13677–13686 (2006).
18. Mazur, A. An enzyme in animal tissues capable of hydrolyzing the phosphorus-fluorine bond of alkyl fluorophosphates. *J. Biol. Chem.* **164,** 271–289 (1946).
19. Draganov, D. I. & La Du, B. N. Pharmacogenetics of paraoxonases: a brief review. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **369,** 78–88 (2004).

20. Aldridge, W. N. Serum esterases. 2. An enzyme hydrolysing diethyl *p*-nitrophenyl phosphate (E600) and its identity with the A-esterase of mamalian sera. *Biochem. J.* **53,** 117–124 (1953).
21. Pejin-grubiša, I. in *Lipoproteins - Roles in Health and Diseases* (Frank, S. & Kostner, G.) 431–444 (InTech, 2012).
22. Kuo, C. & Du, B. N. LA. Calcium binding by human and rabbit serum paraoxonases structural stability and enzymatic activity. *Drug Metab. Dispos.* **26,** 653–660 (1998).
23. Harel, M. *et al.* Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. *Nature Structural & Molecular Biology* **11,** 412–419 (2004).
24. Josse, D. *et al.* Oligomeric states of the detergent-solubilized human serum paraoxonase (PON1). *The Journal of Biological Chemistry* **277,** 33386–33397 (2002).
25. Himbergen, T. M. Van, Tits, L. J. H. Van, Roest, M. & Stalenhoef, A. F. H. The story of PON1 : how an organophosphate-hydrolysing enzyme is becoming a player in cardiovascular medicine. *Netherlands The Journal of Medicine* **64,** 34–38 (2006).
26. Hoskin, F. C. G. & Roush, A. H. Hydrolysis of nerve gas by squid-type diisopropyl phosphorofluoridate hydrolyzing enzyme on agarose resin. *Science* **215,** 1255–1256 (1982).
27. Blum, M., Löhr, F., Richardt, A., Rüterjans, H. & Chen, J. C. H. Binding of a designed substrate analogue to diisopropyl fluorophosphatase: implications for the phosphotriesterase mechanism. *Journal of American Chemical Society* **120,** 12750–12757 (2006).
28. Scharff, E. I., Koepke, J., Fritzsch, G., Lücke, C. & Rüterjans, H. Crystal structure of diisopropylfluorophosphatase from *Loligo vulgaris*. *Structure* **9,** 493–502 (2001).
29. Bigley, A. N. & Raushel, F. M. Catalytic mechanisms for phosphotriesterases. *Biochim. Biophys. Acta* **1834,** 443–453 (2013).
30. Cheng, T., Harvey, S. P. & Stroup, A. N. Purification and properties of a highly active organophosphorus acid anhydrolase from *Alteromonas undina*. *Appl. Environ. Microbiol.* **59,** 3138–3140 (1993).
31. Cheng, T., Defrank, J. J. & Rastogi, V. K. *Alteromonas* prolidase for organophosphorus G-agent decontamination. *Chem.-Biol. Interact.* **119-120,** 455–462 (1999).
32. Vyas, N. K., Nickitenko, A., Rastogi, V. K., Shah, S. S. & Quiocho, F. a. Structural insights into the dual activities of the nerve agent degrading organophosphate anhydrolase/prolidase. *Biochemistry* **49,** 547–559 (2010).
33. Graham, S. C., Bond, C. S., Freeman, H. C. & Guss, J. M. Structural and functional implications of metal ion selection in aminopeptidase P, a metalloprotease with a dinuclear metal center. *Biochemistry* **44,** 13820–13836 (2005).
34. Jao, S.-C., Huang, L.-F., Tao, Y. S. & Li, W.-S. Hydrolysis of organophosphate triesters by *Escherichia coli* aminopeptidase P. *J. Mol. Catal. B: Enzym.* **27,** 7–12 (2004).
35. Newcomb, R. D. *et al.* A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7464–7468 (1997).
36. Jackson, C. J. *et al.* Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10177–10182 (2013).

37. Sethunathan, N. & Yoshida, T. A *Flavobacterium* sp. that degrades diazinon and parathion. *Can. J. Microbiol.* **19,** 873–875 (1973).
38. Yang, H. *et al.* Evolution of an organophosphate-degrading enzyme: a comparison of natural and directed evolution. *Protein Engineering* **16,** 135–145 (2003).
39. Jackson, C. J. *et al.* Anomalous scattering analysis of Agrobacterium radiobacter phosphotriesterase: the prominent role of iron in the heterobinuclear active site. *Biochem. J.* **397,** 501–508 (2006).
40. Omburo, G. a, Kuo, J. M., Mullins, L. S. & Raushel, F. M. Characterization of the zinc binding site of bacterial phosphotriesterase. *J. Biol. Chem.* **267,** 13278–13283 (1992).
41. Seibert, C. M. & Raushel, F. M. Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* **44,** 6388–6390 (2005).
42. Roodveldt, C. & Tawfik, D. S. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **44,** 12728–12736 (2005).
43. Benning, M. M., Kuo, J. M., Raushel, F. M. & Holden, H. M. Three-dimensional structure of phosphotriesterase: an enzyme capable of detoxifying organophosphate nerve agents. *Biochemistry* **33,** 15001–15007 (1994).
44. Holm, L. & Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Structure, Function and Genetics* **28,** 72–82 (1997).
45. Benning, M. M., Shim, H., Raushel, F. M. & Holden, H. M. High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from *Pseudomonas diminuta. Biochemistry* **40,** 2712–2722 (2001).
46. Ely, F. *et al.* Enzymatic bioremediation: organophosphate degradation by binuclear metallo-hydrolases. *Current Topics in Biochemical Research* **9,** 63–78 (2007).
47. Chen, Y., Zhang, X., Wang, Y. & Xia, X. Study on *Pseudomonas* sp. WBC-3 capable of complete degradation of methylparathion. *Wei Sheng Wu Xue Bao* **42,** 490–497 (2002).
48. Cui, Z., Li, S. & Fu, G. Isolation of methyl parathion-degrading strain M6 and cloning of the methyl parathion hydrolase gene. *Appl. Environ. Microbiol.* **67,** 4922–4925 (2001).
49. Yang, C., Liu, N., Guo, X. & Qiao, C. Cloning of *mpd* gene from a chlorpyrifos-degrading bacterium and use of this strain in bioremediation of contaminated soil. *FEMS Microbiol. Lett.* **265,** 118–125 (2006).
50. Li, X. *et al.* Diversity of chlorpyrifos-degrading bacteria isolated from chlorpyrifos-contaminated samples. *International Biodeterioration & Biodegradation* **62,** 331–335 (2008).
51. Xiao, W., Chu, X., Tian, J., Guo, J. & Wu, N. Cloning of a methyl parathion hydrolase gene from *Ochrobactrum* sp. *Journal of Agricultural Science and Technology* **10,** 99–102 (2008).
52. Wu, N. *et al.* Isolation, purification and properties characterization of a new organphosphorus hydrolase OPHC2. *Chinese Science Bulletin* **49,** 268–272 (2004).
53. Ekkhunnatham, A., Jongsareejit, B., Yamkunthong, W. & Wichitwechkarn, J. Purification and characterization of methyl parathion hydrolase from Burkholderia cepacia capable of degrading organophosphate insecticides. *World J. Microb. Biot.* **28,** 1739–1746 (2012).

54. Zhang, Z., Hong, Q., Xu, J., Zhang, X. & Li, S. Isolation of fenitrothion-degrading strain *Burkholderia* sp. FDS-1 and cloning of *mpd* gene. *Biodegradation* **17,** 275–283 (2006).
55. Neuwald, A. F., Liu, J. S., Lipman, D. J. & Lawrence, C. E. Extracting protein alignment models from the sequence database. *Nucleic Acids Res.* **25,** 1665–1677 (1997).
56. Daiyasu, H., Osaka, K., Ishino, Y. & Toh, H. Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **503,** 1–6 (2001).
57. Bebrone, C. Metallo-β-lactamases (classification, activity, genetcis organization, strucuture, zinc coordination) and their superfamily. *Biochemical Phamacology* **74,** 1686–1701 (2007).
58. Chu, X., Zhang, X., Chen, Y., Liu, H. & Song, D. Study on the properties of methyl parathion hydrolase from *Pseudomonas* sp. WBC-3. *Wei Sheng Wu Xue Bao* **43,** 453–459 (2003).
59. Dong, Y.-J. *et al.* Crystal structure of methyl parathion hydrolase from *Pseudomonas sp.* WBC-3. *J. Mol. Biol.* **353,** 655–663 (2005).
60. Yang, J., Yang, C., Jiang, H. & Qiao, C. Overexpression of methyl parathion hydrolase and its application in detoxification of organophosphates. *Biodegradation* **19,** 831–839 (2008).
61. Yang, W. *et al.* Application of methyl parathion hydrolase (MPH) as a labeling enzyme. *Anal. Bioanal. Chem.* **390,** 2133–2140 (2008).
62. Sun, L. *et al.* Crystallization and preliminary X-ray studies of methyl parathion hydrolase from *Pseudomonas* sp. WBC-3. *Acta Crystallographica. Section D, Biological crystallography* **D60,** 954–956 (2004).
63. Jackson, C. J. *et al.* In crystallo capture of a Michaelis complex and product-binding modes of a bacterial phosphotriesterase. *J. Mol. Biol.* **375,** 1189–1196 (2008).
64. Jackson, C., Kim, H.-K., Carr, P. D., Liu, J.-W. & Ollis, D. L. The structure of an enzyme-product complex reveals the critical role of a terminal hydroxide nucleophile in the bacterial phosphotriesterase mechanism. *Biochim. Biophys. Acta* **1752,** 56–64 (2005).
65. Liu, H., Zhang, J.-J., Wang, S.-J., Zhang, X.-E. & Zhou, N.-Y. Plasmid-borne catabolism of methyl parathion and *p*-nitrophenol in *Pseudomonas* sp. strain WBC-3. *Biochem. Biophys. Res. Commun.* **334,** 1107–1114 (2005).
66. Wei, M. *et al.* A transposable class I composite transposon carrying mph (methyl parathion hydrolase) from *Pseudomonas* sp. strain WBC-3. *FEMS microbiology letters* **292,** 85–91 (2009).
67. Zhang, R. *et al.* Cloning of the organophosphorus pesticide hydrolase gene clusters of seven degradative bacteria isolated from a methyl parathion contaminated site and evidence of their horizontal gene transfer. *Biodegradation* **17,** 465–472 (2006).
68. Zhang, R. *et al.* Diversity of organophosphorus pesticide-degrading bacteria in a polluted soil and conservation of their organophosphorus hydrolase genes. *Can. J. Microbiol.* **51,** 337–343 (2005).
69. Horne, I., Sutherland, T. D., Harcourt, R. L., Russell, R. J. & Oakeshott, J. G. Identification of an *opd* (organophosphate degradation) gene in an *Agrobacterium* isolate. *Appl. Environ. Microbiol.* **68,** 3371–3376 (2002).
70. Mulbry, W. W. & Karns, J. S. Parathion hydrolase specified by the Flavobacterium *opd* gene: relationship between the gene and proten. *J. Bacteriol.* **171,** 6740–6746 (1989).

71. Serdar, C. M., Gibson, D. T., Munnecke, D. M. & Lancaster, J. H. Plasmid involvement in parathion hydrolysis by *Pseudomonas diminuta*. *Appl. Environ. Microbiol.* **44,** 246–249 (1982).
72. Horne, I., Qiu, X., Russell, R. J. & Oakeshott, J. G. The phosphotriesterase gene *opdA* in *Agrobacterium radiobacter* P230 is transposable. *FEMS Microbiol. Lett.* **222,** 1–8 (2003).
73. Siddavattam, D., Khajamohiddin, S., Manavathi, B., Pakala, S. B. & Merrick, M. Transposon-like organization of the plasmid-borne organophosphate degradation (*opd*) gene cluster found in *Flavobacterium* sp. *Appl. Environ. Microbiol.* **69,** 2533–2539 (2003).
74. Wetzel, R. What is protein engineering? *Protein Engineering* **1,** 3–6 (1986).
75. Leatherbarrow, R. J. & Fersht, a R. Protein engineering. *Protein Engineering* **1,** 7–16 (1986).
76. Pikkemaat, M. G., Linssen, A. B. M., Berendsen, H. J. C. & Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering* **15,** 185–192 (2002).
77. Hogeweg, P. The roots of bioinformatics in theoretical biology. *PLoS Comput. Biol.* **7,** e1002021 (2011).
78. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darvinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 217–224 (1967).
79. Saiki, R. K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354 (1985).
80. Leung, D. W., Chen, E. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1,** 11–15 (1989).
81. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2,** 28–33 (1992).
82. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassemby: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10747–10751 (1994).
83. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16,** 258–261 (1998).
84. Mcloughlin, S. Y., Jackson, C., Liu, J. & Ollis, D. L. Growth of *Escherichia coli* coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Appl. Environ. Microbiol.* **70,** 404–412 (2004).
85. Cherry, J. R. & Fidantsef, A. L. Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* **14,** 438–443 (2003).
86. Bommarius, A. S., Broering, J. M., Chaparro-Riggers, J. F. & Polizzi, K. M. High-throughput screening for enhanced protein stability. *Curr. Opin. Biotechnol.* **17,** 606–10 (2006).
87. Liu, J., Boucher, Y., Stokes, H. W. & Ollis, D. L. Improving protein solubility: the use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expression Purif.* **47,** 258–263 (2006).
88. Ju, J., Misono, H. & Ohnishi, K. Directed evolution of bacterial alanine racemases with higher expression level. *J. Biosci. Bioeng.* **100,** 246–254 (2005).
89. McLoughlin, S. Y., Jackson, C., Liu, J.-W. & Ollis, D. Increased expression of a bacterial phosphotriesterase in *Escherichia coli* through directed evolution. *Protein expression and purification* **41,** 433–440 (2005).

90. Korman, T. P. *et al.* Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution. *Biotechnology for Biofuels* **6,** 70–82 (2013).
91. May, O., Nguyen, P. T. & Arnold, F. H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat. Biotechnol.* **18,** 317–320 (2000).
92. Liebeton, K. *et al.* Directed evolution of an enantioselective lipase. *Chemistry & Biology* **7,** 709–718 (2000).
93. Oue, S., Okamoto, A. & Yano, T. Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *J. Biol. Chem.* **274,** 2344–2349 (1999).
94. Kim, Y.-W., Lee, S. S., Warren, R. A. J. & Withers, S. G. Directed evolution of a glycosynthase from *Agrobacterium* sp. increases its catalytic activity dramatically and expands its substrate repertoire. *J. Biol. Chem.* **279,** 42787–42793 (2004).
95. Shimazu, M., Mulchandani, A. & Chen, W. Cell surface display of organophosphorus hydrolase using ice nucleation protein. *Biotechnol. Prog.* **17,** 76–80 (2001).
96. Yang, C. *et al.* Surface display of MPH on *Pseudomonas putida* JS444 using ice nucleation protein and its application in detoxification of organophosphates. **99,** 30–37 (2008).
97. Liu, R. *et al.* Development of a whole-cell biocatalyst/biosensor by display of multiple heterologous proteins on the *Escherichia coli* cell surface for the detoxification and detection of organophosphates. *J. Agric. Food Chem.* **61,** 7810–7816 (2013).
98. Yang, C., Freudl, R., Qiao, C. & Mulchandani, A. Cotranslocation of methyl parathion hydrolase to the periplasm and of organophosphorus hydrolase to the cell surface of *Escherichia coli* by the Tat pathway and ice nucleation protein display system. *Appl. Environ. Microbiol.* **76,** 434–440 (2010).
99. Wang, X.-X., Chi, Z., Ru, S.-G. & Chi, Z.-M. Genetic surface-display of methyl parathion hydrolase on *Yarrowia lipolytica* for removal of methyl parathion in water. *Biodegradation* **23,** 763–774 (2012).
100. Cho, C. M., Mulchandani, A. & Chen, W. Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of orga. *Appl. Environ. Microbiol.* **68,** 2026–2030 (2002).
101. Cho, C. M., Mulchandani, A. & Chen, W. Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Appl. Environ. Microbiol.* **70,** 4681–4685 (2004).
102. Ghanem, E., Li, Y., Xu, C. & Raushel, F. M. Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry* **46,** 9032–9040 (2007).
103. Xie, J., Zhao, Y., Zhang, H., Liu, Z. & Lu, Z. Improving methyl parathion hydrolase to enhance its chlorpyrifos-hydrolysing efficiency. *Letters in Applied Microbiology* 1–7 (2013). doi:10.1111/lam.12155
104. Jeong, Y.-S. *et al.* High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. *Protein Eng., Des. Sel.* **25,** 725–731 (2012).
105. POST. Organophosphates. *Parliamentary Office of Science and Technology* **122,** (1998).

106. Mulbry, W., Ahrens, E. & Karns, J. Use of a field-scale biofilter for the degradation of the organophosphate insecticide coumaphos in cattle dip wastes. *Pesticide Science* **52,** 268–274 (1998).
107. David, M. D. & Seiber, J. N. Accelerated hydrolysis of industrial organophosphates in water and soil using sodium perborate. *Environmental Pollution* **105,** 121–128 (1999).
108. Mulbry, W. W., Valle, P. L. Del & Karns, J. S. Biodegradation of the organophosphate insecticide coumaphos in highly contaminated soils and in liquid waste. *Pesticide Science* **48,** 149–155 (1996).
109. Yang, C., Song, C., Freudl, R., Mulchandani, A. & Qiao, C. Twin-arginine translocation of methyl parathion hydrolase in *Bacillus subtilis*. *Environmental Science & Technology* **44,** 7607–7612 (2010).
110. Scott, C. *Landguard A900. An enzyme based remediant for the detoxification of organophosphate insecticide in animal dips*. 1–30 (2012).
111. Chu, X., Tian, J., Wu, N. & Fan, Y. An intramolecular disulfide bond is required for the thermostability of methyl parathion hydrolase, OPHC2. *Appl. Microbiol. Biotechnol.* **88,** 125–131 (2010).
112. Su, Y. *et al.* Improving the thermostability of a methyl parathion hydrolase by adding the ionic bond on protein surface. *Appl. Biochem. Biotechnol.* **165,** 989–997 (2011).
113. Tian, J. *et al.* Enhanced thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 by rational engineering of a glycine to proline mutation. *The FEBS Journal* **277,** 4901–4908 (2010).
114. Tian, J. *et al.* Improving the thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 using a computationally aided method. *Appl. Microbiol. Biotechnol.* **97,** 2997–3006 (2013).
115. Singh, a K. *et al.* Development of sensors for direct detection of organophosphates. Part I: Immobilization, characterization and stabilization of acetylcholinesterase and organophosphate hydrolase on silica supports. *Biosensors & Bioelectronics* **14,** 703–713 (1999).
116. Caldwell, S. R. & Raushel, F. M. Detoxification of organophosphate pesticides using an immobilized phosphotriesterase from *Pseudomonas diminuta*. *Biotechnol. Bioeng.* **37,** 103–109 (1991).
117. Efremenko, E. N. *et al.* Addition of polybrene improves stability of organophosphate hydrolase immobilized in poly(vinyl alcohol) cryogel carrier. *J. Biochem. Biophys. Methods* **51,** 195–201 (2002).
118. Jia, H., Zhu, G. & Wang, P. Catalytic behaviors of enzymes attached to nanoparticles: the effect of particle mobility. *Biotechnol. Bioeng.* **84,** 406–414 (2003).
119. Du, D. *et al.* Covalent coupling of organophosphorus hydrolase loaded quantum dots to carbon nanotube/Au nanocomposite for enhanced detection of methyl parathion. *Biosensors & Bioelectronics* **25,** 1370–1375 (2010).
120. Daumann, L. J., Larrabee, J. a., Ollis, D., Schenk, G. & Gahan, L. R. Immobilization of the enzyme GpdQ on magnetite nanoparticles for organophosphate pesticide bioremediation. *J. Inorg. Biochem.* (2013). doi:10.1016/j.jinorgbio.2013.10.007
121. Mulchandani, P., Mulchandani, A., Kaneva, I. & Chen, W. Biosensor for direct determination of organophosphate nerve agents. 1. potentiometric enzyme electrode. *Biosensors & Bioelectronics* **14,** 77–85 (1999).
122. Mulchandani, P., Chen, W. & Mulchandani, A. Flow injection amperometric enzyme biosensor for direct determination of organophosphate nerve agents. *Environmental Science & Technology* **35,** 2562–2565 (2001).

123. Mulchandani, A., Pan, S. & Chen, W. Fiber-optic enzyme biosensor for direct determination of organophosphate nerve agents. *Biotechnol. Prog.* **15,** 130–134 (1999).
124. Jeyaratnam, J. Acute pesticide poisoning: a major global health problem. *World Health Statistics Quarterly* **43,** 139–144 (1990).
125. Bird, S. B. *et al.* OpdA, a bacterial organophosphorus hydrolase, prevents lethality in rats after poisoning with highly toxic organophosphorus pesticides. *Toxicology* **247,** 88–92 (2008).
126. Rochu, D., Chabrière, E. & Masson, P. Human paraoxonase: a promising approach for pre-treatment and therapy of organophosphorus poisoning. *Toxicology* **233,** 47–59 (2007).
127. Stevens, R. C. *et al.* Engineered recombinant human paraoxonase 1 (rHuPON1) purified from Escherichia coli protects against organophosphate poisoning. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 12780–12784 (2008).
128. Massona, P. *et al.* Enzymes hydrolyzing organophosphates as potential catalytic scavengers against organophosphate poisoning. *Journal of Physiology (Paris)* **92,** 357–362 (1998).
129. Sogorb, M. a, Vilanova, E. & Carrera, V. Future applications of phosphotriesterases in the prophylaxis and treatment of organophosporus insecticide and nerve agent poisonings. *Toxicol. Lett.* **151,** 219–233 (2004).
130. Petrikovics, I. *et al.* Antagonism of paraoxon intoxication by recombinant phosphotriesterase encapsulated within sterically stabilized liposomes. *Toxicol. Appl. Pharmacol.* **156,** 56–63 (1999).
131. Pei, L. *et al.* Encapsulation of phosphotriesterase within murine erythrocytes. *Toxicol. Appl. Pharmacol.* **124,** 296–301 (1994).

MATERIALS AND METHODS

2. Materials and Methods

2.1. Preamble

This chapter describes the techniques used in the experimental work performed for this dissertation. It includes the details of commonly used techniques of molecular biology as well as protein expression and purification. The experimental details of enzyme kinetic and the analysis of the data are also given. A reference to softwares used at various stages in this research is also made. Detailed information regarding materials, protocols and equipment used in the research are given in appendices – as noted in the relevant sections below.

In general, all chemicals and reagents used were of analytical grade. For molecular biology and cell culturing work, glassware, microcentrifuge tubes, micropipette tips, growth media, buffers and solutions were sterilised by autoclaving using ASB270BT autoclave (Astell Scientific). Milli-Q Reagent Water System produced the distilled water used in this study.

2.2. Bacterial strains and their growth

The two strains used for this study are described, below along with the procedures used to culture them and prepare them for transformation. **Appendix A** lists the recipes for culture media and buffers used in this work – much of this material was obtained from the laboratory manual for cloning by Sambrook and Russell[1].

2.2.1. Strains

E. coli DH5α is a commonly used laboratory K12 strain. It has a high transformation efficiency, simple preparation and is easily stored. It has a *recA*$^-$ genotype, which eliminates the possibility of recombination between chromosomal and plasmid genes. This strain has an *endA*$^-$ genotype, implying that EndA, a non-specific endonuclease, is not expressed thus eliminating problems associated with plasmid preparation.

BL21 (DE3) *recA*$^-$ is used for high-level protein expression and is required to express genes cloned into T7-based expression vectors (pET vectors). BL21 is a lysogen of bacteriophage λDE3, which contains the *lacUV5* promoter and the gene for T7 RNA polymerase. Upon introduction of IPTG, the expression of T7 RNA polymerase, that is located downstream of the *lacUV5* promoter, will be induced. The T7 RNA polymerase then transcribes the target gene located downstream of T7 promoter in the plasmid.[2] The strain used in this work is a *recA*$^-$ derivative for the reasons described above. BL21 strains also lack the *lon* and *ompT* genes that code for proteases to reduce proteolytic degradation during expression.[3]

2.2.2. Growth media

The growth media used for most culturing in this work was Lysogeny Broth (LB). Unless stated otherwise, a culture is either an LB broth or LB agar. Recipes for the preparation of all media used are listed in **Appendix A**. Each media was supplemented with antibiotics in accordance to the antibiotic resistance that the various plasmids used confer.

2.2.3. Culture growth

All broth cultures used were inoculated from single colonies isolated from LB plates supplemented with suitable antibiotics. Broth cultures larger than 20 mL were subcultured from an overnight starter culture that was 1% of the final culture volume. Cultures smaller than 5 mL were grown in 10 mL tubes or 15 mL tubes, whereas larger cultures were grown in baffled flasks. Unless stated otherwise, most cultures were grown at 37°C with orbital shaking.

2.2.4. Preparation of electrocompetent cells

The procedure for electrocompetent cell preparation was adopted and modified from the laboratory manual by Sambrook and Russell[1]. A single colony was inoculated in 10 mL of YENB **(Appendix A)** broth and was grown overnight at 37°C. 8 mL of the overnight culture was subcultured in 800 mL of fresh YENB and was grown until OD_{600} reaches 0.35 – 0.4. The culture was then rapidly cooled to 4°C and left to stand for 1 hour. From here onwards, the culture was kept at 4°C as much as possible and all

buffers and solutions used were pre-chilled to 4°C. The culture was centrifuged at 5,000 rpm in a VWR R9A rotor, at 4°C, for 10 mins. The supernatant was discarded and the pellet was resuspended gently in 300 mL of dH_2O for washing. The cells were centrifuged and washed again with 300 mL of 10% glycerol. The cells were then centrifuged and the pellet resuspended in 1 mL of 10% glycerol. The cells were aliquoted into 50 μL aliquots and snap frozen in ethanol dry ice bath. The cells were kept frozen at -80°C until they were used. All competent cells prepared were plated on LBA and LBK plates for contamination check, as well as transformed with 1 ng of pJWL1030folA_c11 plasmid for competency check.

2.3. Plasmid DNA and their manipulations

The plasmids used in this work along with the methods used to manipulate them are described in this section. The methods include isolation, digestion and ligation of DNA fragments into plasmid. Techniques involving PCR are described in the next section.

2.3.1. Description of plasmids

pETMCSI

This vector is a modified vector that was derived from T7 pET vectors originally created by Studier.[4] The construction of pETMCSI can be found in the work done by Neylon and co-workers.[5] The relevant modification was the introduction of an *Nde*I restriction site into the multiple-cloning site, downstream of the RBS. Unlike some pET vectors from Novagen, this vector lacks the *lac*I repressor gene, thus allowing leaky expression to occur. This was most likely due to the presence of lactose in the media, as tryptone, a common component in culture media was prepared from enzymatic digest of casein.[6] pETMCSI confers ampicillin resistance to the host. This vector was a gift from Professor Nick Dixon of the University of Wollongong.

pCY76

pCY76 vector was constructed to overexpress non-toxic genes constitutively in *E. coli.*[7] The translation initiation region, RBS and the MCS component were isolated from pND706,[8] while the backbone of the plasmid was formed from pMTL22P[9]. pCY76 is a high copy plasmid that confers ampicillin resistance to the host.

pJWL1030

pJWL1030 was a modified vector constructed from pCY76 and pJJKmf.[10] The constitutive expression cassette was isolated from pCY76, whose construction was described above, and ligated to the backbone of pJJKmf[11]. A leaky *lac* promoter allows constitutive expression of target gene cloned into pJWL1030. Similar to Studier's T7-based pET vectors, it has a strong RBS and translation initiation region to improve protein expression. pJWL1030 confers kanamycin resistance to the host.

pET47b

This vector is a T7-based expression vector that is commercially available from Novagen. It is an IPTG-inducible vector, which uses kanamycin as a selectable marker. The vector also contains an N-terminal his-tag for purification and detection purposes.[12]

2.3.2. Plasmid isolation

Plasmid isolation was done using QIAprep Spin Miniprep kit from Qiagen. The manufacturer's protocol was followed. In brief, 1 – 10 mL of overnight culture was pelleted and the supernatant removed before resuspending in Buffer P1. Then, Buffer P2 was added to lyse the cells, followed by Buffer N3 to neutralise the solution. The mixture was pelleted and the supernatant was loaded to the spin column to bind DNA. The column was then washed with PB buffer, PE buffer containing ethanol and finally eluted in distilled water.

2.3.3. DNA purification

Two types of DNA purification methods were used in this work, namely DNA gel extraction and DNA PCR purification. Whether DNA gel extraction or PCR purification was done would depend on the DNA and overall workflow. DNA gel extraction was done using Wizard® SV Gel and PCR Clean-Up System from Promega. PCR purification was done with either the same kit used for gel extraction or QIAquick PCR Purification Kit from Qiagen. All DNA purification methods were performed according to manufacturers' protocol.

2.3.4. DNA visualisation

All DNA samples were visualised by subjecting the samples to 1% agarose gel electrophoresis. The agarose gel was supplemented with 5 μg/L of ethidium bromide or 1x RedSafe™ DNA Stain during preparation. 1.0 g of agarose was melted in 100 mL of SB buffer **(Appendix A)** and poured onto 125 mm x 80 mm flat glass plate so that the gel was formed by surface tension. Prior to loading, DNA sample was mixed with bromophenol blue loading dye **(Appendix A)** at a DNA: Loading Dye ratio of 6:1. The electrophoresis was performed at 150 V and 300 mA in SB buffer. SB buffer system is preferred over TAE and TBE buffer system due to its cost effectiveness, higher resolution at higher voltage and a cooler system during electrophoresis.[13] After electrophoresis, gels were viewed under weak UV (365 nm), strong UV (254 nm) or blue LED array box for fluorescence. Images of gel were taken using UVI Pro electronic system.

2.3.5. DNA quantification

DNA concentration was determined using NanoDrop® ND-1000 spectrophotometer by observing the absorption at 260 nm. A_{260}/A_{280} and A_{260}/A_{230} ratios of the DNA samples were also noted to detect the presence of protein and carbohydrate contaminants respectively.

2.3.6. Restriction digests

Most of the restriction endonucleases used in this work were purchased from New England Biolabs (NEB). 10 – 20 Units (U) of each restriction enzyme was added to 2 – 3 μg of DNA in 1x manufacturer's recommended buffer and 1x bovine serum albumin. The mixture was incubated at 37°C for 3 hours before being incubated at 65°C for 20 minutes to heat inactivate the enzymes. Digested vector plasmid was then treated with alkaline phosphatase as described in **section 2.3.7**. Digested PCR product was purified without phosphatase treatment.

2.3.7. Alkaline phosphatase treatment

Digested plasmid DNA was dephosphorylated with calf intestinal alkaline phosphatase (CIP) from NEB to prevent self-ligation of the vector. 10 U of CIP was

added to digested plasmid and was incubated at 37°C for 1 hour. After CIP treatment, the DNA sample was purified via gel extraction as described in **section 2.3.3**.

2.3.8. DNA ligation

Ligation of DNA fragments was performed generally with the molar ratio of insert: vector of 3:1. The molar ratio was altered and optimized when the colony count was not satisfactory. Thermo Scientific, Fermentas and NEB supplied the T4 Ligase used in this work. 1.5 – 7.5 U of T4 ligase was added to the vector-insert mixture and the mixture was incubated overnight at 12-16°C. The ligation mixture was purified using PCR purification as described in **section 2.3.3** after overnight incubation to lower the ionic strength of the ligation mixture that would otherwise lower the transformation efficiency.

2.3.9. Bacterial transformation

Electroporation was preferred over heat shock transformation due to its higher transformation efficiency.[14] When the cells are subjected to high intensity electrical field briefly, pores are formed at the cell membrane due to an increase in cell permeability that causes reversible electrical breakdown of the cell membrane.[15] This increase in permeability facilitates the entry of exogenous DNA molecules into the cell.

1 – 5 μL of DNA samples were added gently to 50 μL of thawed electrocompetent cells (competent cell preparation described in **section 2.2.4**) and were mixed thoroughly by gentle flicking or pipetting. The mixtures were then transferred to pre-chilled 0.2 mm electroporation cuvettes. Electroporation was done on a Bio-Rad Micropulser™ using the built in Ec2 preset (V = 2.5 kV, $\tau = \geq 5$ ms). The electroporated cells were immediately resuspended in 1 mL of YENB broth and were incubated in 37°C for 1 hour to allow the cells to recover. 50 μL of cells, with or without dilution, were then plated on LB plates supplemented with antibiotics and grown overnight at 37°C. Electroporation cuvettes were rinsed and washed with distilled water, soaked in 70% ethanol, and then dried for reuse.

2.4. Polymerase chain reaction (PCR)

This section includes basic PCR operations. It includes procedures for general amplification and DNA sequencing. PCR protocols for mutation introduction will be described in the section dealing with directed evolution (**section 2.5**).

2.4.1. Primer design

Oligo Calc Properties Calculator, a web based computational software, was used to estimate the primer melting temperature (T_m) and primer GC content (http://www.basic.northwestern.edu/biotools/OligoCalc.html).[16] All the primers used in this work are listed in **Appendix C**. All the primers used were ordered from GeneWorks or Integrated DNA Technologies (IDT).

2.4.2. Basic PCR

When accurate amplification was required, e.g., building new plasmid construct and long range PCR, high fidelity DNA polymerases such as *Pfu* (Thermo Scientific; Agilent Technologies) and Phusion (Finnzyme; NEB) were used as per manufacturers' instructions. For routine amplification, Taq polymerase (Roche) and BioTaq™ (Bioline) were used. Thermocyclers used were GeneAmp® PCR System 9700 (Applied Biosystems), iCycler® (Bio-Rad), C1000™ Thermocycler (Bio-Rad), MJ Mini Personal Thermal Cycler (Bio-rad) and Veriti® Thermal Cycler (Applied Biosystems). The reaction mixture contained 5 – 30 ng of template DNA, 1 μM of each forward and reverse primer, 1.5 mM $MgCl_2$ for Taq polymerase, BioTaq™ and Phusion or 2 mM $MgSO_4$ for *Pfu* polymerase, 0.2 mM of each dNTP and 1.25 U – 2.5 U of DNA polymerase. The cycling conditions were:

Initial denaturation	94°C (or 98°C for Phusion)	2-5 mins
30 – 35 cycles		
Annealing	Average T_m of 2 primers - 5°C	10-30s
Extension	72°C	1 min/kb DNA amplified for Taq, 2 mins/kb for *Pfu*, 15 – 45 s/kb for Phusion, 15 – 30 s/kb for BioTaq™
Denaturation	94°C (or 98°C for Phusion)	10-30s
Final annealing	Average Tm of 2 primers - 5°C	2 mins
Final extension	72°C	5 mins

2.4.3. Colony PCR

Colony PCR was performed to provide a rapid and easy check for cells harbouring plasmid with the intended insert before proceeding to downstream applications, typically done after transforming a new plasmid construct or a new mutagenic library. A detailed protocol for colony PCR was described by Woodman.[17] In brief, a master mix for the reaction containing 1 μM of each forward and reverse primer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 1.25 U of Taq Polymerase (Roche/Bioline) was prepared. Single colonies were used as templates instead of plasmid DNA. A long initial (5 minutes) denaturation step is important to ensure lysis of the cell and release of the plasmid DNA. The cycling conditions were as described in 2.4.1.

2.4.4. DNA sequencing

DNA sequencing was done in the ACRF Biomolecular Resource Facility (BRF) at The John Curtin School of Medical Research (JCSMR) in the Australian National University. Samples were prepared according to the facility's guidelines given on their website[18]. Each PCR reaction consisted of 150 – 350 ng plasmid DNA, 3.2 pmol primer, 1 μL of BigDye® Terminator and 1 x Sequencing Buffer. Distilled water was added into the mixture to a final volume of 20 μL. The cycling conditions were:

Initial denaturation	94°C	5 mins
30 – 50 cycles		
Denaturation	96°C	10 s
Annealing	50°C	5 s
Extension	60°C	4 mins

The DNA was cleaned up with ethanol precipitation prior to sequencing. A clean up mix consisted of 62.5 μL 95% ethanol, 3 μL 3 M sodium acetate pH 4.6 and 14.5 μL distilled water that was added to the PCR product and mixed thoroughly before being left to stand at room temperature for 15 mins. The sample was then centrifuged at max speed for 30 mins and the supernatant discarded. If the sequencing sample was prepared in a microtitre plate, the supernatant was discarded by tipping inverted plate, and the inverted plate was centrifuged at 300 rpm for 1 min. The pellet was washed with 200 μL of freshly prepared 70% ethanol and was centrifuged for 10 mins. The supernatant was then removed, and the sample was air-dried before being sent to BRF

for sequencing. DNA sequencing chromatograms were viewed using Sequencher 4.10.1 (Gene Codes Corp).

2.5. Directed evolution and protein engineering

Chapter 1 gave a brief introduction to directed evolution. Most of the experimental techniques used for directed evolution and mutagenesis experiments are discussed in this chapter. Details specific to a particular set of experiments are given in the relevant chapters.

There are two main steps in protein evolution – mutagenic library generation and screening. Mutagenic library generation involves the introduction of genetic diversity via ePCR, StEP, StEP-ePCR, and SSM are described in sections below. The resulting PCR product from these mutagenic PCR methods was digested with appropriate restriction enzymes and cloned into desired vector as described in **sections 2.3** and **2.4**. The library produced was then subjected to primary and secondary screen for a desired improvement. A tertiary screening was carried out when necessary to verify the improvement seen.

2.5.1. Error-prone PCR (ePCR)

ePCR was used to introduce random mutations into a gene due to its simplicity. There are numerous approaches in lowering the fidelity of Taq polymerase; we chose to use unbalanced concentrations of $MgCl_2$ and $MnCl_2$. The protocol was adapted from a protocol by Cadwell et al.[19] and Arnold et al.[20]. ePCR has higher $MgCl_2$ (5 mM) than standard PCR (1.5 mM) to stabilise non-complementary pairs due to mutation incorporation while $MnCl_2$ was added to reduce the specificity and therefore fidelity of the DNA polymerase. Increased amount of Taq polymerase (5 U) was used to encourage nucleotide extension beyond the positions of base mismatch.[19]

In this work, the ePCR reaction consisted of 30 ng of plasmid template, 1 μM of each forward and reverse primer, 5 mM of $MgCl_2$, 0.2 mM of each dNTP, PCR buffer without $MgCl_2$ at a final concentration of 1x and 5 U of Taq Polymerase (Roche/Bioline). The reaction also contained 0.05 mM to 0.3 mM of $MnCl_2$ depending on the desired mutation rate. The typical strategy employed in this work was to prepare multiple ePCR reactions with various $MnCl_2$ concentrations, and to pool the ePCR

products to obtain a library with wide range of mutation rates. The cycling condition was identical to standard PCR and was described in **section 2.4.2.**

2.5.2. Staggered extension process (StEP)

StEP is an *in vitro* recombination process and was first described by Zhao *et al.*[21,22] The procedure is much simpler than classical DNA shuffling by utilising shorter extension time during PCR cycle to allow incomplete extension and template swapping. In this thesis, StEP was introduced to shuffle mutations generated by site saturation libraries of various key residues surrounding MPH active site for the work done in **Chapter 5**. Both high fidelity polymerase (Phusion) and Taq polymerase were used in an attempt to increase the shuffling rate of the library.

The StEP reaction consisted of 30 ng of template, 1 μM of each forward and reverse primer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$ for Taq polymerase or 1.5 mM $MgCl_2$ for Phusion and 2.5 U of Taq Polymerase (Roche) or 1 U of Phusion. The cycling conditions were as below:

Initial denaturation	94°C (or 98°C for Phusion)	5 mins
120 cycles		
Denaturation	94°C (or 98°C for Phusion)	10 s
Annealing	55°C	10 s
Extension	72°C	2 s
Final annealing	55°C	2 mins
Final extension	72°C	5 mins

The extension rates for Taq polymerase and Phusion are 1 min/kb and 15-45 s/kb respectively. Taq polymerase was used in later shuffling procedures to increase the shuffling frequency as the extension rate is slower. An extension time of 2 s was used to strike a balance between template swapping and the yield of the reaction.

2.5.3. StEP-ePCR

StEP-ePCR is a one-step procedure that combines both shuffling and error-prone PCR in a single reaction. The protocol was conceived after template switching was observed at very low frequency in ePCR.[23] The PCR reaction recipe and cycling condition was identical to ePCR with the exception of much shorter extension time (5 s) and more reaction cycles (60x). In this work, StEP-ePCR was typically introduced from the second round of evolution onwards, to introduce new mutations in addition to shuffling the mutations obtained during the first round.

2.5.4. Site-specific mutagenesis

Two types of site-specific mutagenesis were performed in this work, site directed mutagenesis (SDM) and site saturation mutagenesis (SSM). The protocols described below are interchangeable for SDM and SSM with the exception - SDM uses primers that contain a specific mutation for the targeted site while SSM uses primers that contain degenerate sequences at the targeted site.

Site-specific mutagenesis was carried out using protocols adapted from the lab manual by Arnold and Georgiou.[24,25] The methods used were sequence overlap extension (SOE)[26], mutagenic plasmid amplification adapted from Stratagene's Quikchange and megaprimed ligase-free site specific mutagenesis.[27,28]

2.5.4.1. Method adapted from Stratagene's Quikchange method (QCM)

Stratagene's Quikchange method (QCM) allows introduction of point mutations without going through routine cloning procedures by amplifying the whole plasmid. A simplified diagram of the method was illustrated in **Figure 2.1a**. A high fidelity polymerase was essential as amplifying long stretches of DNA would be prone to replication errors.

The PCR reaction mixture consisted of 10 – 20 ng of plasmid template, 1 μM of each forward and reverse primer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$ and 2 U of Phusion. The cycling conditions were as follows:

Initial denaturation	98°C	30 s
20 cycles		
Denaturation	98°C	30 s
Annealing	60°C	60 s
Extension	72°C	30 s/kb DNA
Final annealing	55°C	2 mins
Final extension	72°C	5 mins

The PCR product was then treated with *Dpn*I to remove the wild type plasmid template that has methylated DNA and purified with PCR purification as described in **section 2.3.3**. 5 μL of the purified PCR product was then transformed to replicate and repair the DNA nick resulted from the PCR.

2.5.4.2. Sequence overlap extension (SOE)

Sequence overlap extension is a more laborious site-specific mutagenesis procedure. Three PCR reactions, as well as routine cloning procedures such as restriction digestion and ligation are required. Two separate PCR reactions are required to generate the template for a third PCR reaction to assemble the DNA with the desired mutation (**Figure 2.1b**).

A forward flanking primer (denoted A) and a reverse mutagenic primer (denoted C) were used to generate the first fragment (denoted AC). A reverse flanking primer (denoted D) and a forward mutagenic primer (denoted B) were used to generate the second fragment (denoted BD). 10 μL of unpurified AC and BD were then used as templates for the third PCR to assemble the fragments that contained the mutation. All the PCR steps were performed based on protocols described in **section 2.4.2**. The PCR reactions can be performed with any high fidelity polymerase such as Phusion and *Pfu*. The full length PCR product was then subjected to routine cloning procedure.

2.5.4.3. Megaprimed ligase-free site-specific mutagenesis

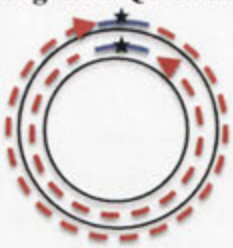
The megaprimed ligase-free site-specific mutagenesis method is similar to Stratagene's QCM, but the method generates a megaprimer first before amplifying the whole plasmid and requires only one mutagenic primer for single site mutagenesis. Tseng et al.[28] and Sanchis et al.[27] independently reported the method, at around the same time, as a better site-specific mutagenesis method than Stratagene's Quikchange method. The method does not require cloning procedures and is more versatile as it can be used to introduce multiple mutations in one round. A simplified diagram of the method was illustrated in **Figure 2.1c** below. In a much earlier work, Miyazaki and Takenouchi reported the megaprimer PCR of whole plasmid (MEGAWHOP) method that can be used for generating error-prone, DNA shuffling and site saturation libraries, if the megaprimer was generated based on one of those respective procedures.[29] While the MEGAWHOP procedure is more general and can accommodate a wide range of library creation methods, it requires two PCR reactions. Tseng *et al.* and Sanchis *et al.* both reported a single-tube protocol that simplifies the procedure.

The PCR reaction mixture contained 30 ng of template DNA, 1x HF buffer, 1.5 mM $MgCl_2$, 0.1 μM of each mutagenic primer and flanking primer, 1 U of Phusion and 0.2 mM of each dNTP. The cycling condition were as below:

Initial denaturation	98°C	5 mins
18 cycles		
Denaturation	98°C	1 min
Annealing	55°C	1 min
Extension	72°C	30 s/kb DNA
25 cycles		
Denaturation	98°C	1 min
Annealing and extension	68°C	45 s/kb DNA
Final extension	68°C	10 mins

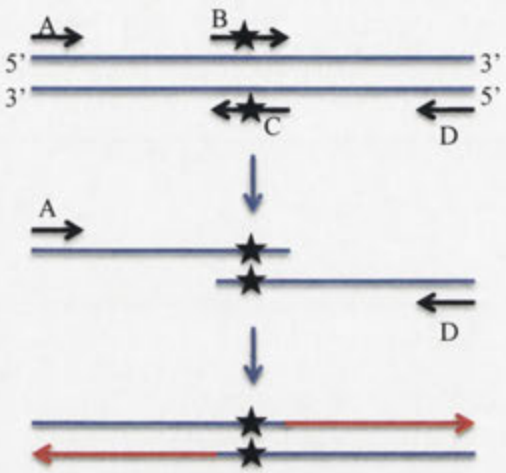
The PCR product was then treated with 20 U of *Dpn*I twice at 37°C for 1 hour each time to remove the template plasmid. The PCR product was then subsequently PCR purified and transformed.

(a) Stratagene's Quikchange



A pair of mutagenic complementary primers were used to amplify the whole plasmid, with the WT plasmid template removed by *Dpn* I digestion

(b) Sequence Overlap Extension

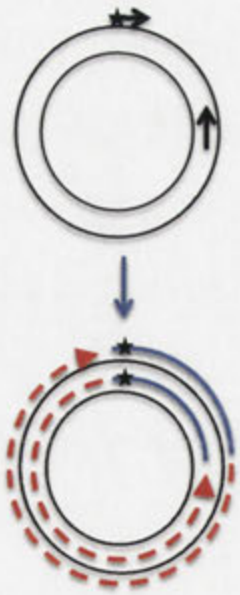


Two PCR reactions are required for SOE. 2 PCR reactions with primers A+C and B +D, with primers B and C being the mutagenic primers

PCR products from the 1st step were used as templates for the overlap extension PCR with flanking primers A+D

Completed mutagenic PCR product was cloned as usual

(c) Megaprimed ligase-free site specific mutagenesis



Requires the generation of a megaprimer by a mutagenic primer and a flanking primer

The megaprimer was then used to amplify the whole plasmid and the template was removed by *Dpn* I digestion

Figure 2.1: Site-specific mutagenesis strategies.
Simplified figures showing different methods used to generate site-specific mutants. (a) Stratagene's Quikchange. (b) Sequence Overlap Extension (SOE). (c) Megaprimer ligase-free site-specific mutagenesis.

2.5.5. Analysis of mutant library

Whenever a mutant library was generated, a few basic analyses such as library size, average mutation frequency and presence of the plasmid harbouring the mutated gene of interest were done to assess the quality of the library.

The library size was quantified by estimating the colony-forming unit (cfu) on agar plates supplemented with antibiotics. Usually, several dilutions of the transformed *E. coli* were plated on agar plates. The library size per transformation was then estimated from the mean of cfu of the agar plates with different cell dilutions. Dilutions that gave too many colonies (>800 cfu) were discounted from the estimation as they might skew the library size estimation.

The average mutation frequency was estimated by looking at the number of mutations in ten randomly selected colonies. The mutation frequency was expressed as the mean and standard deviation of the number mutations found. Colony PCR was also performed as described in **section 2.4.3** to ascertain the presence of plasmid harbouring the mutant gene.

2.5.6. Library screening – direct activity measurement

The primary screen master plate was prepared either by picking single-colonies that were then inoculated into sterile 96-well microtitre plates or by culture dilution method developed by Dr Bradley Stevenson during his time in this laboratory[23]. The former was employed when the library to be screened was small (<1000) and the latter was employed when the library to be screened was larger than 1000 colonies. Dr Stevenson introduced the culture dilution method largely due to the laborious nature of picking single colonies manually, especially for larger libraries.

In the usual single-colony method, the colonies were picked by hand, and inoculated into 96-well round bottom microtitre plates containing 100 – 200 µL of LB broth supplemented with desired antibiotics. With the culture dilution method, the cell density of the transformed *E. coli* was first estimated as described in **section 2.5.5** while the remainder was either stored in either 4°C or -80°C. Once the cell density was estimated, the transformed *E. coli* culture was thawed and diluted in LB broth to ~3 – 4

cells per 100 – 200 μL and was dispensed into sterile microtitre plates with 100 – 200 μL culture per well. The plates were incubated in orbital shaking incubator at 37°C for 24 hours.

Depending on the substrate screened, the cultures were treated differently before lysis to maintain the linear range of the assay. The amount of culture used and lysis procedure would be described in the latter chapters of this thesis. Unless otherwise stated, the proteins were assayed in 96-well flat bottom plates with 119 μM methyl parathion (MPS), ethyl parathion (EPS) or ethyl paraxon (EPO) in 50 mM Tris-HCl or HEPES, pH 7.6, 100 mM NaCl, 0.1% Tween 20. The activity was monitored by the release of *p*-nitrophenolate at 405 nm using a Spectramax M2e microplate reader at 30°C. This assay procedure was adapted and modified from various references[30–32]. In the later stage of the work, the proteins were assayed with 119 μM OP substrates in 50 mM HEPES, pH 7.6, 100 mM NaCl and the rationale would be explained in **Chapter 4.** 0.1 mM $ZnCl_2$ was included in the assay buffer for all activity measurements done in **Chapter 5.** The activities were compared to those of wild type, and in some cases, the best mutant from previous round, and mutants with the best enhancements or with the desired substrate specificity were selected.

The selected mutants were streaked on LB plates supplemented with antibiotics to obtain single colonies. One colony was picked from each agar plate for primary screen master plates prepared with single colony picking method, whereas up to six colonies were picked from each agar plate for primary screen master plates prepared with culture dilution method. Mutants selected from the secondary screen were streaked and single colonies were grown in 3 mL LB broth supplemented with antibiotics for plasmid isolation and DNA sequencing. Unique mutants were then selected and subjected to the next round of evolution.

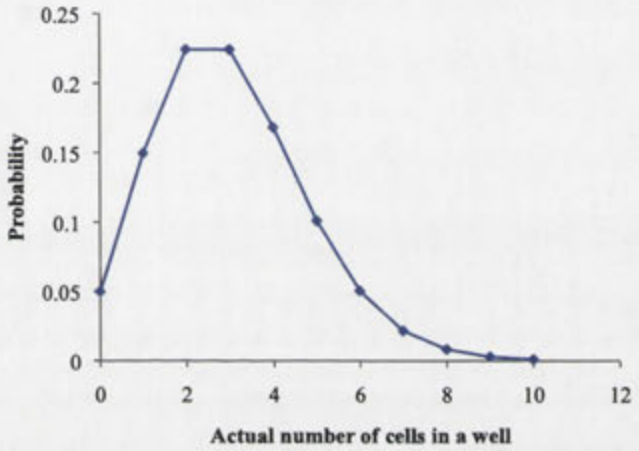
With the culture dilution method, the method assumes that a mutant clone with enhanced activity can be detected even in the presence of other non-performing mutants. The method also assumes that the cells are evenly distributed in the culture dilution and the cell division is at minimal during the process of dilution and dispensation of the culture.

The Poisson distribution given by **Equation 2.1** describes the probability of finding viable cells in any well:

$$P(r) = \frac{e^{-\mu}\mu^{r}}{r!} \tag{2.1}$$

Where $P(r)$ is the probability of a well having r number of cells, when the mean number of cells per well is μ. Based on the number of sterile wells in the microtitre plate (i.e. $r = 0$), the actual mean number of cells per well can be calculated.

(a)



(b)

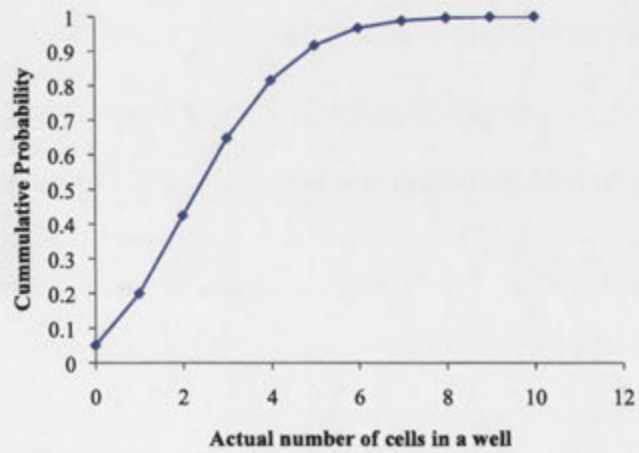


Figure 2.2: Poisson distribution of cells in liquid culture.
(a) shows the distribution of cells in wells of microtitre plate culture with a mean of dilution of three cells per well and (b) shows the cumulative distribution of cell distribution of the same dilution.

As illustrated in **Figure 2.2b**, for a mean of three cells per well, ~90% of wells have five or fewer cells and ~9% of wells have six or more cells after dilution. As a compromise to isolate all variants present in a well, up to six colonies were picked from the streaking of each chosen well for secondary screen.

2.6. Protein manipulations

This section will focus on the general methods used for protein studies such as protein concentration quantification and protein visualisation. Description of procedures that are experiment-specific such as protein expression for certain experiments and protein purification would be described in **Chapters 3**.

2.6.1. Recombinant protein overexpression

Proteins were expressed by 'leaky expression', IPTG induction or constitutive expression, depending on the plasmid vectors as described in **section 2.2**. BL21 (DE3) *RecA*$^-$ was used for expression gene in pET based vectors and DH5α was used for *lac* promoter based vectors. For large-scale expression, a single colony was inoculated into 10 mL of LB broth supplemented with antibiotics and grown overnight at 37°C as starter culture. The next day, the 10 mL starter culture was subcultured to a fresh 1 L LB broth supplemented with antibiotics (starter culture volume: final culture volume ratio of 1:100). Three different growth conditions were used after the subculture to 1 L fresh media:

- pETMCSI is an IPTG inducible vector. The cells were grown at 37°C for 4-5 hours to an OD_{600} of 0.6 – 0.8. IPTG was added to a final concentration of 0.1 – 0.5 mM and the cells were then incubated by orbital shaking at 18-25°C for either overnight or 16 hours.
- pJWL1030 and pCY76 express constitutively. The 1 L culture was further grown overnight at 37°C.
- pET47b is an IPTG inducible vector used in **Chapter 3** to investigate MPH proteolysis. The cells were grown at 37°C for 4 – 5 hours to an OD_{600} of 0.6 – 0.8. IPTG and $ZnCl_2$ were added to final concentrations of 0.5 mM and 0.1 mM respectively, and the cells were then incubated by orbital shaking at 37°C for either overnight or 16 hours.

For small-scale expression that was used for expression confirmation, crude lysate assay and mutant activity validation, the final culture volume was scaled down to typically 5 – 20 mL. The starter culture was also scaled down accordingly, based on the starter culture volume: final culture volume ratio of 1:100, and was grown similarly as described above.

2.6.2. Cell lysis

For large-scale culture to be used in protein purification, the cultured cells were harvested by centrifugation at 5,000 rpm in a VWR R9A rotor for 15 minutes and were either used immediately or stored at -80°C until required. The centrifugation and all subsequent steps were done at 4°C. The cells were resuspended in 10 mL of column binding buffer per gram of wet pellet until it reached homogeneity. The column binding buffer was supplemented by either 1 mM PMSF or AEBSF to minimise proteolysis. The column binding buffer used for each column are specified in **Chapter 3.** The resuspended cells were lysed by passage through a French Pressure Cell Press (12,000 psi cell pressure, two repeats) and centrifuged in a VWR R20A2 rotor at 12,000 rpm for 45 minutes to pellet cell debris. The clarified lysate was then applied to a purification column using a peristaltic pump and was purified. The columns used for each purification protocol are described in **Chapter 3.** Unless stated, all purifications were done using an AKTA™ FPLC System.

For small-scale cultures, cell lysis was done with Bugbuster® protein extraction reagent (Novagen) or rLysozyme™ (Novagen) according to the suppliers' protocols. The cells were harvested with an Eppendorf 5415D bench top centrifuge at maximum speed for five minutes. In the case of using Bugbuster™, the cells were resuspended in 5 mL 50 mM HEPES or Tris-HCl, 100 mM NaCl, 1x BugBuster®, pH 7.6 per gram of wet cell paste. The mixture was kept at room temperature for 15 – 30 minutes prior to centrifugation. For rLysozyme™ lysis, the harvested cells were frozen completely at -20°C or -80°C. The cells were then completely thawed and resuspended in 7 mL of 50 mM HEPES or Tris, 100 mM NaCl, pH 7.6 per gram of well cell paste. rLysozyme™ was then added to the resuspended cells at a ratio of 7.5 kU rLysozyme™ per mL of resuspended cells. The cell suspension was then incubated on a shaking platform for 10

– 20 minutes at room temperature. The mixture was then centrifuged for 20 minutes at max speed to remove insoluble cell debris.

2.6.3. Protein visualisation

Proteins were visualised using Laemmli SDS-PAGE.[33] The SDS-PAGE used composed of a 5% (w/v) stacking gel at pH 6.8 and a 15% (w/v) resolving gel at pH 8.8. The two component gels were prepared as below:

	15% (w/v) Resolving gel	5% (w/v) Stacking gel
Acryl/Bis 37.5: 1, 40% (w/v)	3000 µL	375 µL
0.5 M Tris, pH 6.8	-	833 µL
1.5 M Tris, pH 8.8	2000 µL	-
10% (w/v) SDS	80 µL	33.3 µL
dH_2O	2850 µL	2069 µL
TEMED	10 µL	3 µL
10% (w/v) APS	60 µL	20 µL
Total volume	8000 µL	3333.3 µL

Low range SDS-PAGE molecular weight standard (Bio-Rad) and Dual Color Precision Plus Protein™ Prestained Standards (Bio-Rad) were used for all protein visualisation described in this thesis. Low range SDS-PAGE marker contains six different proteins with molecular weights ranging between 14.4 kDa and 97.4 kDa, while Dual Color Precision Plus contains ten different proteins of molecular weights ranging from 10 kDa to 250 kDa. The stock for low range SDS-PAGE marker was diluted 20-fold with β-mercaptoethanol and stock sample buffer (**Appendix A**) mixture as recommended by the supplier's protocol. Diluted protein standards were heated for five minutes at 95°C and stored at -20°C until use. Dual Color Precision Plus marker was used without dilution and heat treatment.

Protein samples were mixed with SDS-reducing sample buffer (**Appendix A**) at a ratio of 1:1 and heat treated at 95°C for five minutes prior to loading onto the SDS-PAGE gel. The 5 mm well width (10-lane combs) and 3.5 mm well width can accommodate a maximum of 20 µL and 15 µL of protein sample respectively. The electrophoresis was ran in SDS running buffer (**Appendix A**) at variable voltage with a constant current of 30 mA per gel. A typical SDS-PAGE run required 35 – 45 minutes to allow the dye front to migrate to the end of the gel.

2.6.4. Protein quantification

Protein concentrations were determined by NanoDrop® by measuring the absorbance at 280 nm. 1.5 – 2 µL of protein sample was used for each measurement. Each protein sample was measured at least thrice and the mean and standard deviation of the measurements were calculated. Only a mean with less than 10% standard deviation would be accepted. The protein concentration was calculated from the A_{280} values obtained based on the Beer-Lambert equation[34] that is given in **Equation 2.2:**

$$A = \varepsilon l c \qquad (2.2)$$

Where A is the absorbance, ε is the molar absorption coefficient ($M^{-1}cm^{-1}$), l is the cell path length (cm) and c is the molar concentration (M). Protein extinction coefficient was estimated by ProtParam tool (http://web.expasy.org/protparam/).

2.6.5. Protein molecular weight determination

Size exclusion chromatography and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) were the two main methods used for molecular weight determination.

To determine protein molecular weight with size exclusion chromatography, the column was calibrated with commercial protein standards with known molecular weights. Superdex 200 size exclusion column was calibrated with four protein standards and blue dextran from Gel Filtration Markers Kit for Protein Molecular Weights 12,000 – 200,000 Da (Sigma) according to the supplier's protocols and the calibration curve obtained is shown in **Figure 2.3.** To determine the molecular weight of MPH, MPH enzyme purified from purification workflows that will be described in **Chapter 3** was used. The purified MPH fractions were collected and concentrated down to 1 mL and loaded onto the Superdex 200 size exclusion column. The elution volume for MPH obtained was then fitted into the linear equation obtained from the calibration curve to calculate the molecular weight. Size exclusion chromatography could also have been used to estimate the oligomeric state of the enzyme since the chromatography was non-denaturing.

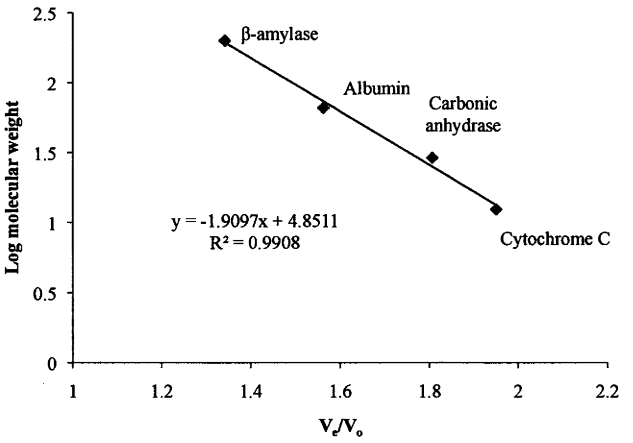


Figure 2.3: Calibration curve obtained with protein standards on Superdex 200.
The molecular weights of β-amylase, albumin, carbonic anhydrase and cytochrome c are 200 kDa, 66 kDa, 29 kDa and 12.4 kDa respectively. Blue dextran (2,000 Da) was used to obtain the void volume. V_e is the elution volume and V_o is the void volume of the column.

To determine the molecular weight with FTICR-MS, purified MPH sample was dialysed exhaustively (two 3-hour dialysis, one overnight dialysis) against water at 4°C to remove any salt from the purification or storage buffer. A final protein concentration of 20 μM in at least 100 μL is required for the measurement. The FTICR-MS measurement and analysis were performed by Professor Thomas Huber.

2.6.6. MPH sample preparation for zinc content measurement.

Zinc composition of MPH was measured by inductively coupled plasma – optical emission spectroscopy (ICP-OES). In **Chapter 4**, the zinc content of MPH, purified as it is, and MPH incubated with excess zinc are measured. The incubation of the MPH samples with zinc prior to excess metal removal will be described in **section 4.3.2.4**. In this section, the preparation i.e. the treatment of purified MPH enzyme with zinc is described. To remove the excess zinc after incubation, a 5 mL HiTrap Desalting column (GE Healthcare) was used. The column was washed with 10 mL chelating buffer consisted of 5 mM each of EDTA, 1,10-phenanthroline, 2,6-pyridine dicarboxylate, 8-hydroxyquinoline-5-sulfonic acid and β-mercaptoethanol in 20 mM HEPES, pH 7 to remove any metal in the column. The chelators were then removed by washing the column with Chelex® 100-treated 50 mM HEPES, 100 mM NaCl, pH 7.6 buffer. The Chelex-treated buffer was prepared in 1 M nitric acid rinsed glassware. 10 g of Chelex 100 was added to 1 L of buffer and was stirred overnight to prepare the metal free buffer. After HiTrap Desalting column was washed with Chelex-treated buffer, 1.5

mL of protein sample was injected into the column with a syringe. 2 mL of buffer was then injected into the column and the flow through was collected. Based on the manufacturer's protocol, the flow through corresponds to 95% yield recovery with a sample dilution factor of 1.3x. The concentration of the flow through was estimated with a Nanodrop® as described in **section 2.6.4** and the protein was diluted to 1 μM with metal free buffer. The protein was then further diluted 10-fold with nitric acid to obtain a final nitric acid concentration of 4% (w/v). The zinc content measurement and analysis were then performed by Ms. Nur Hafizah Azizan.

2.7. Spectrophotometric assay

2.7.1. Standard MPH activity assay

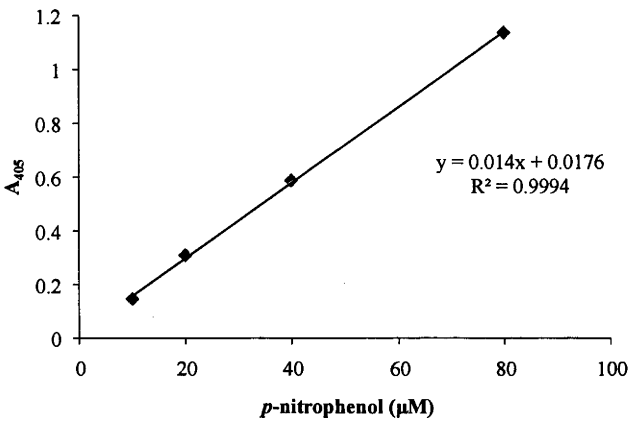
The standard MPH activity assay was similar to the methods for screening libraries described in **section 2.5.6.** A standard assay mixture consisted of 119 μM of OP substrates in 50 mM Tris-HCl or HEPES, 100 mM NaCl, pH 7.6. All assays described in Chapter 5 were done with the supplementation of 0.1 mM $ZnCl_2$. Activity assay was done with either 200 μL reaction mixture per well in 96-well microtitire plates or 1000 μL reaction mixture in micro-cuvettes. The 200 μL reactions were done using Spectramax M2e microplate and the 1000 μL reactions were done using Cary 1E UV-Vis spectrophotometer (Varian). The rate of hydrolysis of MPS, MPO, EPS, EPO, *p*-nitrophenyl acetate (pNPA), *p*-nitrophenyl phosphate (pNPP) and bis-(*p*-nitrophenyl) phosphate (bpNPP) were monitored through the release of *p*-nitrophenol (pNP) at 405 nm while hydrolysis of MCO and MCS were monitored through the release of 3,5,6-trichloro-2-pyridinol (TCPy) at 310 nm. All measurements were done in either duplicates or triplicates.

2.7.2. Enzyme kinetics

To determine the kinetic parameters for the MPH variants that have been obtained in **Chapter 5,** purified MPH enzymes were assayed in 50 mM HEPES, 100 mM NaCl, 0.1 mM $ZnCl_2$, pH 7.6 with various concentrations of different substrates. The substrates assayed were methyl parathion (MPS) (7.6 – 189.9 μM), methyl paraoxon (MPO) (15.2 -379.8 μM), ethyl parathion (EPS) (15.2 – 60.8 μM), ethyl paraoxon (EPO) (15.2 – 379.8 μM), methyl chlorpyrifos oxon (MCO) (30.4 – 379.8 μM) and ethyl chlorpyrifos oxon (ECO) (30.4 – 379.8 μM). The rate of hydrolysis of

MPS, MPO, EPS and EPO were monitored through the release of pNP at 405 nm (ε_{405} = 14696.56 $M^{-1}cm^{-1}$) while hydrolysis of MCO and MCS were monitored through the release of TCPy at 310 nm (ε_{310} = 5800.67 $M^{-1}cm^{-1}$) with Cary 1E UV-Vis spectrophotometer. All kinetic measurements were done in triplicates at 30°C. Standard curves of pNP and TCPy (**Figure 2.4**) were constructed to determine the extinction coefficients of pNP and TCPy by using the Beer-Lambert equation (**Equation 2.2**) since the absorption spectrum of a given compound is affected by pH and temperature. The pNP ε_{405} value was obtained by taking the average of ε_{405} values calculated from 10 – 80 μM of pNP while the TCPy ε_{310} value was obtained by taking the average of ε_{405} values calculated from 10 – 320 μM of TCPy. The measurements were done in triplicates.

(a)



(b)

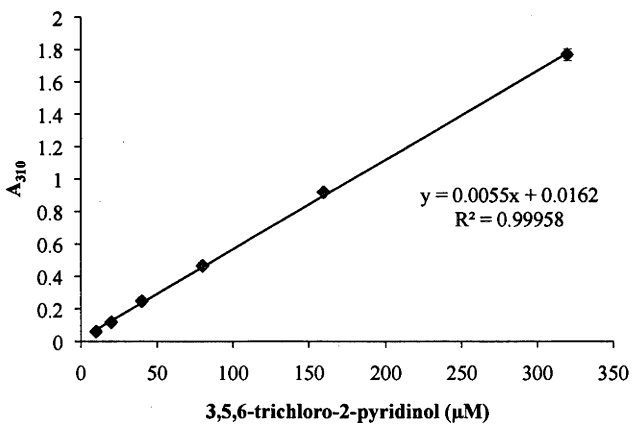


Figure 2.4: The standard curves used to calculate the extinction coefficient values.
(a) is the standard curve built using *p*-nitrophenol (pNP) and (b) is the standard curve built using 3,5,6-trichloro-2-pyridinol (TCPy) and their respective R^2 values.

The kinetics parameters were determined by fitting the data to the modified Michaelis-Menten equation, using curve-fitting software KaleidaGraph. **Equation 2.3** gives the Michaelis-Menten equation.

$$v_o = \frac{V_{max} S}{K_m + S} \tag{2.3}$$

Where v_o is the iniial velocity, V_{max} is the maximum velocity; K_m is the Michaelis-Menten constant and S is the substrate concentration. And since V_{max} is the product of k_{cat} and enzyme concentration, E (**Equation 2.4**), the Michaelis-Menten equation can be rewritten as **Equation 2.5**.

$$V_{max} = k_{cat} E \tag{2.4}$$

$$v_o = \frac{k_{cat} E S}{K_m + S} \tag{2.5}$$

2.8. Computer software programs

DNA sequences were analysed using Sequencher 4.10.1. Multiple DNA and protein sequence alignments were done with Clustal-Omega[35] (http://www.ebi.ac.uk/Tools/msa/clustalo/). ConSurf Server was used for estimating the evolutionary conservation of amino acid positions in a protein based on the phylogenetic relationship between homologous sequences (http://consurf.tau.ac.il/).[36] Protein structures were obtained from Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). MacPyMOL was used to visualise the protein structures and generated the protein structural figures used in this thesis. SWISS-MODEL[37–39] (http://swissmodel.expasy.org/) was used to generate structures of MPH mutants that will be found in **Chapters 4** and **5** by using WT MPH as template. AutoDock Vina[40,41] and Swiss-Dock[41,42] were used to perform molecular docking of MPS to MPH, which is described in **Chapter 5.** The structure of MPS used for docking experiments was obtained from ZINC database (http://zinc.docking.org/).

2.9. References

1. Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual.* (Cold Spring Harbor Laboratory Press, 2001).
2. Novagen. *pET System Manual.* (Novagen, 2003). at <www.novagen.com>
3. Ratelade, J. *et al.* Production of recombinant proteins in the *lon*-deficient BL21(DE3) strain of *Escherichia coli* in the absence of the DnaK chaperone. *Applied and Environmental Microbiology* **75,** 3803–3807 (2009).
4. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189,** 113–30 (1986).
5. Neylon, C. *et al.* Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39,** 11989–11999 (2000).
6. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41,** 207–234 (2005).
7. Yang, H. *et al.* Evolution of an organophosphate-degrading enzyme: a comparison of natural and directed evolution. *Protein Engineering* **16,** 135–145 (2003).
8. Love, C. A., Lilley, P. E. & Dixon, N. E. Stable high-copy-number bacteriophage λ promoter vectors for overproduction of proteins in *Escherichia coli. Gene* **176,** 49–53 (1996).
9. Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. The pMTL *nic*- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68,** 139–149 (1988).
10. Liu, J., Boucher, Y., Stokes, H. W. & Ollis, D. L. Improving protein solubility: the use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expression and Purification* **47,** 258–263 (2006).
11. Kirschman, J. a & Cramer, J. H. Two new tools: multi-purpose cloning vectors that carry kanamycin or spectinomycin/streptomycin resistance markers. *Gene* **68,** 163–165 (1988).
12. Novagen. pET47b(+) Vector. at <http://www.emdmillipore.com/life-science-research/pet-47b-plus-dna/EMD_BIO-71461-3/p_uuid>
13. Brody, J. R. & Kern, S. E. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques* **36,** 214–216 (2004).
14. Dower, W. J., Miller, J. M. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research* **16,** 6127–6145 (1988).
15. Prasanna, G. L. & Panda, T. Electroporation: basic principles, practical considerations and applications in molecular biology. *Bioprocess Engineering* **16,** 261–264 (1997).
16. Kibbe, W. a. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35,** W43–W46 (2007).
17. Woodman, M. E. in *Current Protocols in Microbiology* A.3D.1–A.3D.6 (John Wiley & Sons, Inc, 2008). doi:10.1002/9780471729259.mca03ds9
18. ACRF Biomolecular Resource Facility. Sanger sequencing - user prep samples. at <http://brf.anu.edu.au/sanger-sequencing-user-prep-samples>
19. Cadwell, R. C. & Joyce, G. F. Mutagenic PCR. *PCR Methods and Applications* **3,** S136–S140 (1994).

20. Cirino, P. C., Mayer, K. M. & Umeno, D. in *Methods in Molecular Biology Volume 231* (Arnold, F. H. & Georgiou, G.) **231,** 3–9 (Human Press).
21. Zhao, H. & Zha, W. *In vitro* "sexual" evolution through the PCR-based staggered extension process (StEP). *Nature Protocols* **1,** 1865–1871 (2006).
22. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology* **16,** 258–261 (1998).
23. Stevenson, B. Directed evolution of pyruvate decarboxylase for *in vitro* glycolysis. (2006).
24. Miyazaki, K. in *Methods in Molecular Biology Volume 231* (Arnold, F. H. & Georgiou, G.) 23–28 (Human Press).
25. Georgescu, R., Geethani, B. & Sun, L. in *Methods in Molecular Biology Volume 231* (Arnold, F. H. & Georgiou, G.) 75–83 (Human Press).
26. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59 (1989).
27. Sanchis, J. *et al.* Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. *Applied Microbiology and Biotechnology* **81,** 387–397 (2008).
28. Tseng, W.-C., Lin, J.-W., Wei, T.-Y. & Fang, T.-Y. A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method. *Analytical Biochemistry* **375,** 376–378 (2008).
29. Miyazaki, K. & Takenouchi, M. Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *BioTechniques* **33,** 1033–1038 (2002).
30. Fu, G., Cui, Z., Huang, T. & Li, S. Expression, purification, and characterization of a novel methyl parathion hydrolase. *Protein Expression and Purification* **36,** 170–176 (2004).
31. Huang, L. *et al.* Improving the acidic stability of a methyl parathion hydrolase by changing basic residues to acidic residues. *Biotechnology Letters* **34,** 1115–1121 (2012).
32. Yu, H. *et al.* Expression of methyl parathion hydrolase in *Pichia pastoris*. *Current Microbiology* **59,** 573–578 (2009).
33. Laemmli, U. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).
34. Grimsley, G. R. & Pace, C. N. in *Current Protocols in Protein Science* 3.1.1–3.1.9 (2003).
35. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7,** 1–6 (2011).
36. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* **38,** W529–W533 (2010).
37. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* **31,** 3381–3385 (2003).
38. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).
39. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

40. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Compuational Chemistry* **31,** 455–461 (2010).
41. Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *Journal of Compuational Chemistry* **32,** 2149–2159 (2011).
42. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* **39,** W270–W277 (2011).

PROTEIN PURIFICATION: DEALING WITH PROTEOLYTIC DEGRADATION

3. Protein Purification: Dealing with Proteolytic Degradation

3.1. Preamble

In the initial attempts to purify MPH, it became apparent that proteolysis would be a problem. Hence, much of the work in developing a purification protocol was concerned with avoiding or minimising degradation due to proteolysis.

Proteases are enzymes that cleave proteins and play a central role in cellular homeostasis. However, their presence in protein expression hosts can cause unwanted protein degradation during protein expression and purification. This degradation will result in reduced yield, poorer protein quality and introduce complications to downstream analyses of proteins such as kinetic characterisation and protein crystallography, due to heterogeneity in the protein sample. In *Escherichia coli,* one of the most common protein expression hosts, many proteases were identified in both cytoplasmic and periplasmic compartments. A list of proteases identified in *E. coli* can be found in the *E. coli* Protease Database (http://www.cf.ac.uk/biosi/staffinfo/ehrmann/tools/proteases.index.html), and numerous reviews found in the literature.[1–3] It is estimated that more than 3% of enzymatic activity in *E. coli* at any given time is proteolytic.[1] The number of proteases that are present in various compartments of *E. coli* underlines the challenge a protein chemist faces when expressing proteins in *E. coli.* While there are strategies available to minimise proteolysis, the diverse substrate specificity and sheer number of proteases present in *E. coli* often complicate the troubleshooting process when protein degradation is observed.

In this chapter, a detailed discussion of MPH proteolytic degradation and attempts to overcome it is given.

3.2. Protein expression and purification

Otherwise stated, WBC3-MPH was expressed in *E. coli* BL21(DE3)$^{RecA-}$ for the work described in this section. 1 L cell cultures grown in LB supplemented with

suitable antibiotics were used for the each purification. The expression conditions were described in **Chapter 2 (section 2.6.1),** and the exact conditions used will be noted at each of the purification workflow described in this section.

The first 34 amino acid residues of the intact MPH sequence that constitute the signal peptide were removed and the resulting sequence coding the mature MPH was designated M1T (**M**PH-**1**-**T**runcation) in the plasmid construct. The MPH signal peptide was predicted to be a Sec-type signal peptide by PRED-TAT[4] web server and was consistent with the signal peptide found in MPH from *Pleisiomonas* sp. M6[5]. PSORTb 3.0[6] web server predicted that MPH is localised to the periplasm when expressed with the signal peptide. However, the signal peptide was removed to increase the yield of protein, as the posttranslational cleavage of the signal peptide was the bottleneck of obtaining soluble enzyme. That is, the yield of protein export to the periplasm was low, and the uncleaved protein in the cytoplasm formed inclusion bodies, as observed by Fu *et al*[5] and in this work as shown in **Figure 3.1.** A reasonable yield of soluble protein was obtained by expressing the mature protein in the cytoplasm. However, along with the soluble protein, there was still a large quantity of insoluble protein in the cytoplasm, especially when the protein was expressed with an inducible promoter (pET vectors) at 37°C. Decreasing the expression temperature to 18°C or using a constitutive vector (pCY76 or pJWL1030) increased the fraction of soluble protein obtained. For the sake of simplicity, unless stated otherwise, MPH refers to the enzyme expressed in the cytoplasm with the sequence coding for signal peptide removed.

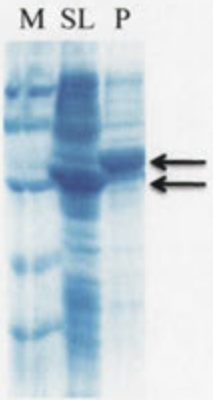


Figure 3.1: The soluble and insoluble fractions of MPH expressed without prior removal of signal sequence.
The difference in the MPH mass shown by soluble and insoluble fractions, as indicated by the arrows, suggests that the MPH present in the insoluble fraction has the intact signal peptide. Molecular weights of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, and 21.5 kDa. SL denotes the soluble lysate after clarification from the total lysate and P denotes the insoluble fraction.

pETMCSI_M1T_WT was constructed by amplifying the sequence coding for mature MPH (M1T) with F-SSM-Flank_Ase and R-SSM-Flank-M1T-Eco primers (**Appendix C**) containing *Ase*I and *Eco*RI respectively. The PCR product was digested and ligated into *Nde*I and *Eco*RI sites of pETMCSI vector. *Nde*I was not introduced when amplifiying M1T due to the presence of *Nde*I site in the M1T sequence. *Nde*I and *Ase*I produce the same overhang after digestion and therefore can be used for ligation. The digestion and ligation steps were performed as described in **sections 2.3.6 – 2.3.8.**

3.2.1. Purification workflow I – cationic exchange column and gel filtration

The purification workflow described here was adopted and modified from Sun and colleagues' work in crystallising MPH.[7] The purification procedure involved a cationic exchange column and a gel filtration column. In Sun and co-workers' protocol, MPH was expressed in the native host, *Pseudomonas* sp. WBC-3, and purified with CM-Sepharose Fast Flow column. For purification workflow I, MPH was expressed in *E. coli* BL21(DE3)$^{RecA-}$ harbouring pETMCSI_M1T_WT with 0.1 mM IPTG induction at 18°C for 16 hours. The cells were harvested and resuspended in 20 mM KH_2PO_4 or HEPES, pH 7 and were lysed with French Pressure Cell Press prior to purification **(section 2.6.2)**. The clarified lysate was applied to a 20 mL SP Sepharose or a 5 mL HiTrap SP Fast Flow column and the bound MPH was eluted using a linear gradient of 0 to 1 M NaCl in 20 mM KH_2PO_4 or HEPES, pH 7 over 10 column volumes. The eluted fractions with the highest MPH activity were pooled, concentrated and buffer exchanged to 50 mM Tris-HCl or HEPES, 100 mM NaCl, pH 7. The sample was then applied to a 120 mL Superdex 200 gel filtration column and eluted with 50 mM Tris-HCl or HEPES, 100 mM NaCl, pH 7.6. The workflow for this purification is summarised in **Figure 3.2.** The purity of the purification fractions was assessed by SDS-PAGE (**Figure 3.3**). The molecular weight and isoelectric point of MPH were predicted to be 31538.9 Da and 7.87 respectively by ProtParam (http://web.expasy.org/protparam/).

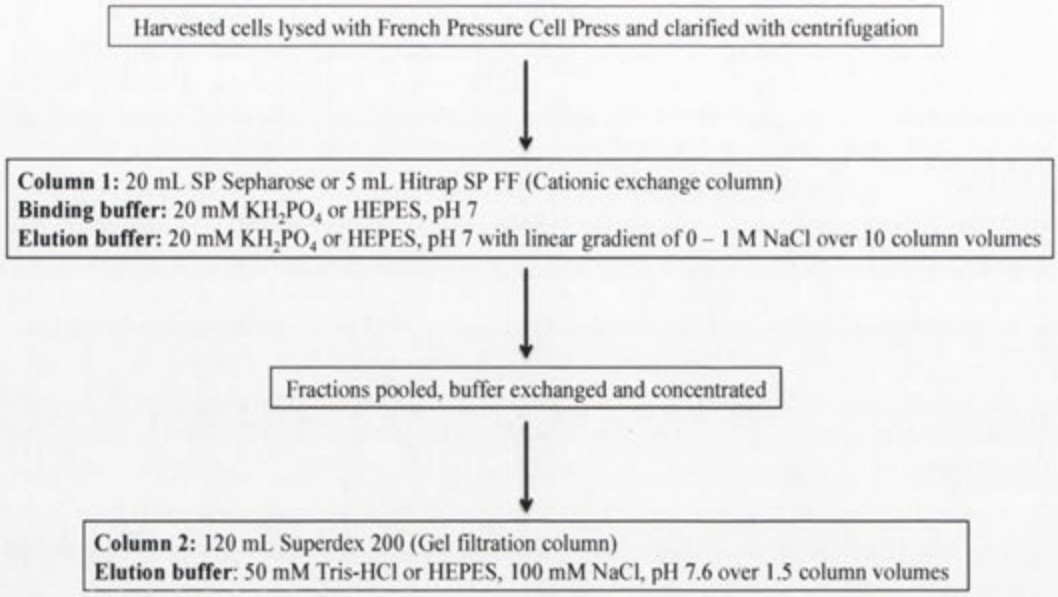


Figure 3.2: Purification workflow I.

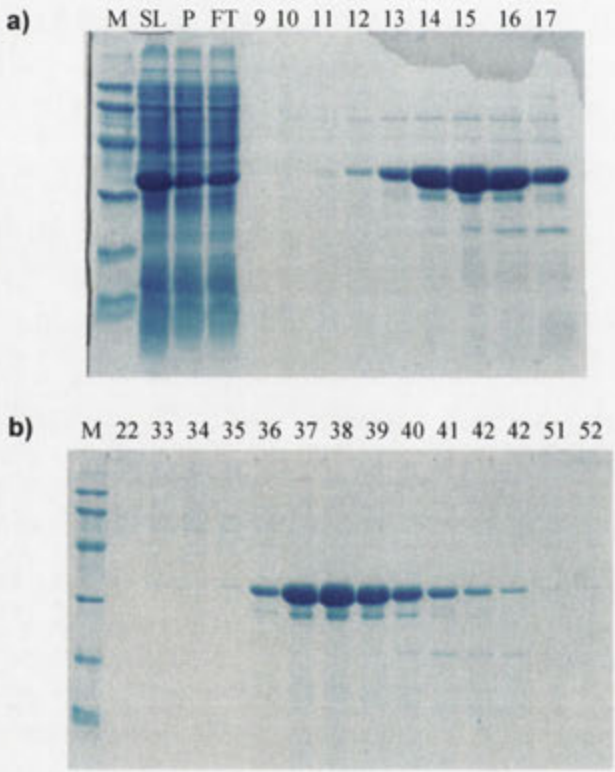


Figure 3.3: SDS-PAGE profiles of the purified fractions done with purification workflow I.
a) shows the eluted fractions from SP Sepharose column and b) shows the eluted fractions from Superdex 200. Molecular weights of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, P denotes the insoluble fraction, FT denotes column flow through and the numbers denote the eluted fractions.

From **Figure 3.3**, a band that corresponded to MPH was observed in the SDS-PAGE profiles for eluted fractions obtained from the cationic exchange column and gel filtration column. However, a slightly smaller band below the MPH band was also observed for both SDS-PAGE profiles, which indicated proteolysis. It was interesting that the smaller band was not separated from MPH when passed through the gel filtration column, suggesting the proteolysis happened throughout the purification. It is also possible that the degraded product was not significantly different enough to be separated by gel filtration chromatography. Supplementing protease inhibitors such as phenlymethylsulfonyl fluoride (PMSF) or 4-(2-Aminoethyl) benzenesulfonyl fluoride (AEBSF) to the lysis buffer did not stop the degradation, suggesting that the protease involved was not a serine protease (data not shown). At this stage, it was uncertain if the proteolysis was due to protease action from the expression host or due to autoproteolysis. One example of enzyme exhibiting promiscuous protease activity was shrimp alkaline phosphatase (SAP).[8] It was suggested that the autoproteolytic activity exhibited by metal-depleted SAP was to regulate the level of phosphatase activity in response to changes in physiological condition. Therefore, the possibility of autoproteolysis cannot be ruled out at this point.

MPH was eluted from Superdex 200 column at 82.425 mL. By fitting the elution volume of MPH to the Superdex 200 calibration curve (**section 2.6.5, Figure 2.3**), MPH was estimated to be 60.3 kDa, closer to the calculated dimeric molecular weight of 63.1 kDa, indicating that MPH was a dimer in solution. The observation was consistent with the dimeric crystal structure solved by Dong *et al.*, but differed from the oligomeric state obtained by Fu *et al.*[5] and Yang *et al.*[9] Fu *et al.* predicted MPH from *Pseudomonas* sp. M6 to be a monomer, with gel filtration chromatography while Yang *et al.* predicted MPH from *Pseudomonas* sp. WBC-3 to be a monomer with analytical ultracentrifugation. The MPH monomer observed by Fu *et al.* might be due to the presence of C-terminal His-Tag affecting the dimeric structure. The N and C-termini of MPH form the dimerisation interface of the dimer. The MPH monomer observed by Yang *et al.* might be due to the dilute MPH sample (~25 nM) used for the ultracentrifugation analysis. Proteins that primarily exist in oligomeric state generally have dissociation constants in the nanomolar range while some proteins have dissociation constants in the micromolar or milimolar range, thus the oligomeric structure might dissociate at concentrations lower than the dissociation constant.[10]

During gel filtration or crystallisation, the protein samples used are often very concentrated.

3.2.2. Attempts to overcome proteolytic degradation

3.2.2.1. Identifying proteolytic cleavage site

In order to identify the cleavage site in MPH, a purified MPH sample that contained both MPH and the smaller band was analysed with Fourier transform ion cyclotron resonance mass spectroscopy (FTICR-MS). Mass spectroscopy can give accurate measurement of protein mass and therefore can be used to measure the mass of the smaller band. The mass obtained was then used to predict the cleavage site of the protease. The protein sample for FTICR-MS was prepared as described in **section 2.6.5,** and the FTICR-MS measurement and analysis were performed by Professor Thomas Huber.

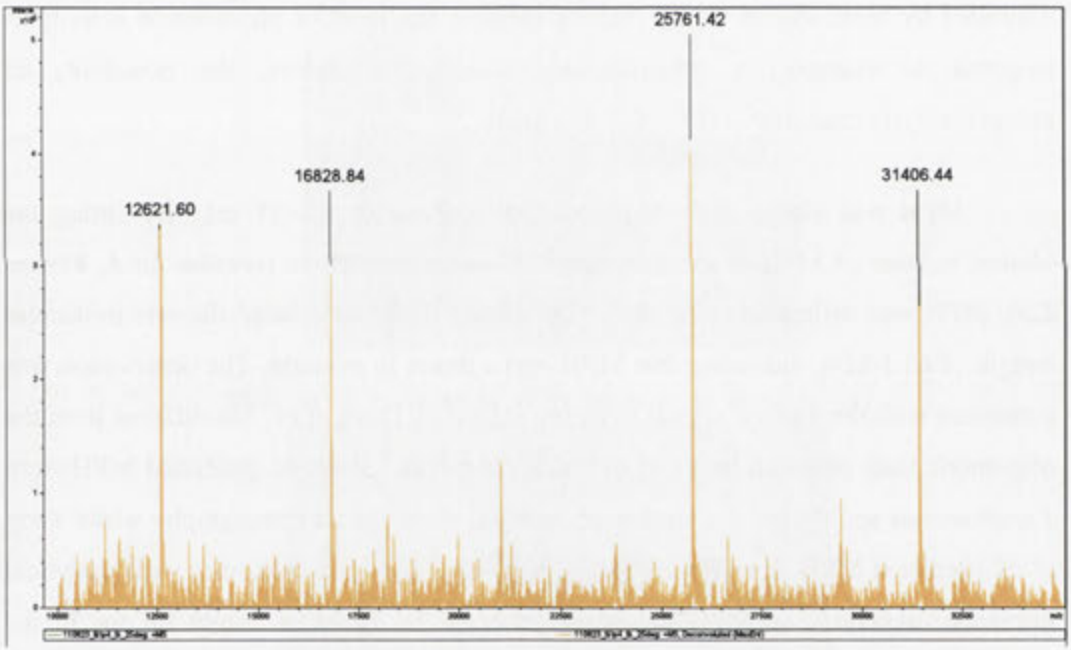


Figure 3.4: Deconvoluted FTICR-MS spectrum of MPH purified with purification workflow I.
The numbers on the peak indicates the molecular weight of the detected species in Dalton.

According to the deconvoluted mass spectrum of the MPH sample (**Figure 3.4**), four species with masses 31406.44 Da, 25761.42 Da, 16828.84 Da and 12621.60 Da were found. The 31406.44 Da species corresponded to the mass of MPH without the N-terminal methionine (149.21 Da). The loss of N-terminal methionine was most likely due to the activity of methionine aminopeptidase (MetAP1) that is responsible for catalysing the N-terminal methionine cleavage of newly synthesised polypeptides.[11]

The specificity of MetAP1 depends on the amino acid residue at P1' and P2' positions (N-terminal methionine being P1 position), and it was found that *E. coli* MetAP1 prefers small amino acid residues with small side chain at P1' position and hydrophilic or small residues at P2' position. Both the amino acid residues at P1' and P2' positions of MPH are Ala, suggesting that MetAP1 was most likely reason for the loss of N-terminal methionine.[11]

The species with masses 16828.84 Da and 12621.60 Da were not observed from the SDS-PAGE profiles (**Figure 3.3**) and the masses were deemed too small to correspond to the smaller band below the MPH band. Moreover, sequential removal of amino acid residues from either terminus to calculate the molecular weight of MPH fragment with ProtParam did not produce molecular weights corresponding to those two values. Therefore, the most likely species that corresponded to the smaller band observed in the SDS-PAGE was the species with a mass of 25761.42 Da, and this was used for the prediction of cleavage site. It is worth noting that the apparent molecular weight of the MPH band was heavier than expected in the SDS-PAGE profile. The protein molecular weight estimation from SDS-PAGE was done by constructing a standard curve based on the migration of the molecular weight markers. The molecular weight for MPH and the smaller band obtained from this estimation were approximately 36 kDa and 28 kDa respectively. Anomalous SDS-PAGE migration is not uncommon and can be attributed to the variation in SDS binding of different proteins,[12,13] hence the accuracy of molecular weight estimation from SDS-PAGE is an approximation at best.

The proteolytic cleavage site was then predicted by removing amino acid residues sequentially from the N or C-terminus and matching the resulting molecular weight to the fragment obtained from FTICR-MS. Two possible cleavage sites were identified, one from each terminus and were listed below. The blue bolded font indicates the cleaved fragment.

MAAPQVRTSAPGYYRMLLGDFEITALSDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNT
GSKLVLVDTGAAGLFGPTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQK
EADFWLSQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTTYVVE
SQGQKLALLGDLILVAAVQFDDPSVTTQLDSDSKSVAVERKKAFADAAKGGYLIAASHLSFPGIGHIRAE
GKGYRFVPVNYSVVNPK
Calculated Molecular Weight: 25762.2 Da

MAAPQVRTSAPGYYRMLLGDFEITALSDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNT
GSKLVLVDTGAAGLFGPTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQK
EADFWLSQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTTYVVE
SQGQKLALLGDLILVAAVQFDDPSVTTQLDSDS**KSVAVERKKAFADAAKGGYLIAASHLSFPGIGHIRAE
GKGYRFVPVNYSVVNPK**
Calculated Molecular Weight: 25783.2 Da

The N-terminus cleavage site was thought to be the likely cleavage site as the calculated molecular weight was in good agreement with the molecular weight obtained from FTICR-MS (25762.2 Da and 25761.4 Da).

3.2.2.2. Site directed mutagenesis

One approach in overcoming the proteolysis is to engineer MPH to be more proteolytic resistant. Several successful examples of proteolytic resistance enhancement were achieved by altering the amino acid residues at the proteolytic cleavage site.[14,15] In MPH, the cleavage site was predicted to be between F88 and Q89. Therefore, F88A site directed mutagenesis (SDM) was performed using sequence overlap extension (SOE) method as described in **section 2.5.4.2** and the resulting mutant was designated MPH_F88A. The forward and reverse flanking primers used were F-SSM-Flank-M1T-Ase and R-SSM-Flank-M1T-Eco while the forward and reverse mutagenic primers used were F-SDM-M1T-F54A and R-SDM-M1T-F54A (**Appendix C**). The resulting construct, pETMCSI_M1T_F88A, was transformed into BL21(DE3)$^{RecA-}$, and protein expression was done similarly as WT MPH (**section 3.2.1**). MPH_F88 was then purified with column I (20 mL SP Sepharose) with the condition described in **section 3.2.1.** Superdex 200 was not used in this purification as it was deemed just a polishing step for the purification. The purity of the purification fractions was assessed by SDS-PAGE (**Figure 3.5**). From **Figure 3.5,** the proteolysis persisted even after the predicted cleavage site was mutated (Fraction 29).

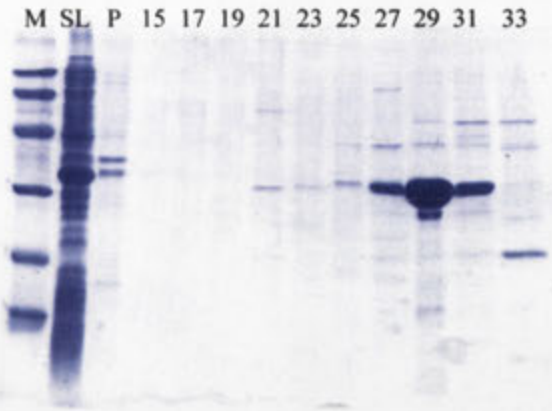


Figure 3.5: SDS-PAGE profile of the purified MPH_F88A fractions done with 20 mL SP Sepharose column.
Molecular weights of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, P denotes the insoluble fraction and the numbers denote the eluted fractions.

3.2.2.3. Purification workflow II – cationic exchange column and anionic exchange column

Another approach in getting around proteolysis is to optimise the purification procedure so that the protease and the degraded fragment can be removed, giving a homogenous MPH enzyme. This is of course assuming that the cause of degradation is due to host protease rather than autoproteolysis of MPH. This purification protocol was adopted from Ms. Nur Hafizah Azizan, who is also working with MPH. The purification procedure involved a cationic exchange column and an anionic exchange column. For this purification workflow, MPH was expressed in *E. coli* BL21(DE3)$^{RecA-}$ harbouring pETMCSI_M1T_WT and harvested as described in purification workflow I (**section 3.2.1**). The purification of MPH with the first column was performed as described in purification workflow I (**section 3.2.1**). The eluted fractions with the highest MPH activity were pooled, concentrated and buffer exchanged to 20 mM piperazine, pH 10. The sample was then applied to a 90 mL Q Sepharose column and the bound protein was eluted using a linear gradient of 0 to 1 M NaCl in 20 mM piperazine, pH 10 over five column volumes. The workflow for this purification was summarised in **Figure 3.6.** The purity of the purification fractions was assessed by SDS-PAGE (**Figure 3.7**).

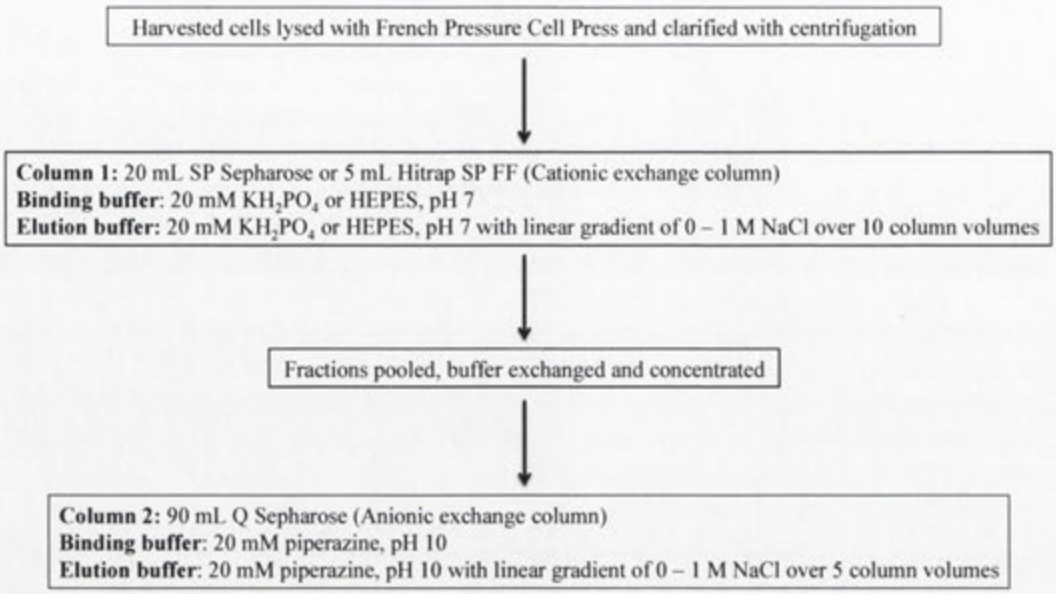


Figure 3.6: Purification workflow II.

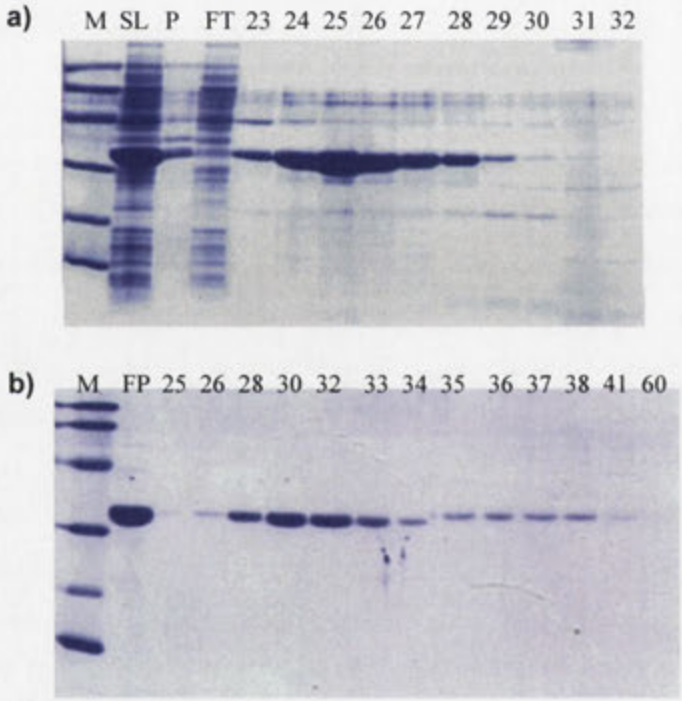


Figure 3.7: SDS-PAGE profiles of the purified fractions done with purification workflow II.
a) shows the eluted fractions from SP Sepharose column and b) shows the eluted fractions from Q Sepharose. Molecular weight of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, P denotes the insoluble fraction, FT denotes column flow through, FP denotes the sample loaded into Q Sepharose column and the numbers denote the eluted fractions.

From the SDS-PAGE profiles (**Figure 3.7**), the purity of the purified enzyme seemed acceptable and no degradation was observed from the purified fractions from Q Sepharose column. However, when examining the final purity of the purified MPH after concentration and buffer exchanged into 50 mM HEPES, 100 mM NaCl, 50% (v/v) glycerol, pH 7.6, a small amount of degraded product was observed below the MPH band (**Figure 3.8**).

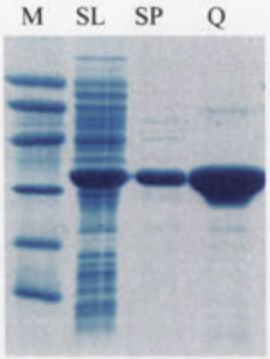


Figure 3.8: MPH purity progression with purification workflow II.
Molecular weights of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, SP denotes the pooled purified fractions from SP Sepharose column and Q denotes the purified enzyme from Q Sepharose

While purification workflow II provided better purity and less degraded product, there were several drawbacks with the purification method. One major drawback was that the enzyme was exposed to denaturing pH range for a considerable duration. Due to deviation from calculated isoelectric point, lower pH values cannot be used when purifying MPH with Q Sepharose. It was found that MPH would not bind to Q Sepharose column with a pH 9 binding buffer, indicating an upward deviation from the theoretical isolectrical point of 7.87 (data not shown). The workflow also gave a lower yield in purified enzyme. It was pointed out that the pI calculations in ProtParam were defined by pKa values obtained by examining polypeptide migration between pH 4.5 and 7.3 under denaturing conditions and that the software might not be adequate for pI prediction of highly basic proteins.[16–18]

3.2.2.4. Purification workflow III – anionic exchange column and cationic exchange column

Purification workflow III also involved an anionic exchange column and a cationic exchange column. However, the strategy used in purification workflow III was different from that of purification workflow II. For this purification workflow, MPH was expressed in *E. coli* BL21(DE3)$^{RecA-}$ harbouring pETMCSI_M1T_WT and

harvested as described in purification workflow I (**Section 3.2.1**). Clarified lysate was loaded onto a 5 mL HiTrap DEAE Fast Flow column and the flow through was collected without NaCl gradient. The flow through was then loaded onto a 5 mL HiTrap SP Fast Flow column and bound MPH was eluted with 0 – 1 M NaCl linear gradient in 20 mM HEPES, pH 7 over 10 column volumes. The workflow for this purification is summarised in **Figure 3.9.** The purity of the purification fractions was assessed by SDS-PAGE (**Figure 3.10**).

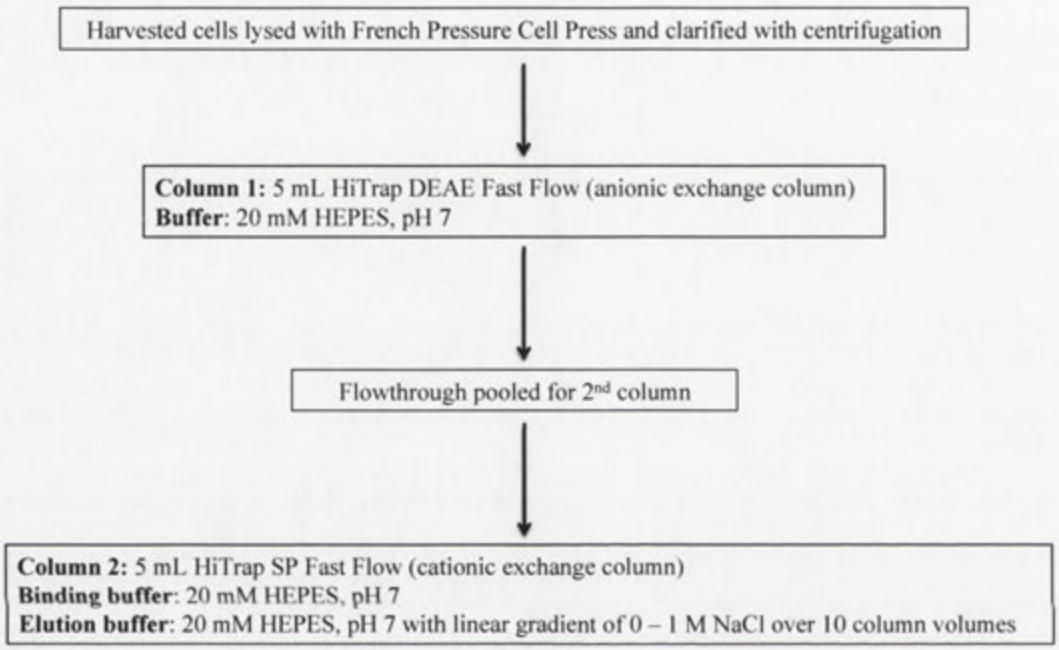


Figure 3.9: Purification workflow III.

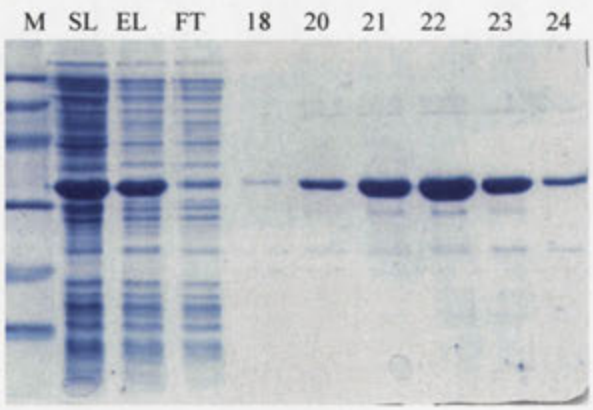


Figure 3.10: SDS-PAGE profile of the purified fractions done with purification workflow III.
Molecular weights of proteins in the molecular marker, M, from top to bottom, are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, EL denotes the eluate from HiTrap DEAE FF column, FT denotes column flow through from HiTrap SP FF column, and the numbers denote the eluted fractions from HiTrap SP FF.

From the SDS-PAGE profile (**Figure 3.10**), a small amount of impurities can be observed. While the purity of this purification workflow was not better than purification workflow II, this workflow allowed rapid purification of MPH with reasonable purity. Compared to purification workflow II, the whole purification workflow III only required 6 – 7 hours, while purification workflow II required overnight purification. The shorter time was possible due to the use of smaller columns and buffers of the same pH on both columns, eliminating the need to exchange buffer in between columns. In the first column (HiTrap DEAE FF), MPH, which has an alkaline pI, was not bound to the column and was collected in the flow through. The flow through, without any buffer exchanging step, was then applied onto the second column (HiTrap SP FF) for purification. This method was used since a big bulk of *E. coli* proteins has acidic pI, as shown in a theoretical 2D map of *E. coli* proteins obtained from the *E. coli* Protein Database (http://eecoli.kaist.ac.kr/theoretical.html) (**Figure 3.11**). Therefore, when using a buffer of pH 7, a large amount of *E. coli* host protein was eliminated prior to the purification with HiTrap SP FF. The simple purification procedure for this workflow also cuts down the time needed between cell lysis and storage of purified enzyme and this would minimise proteolysis. As shown in later sections, purified MPH is stable when stored in -20°C.

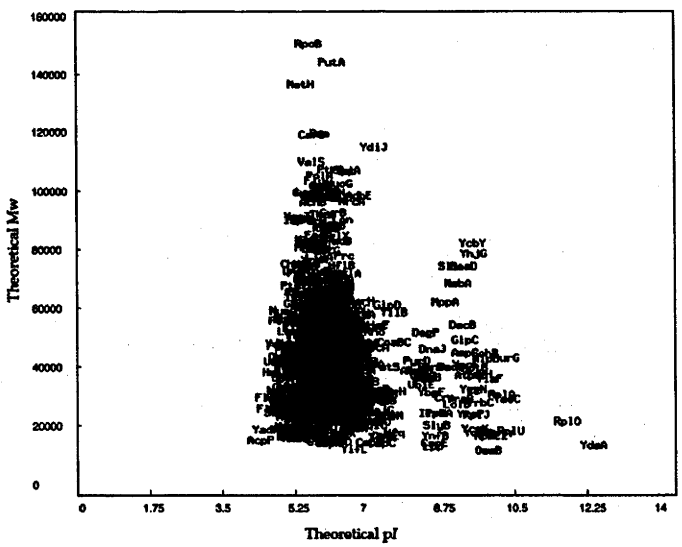


Figure 3.11: 2D map of *E. coli* protein obtained from *E. coli* Protein Database.
The x-axis shows the theoretical pI and the y-axis show the theoretical molecular weight of proteins found in *E. coli*. The map was obtained from the *E. coli* Protein Database (http://eecoli.kaist.ac.kr/theoretical.html)

Purification workflow III was used to purify the MPH variants isolated in **Chapter 4**. The purity and homogeneity of the purified enzyme produced with this workflow was deemed sufficient and the proteolysis should not affect the results if the purification was performed quickly and the purified MPH was stored in -20°C as soon as possible. A simplified version of purification workflow III was also used for the rapid purification of MPH variants in **Chapter 5**. Due to the smaller culture volume used for the purification of MPH variants in **Chapter 5**, column 1 of purification workflow III was omitted and the purification was done with just column 2.

3.2.2.5. Purification workflow IV – nickel affinity column

Immobilised metal-affinity chromatography (IMAC) is widely used to purify recombinant proteins containing polyhistidine affinity tag. This method provides a simple one-step purification process with up to 95% purity[19]. However, the addition of a His-tag to N or C-terminus of MPH might affect the dimeric assembly of MPH since the N and C-termini of each monomer form the dimerisation interface as discussed in **Chapter 1**. In some cases, polyhistidine affinity tag was also found to affect enzymatic activity[20] and crystal structure[21]. It was with these considerations in mind that traditional purification procedures were pursued instead of the more convenient IMAC. Nevertheless, an MPH construct containing polyhistidine tag can be used to probe the cause of the proteolysis. If the proteolysis was due to a protease, IMAC might be able to separate it from His-tagged MPH and the proteolysis would cease. On the other hand, if the proteolysis was due to autoproteolytic activity of MPH, the proteolysis would persist. This purification workflow was developed for the sole purpose of identifying the cause of proteolysis.

For this purification workflow, MPH was expressed in *E. coli* BL21(DE3)$^{RecA-}$ harbouring pET47b_M1T_WT with 0.5 mM IPTG induction at 37°C for 16 hours. The growth media was supplemented with 0.1 mM $ZnCl_2$ to ensure that sufficient Zn^{2+} ions were available to the cell due to the findings in **Chapter 4**. The MPH construct was made by cloning the MPH sequence coding for the mature protein downstream of the N-terminus His-Tag. The cells harvested were resuspended in 50 mM HEPES, 20 mM imidazole, 500 mM NaCl, 1 mM AEBSF, pH 7.6 prior to lysis with French pressure cell press. Two different operation methods were used in this purification workflow: manual elution with syringe (purification workflow IVa) and automated elution with

FPLC purification system (purification workflow IVb). In purification workflow IVa, the clarified lysate were loaded onto a 5 mL Histrap FF column equilibrated with 50m mM HEPES, 20 mM imidazole, 500 mM NaCl, pH 7.6 and bound MPH was eluted with a one-step elution with 500 mM imidazole in 50 mM HEPES, 500 mM NaCl, pH 7.6. Whereas for purification workflow IVb, the clarified lysate was loaded onto the column similarly, and bound MPH was eluted using a linear gradient of 0 to 500 mM imidazole in 50 mM HEPES, 500 mM NaCl, pH 7.6. The workflow for this purification was summarised in **Figure 3.12** and the purity of the eluted enzyme was assessed by SDS-PAGE (**Figure 3.13**).

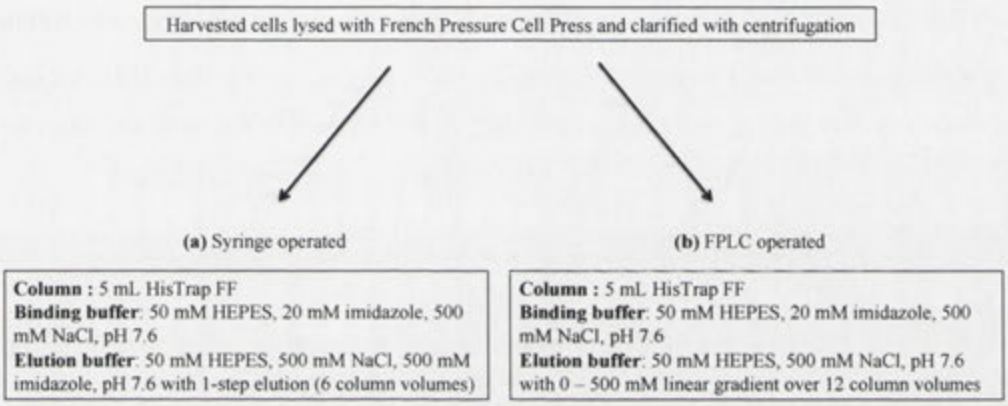


Figure 3.12: Purification workflow IV – (a) syringe and (b) FPLC operated.

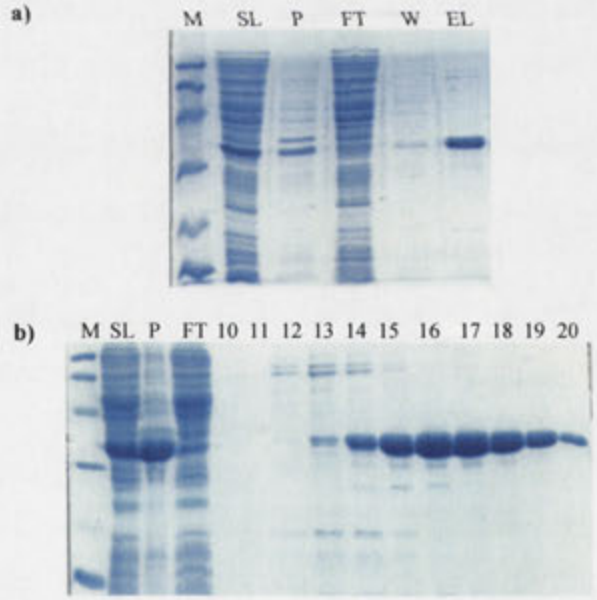


Figure 3.13: SDS-PAGE profiles of purified fractions done with purification workflow IV.
a) shows the profile for syringe operated elution and b) shows the profile for FPLC operated elution. Molecular weights of proteins in molecular marker, M, from top to bottom are 97.4, 66.2, 45.0, 31.0, 21.5, 14.4 kDa. SL denotes the soluble lysate after clarification from the total lysate, P denotes the insoluble fraction, FT denotes the column flow through, W denotes the sample collected from washing after sample application, EL denotes the eluate from the syringe-operated elution and numbers denote the eluted fractions from FPLC.

From the SDS-PAGE profiles (**Figure 3.13**), the purified enzyme obtained from syringe elution gave better purity than FPLC operated elution. However, this was most likely due to dilution of the purified enzyme during one-step MPH elution with syringe. The bound protein was eluted with six column volumes (~30 mL) of elution buffer. **Figure 3.14** shows the purified MPH from purification workflow IVa after concentration. The purity of FPLC eluted MPH would be better since some of the contaminants were eluted at different fractions than MPH. It should be noted that thorough optimisation was not performed for this purification workflow due to time constraints and the purity would be better after the purification is optimised. The smaller bands below the MPH band can still be seen from samples of both purifications workflows, albeit in small amounts. At this point, it was not conclusive if the smaller band observed here was a result of contaminant co-purification or degradation. In order to further investigate this, a series of MPH degradation studies were conducted.

Figure 3.14: Purity of syringe purified MPH with purification workflow IV after concentration. Molecular weights of proteins in molecular marker, from top to bottom are 97.4, 66.2, 45.0, 31.0, 21.5, 14.4 kDa.

3.2.2.6. MPH degradation study

To observe if further degradation occurred after purification, a series of MPH degradation studies were conducted. The outcome of the study might also be able to implicate the reason of degradation. For this stability study, MPH was expressed in *E. coli* DH5α or BL21(DE3)$^{RecA-}$ and purified with the purification workflows described in sections above. The purified MPH was then kept at different temperatures and a small aliquot of the enzyme was taken out at different intervals and subjected to SDS-PAGE to observe if degradation has taken place. The samples were clarified via centrifugation prior to sampling the aliquots to remove any aggregate or precipitate that has formed. Three purified MPH preparations were made and the conditions are summarised in

Table 3.1. The degradation profiles of the three samples are shown in **Figures 3.15**, **3.16** and **3.17**.

Table 3.1: The three purified MPH samples prepared for degradation study.

	Sample I	Sample II	Sample III
Expression system	pCY76 vector in *E. coli* DH5α	pET47b vector in *E. coli* BL21(DE3)$^{RecA-}$	pET47b vector in *E. coli* BL21(DE3)$^{RecA-}$
Expression condition	16 hours constitutive expression, 37°C	16 hours, 0.5 mM IPTG inductive expression, 37°C	16 hours, 0.5 mM IPTG inductive expression, 37°C
Purification workflow	III	IVa	IVb
Incubation temperature	-20, 4, 25, 37	-20, 4, 25, 37	25, 37

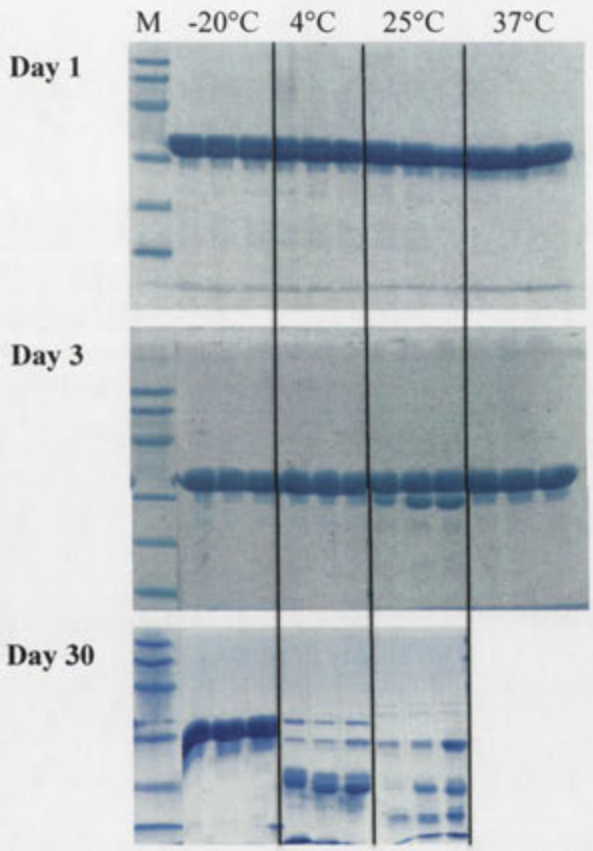


Figure 3.15: Degradation study of Sample I.
The purified MPH samples were incubated at -20°C, 4°C, 25°C and 37°C for up to 30 days with an aliquot of the sample taken out at days 1, 3 and 30. Day 0 was the start of the incubation. The molecular weights of proteins in the molecular marker, from top to bottom are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. Samples from each temperature were loaded into the gel in triplicates for this experiment.

From **Figure 3.15**, degradation can be observed for Sample I from day 3's 25°C and 37°C samples. At day 30, the MPH bands for 4°C and 24°C had almost totally disappeared. The degradation product that first appeared in day 3's samples had degraded into even smaller products. The sample that was kept frozen in -20°C

remained largely unchanged throughout the 30-day period. The 37°C day 30's sample was not available due to the shutdown of 37°C incubators for lab relocation.

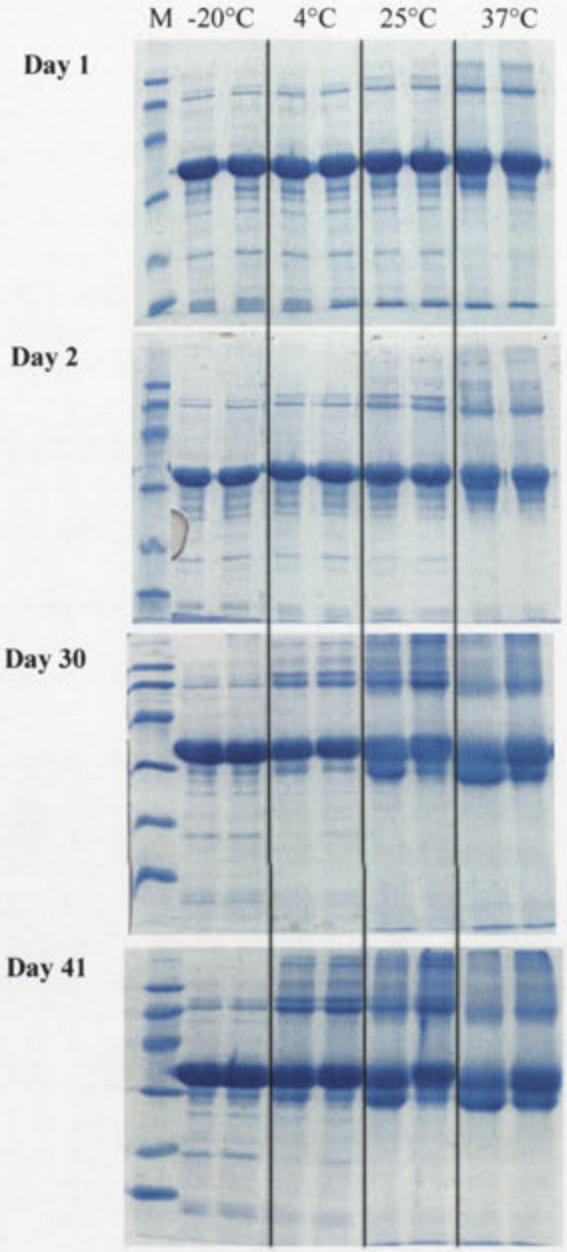


Figure 3.16: Degradation study of Sample II.
The purified MPH samples were incubated at -20°C, 4°C, 25°C and 37°C for up to 41 days with an aliquot of the sample taken out at days 1, 2, 30 and 41. Day 0 was the start of the incubation. The molecular weights of proteins in the molecular marker, from top to bottom are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa. Samples from each temperature were loaded into the gel in duplicates for this experiment.

For Sample II (**Figure 3.16**), degradation product was first observed for day 2's 37°C sample. By day 30, significant degradation product can be observed for 24°C and 37°C samples. On day 41, the amount of degradation product for 24°C and 37°C

samples remained largely similar. However, more degradation product can be observed for 4°C sample on day 41. Interestingly, the appearance of the degradation product coincided with the appearance of heavier bands, observed in the upper region of the SDS-PAGE profiles corresponding to 66 – 99.4 kDa of the molecular weight marker.

For Sample III (**Figure 3.17**), no significant degradation was observed for samples incubated at 25°C and 37°C throughout the incubation period. The SDS-PAGE profile of the samples remained largely similar by day 34 of the degradation study, indicating the source of proteolysis was eliminated. The heavier bands observed for sample II were also not observed in for Sample III.

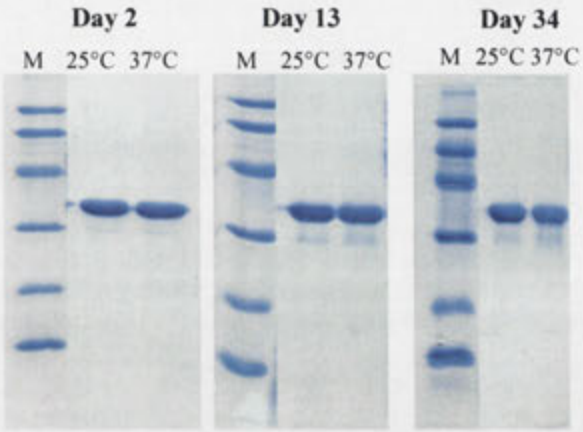


Figure 3.17: Degradation study of Sample III.
The purified MPH samples were incubated at 25°C and 37°C for up to 34 days with an aliquot of the sample taken out at days 2, 13 and 34. Day 0 was the start of the incubation. The molecular weights of proteins in the molecular marker, from top to bottom are 97.4, 66.2, 45.0, 31.0, 21.5 and 14.4 kDa.

3.3. Discussion

Proteolysis proved to be a difficult problem to troubleshoot due to the many possible routes in which proteolysis can occur. In this study, we narrowed down the possibilities for proteolysis to the protease from expression host or autoproteolysis. The possibility of N-end rule degradation was considered and ruled out due to the absence of destabilising residues at the N-terminus for the rule.[22,23] Due to the various types of proteases that are present in *E. coli,* as shown in the *E. coli* Protease Database, identifying the protease that could have been responsible for the degradation is no easy feat. One of the most common approaches to this problem was to use commercially available protease inhibitor cocktails that contain various protease inhibitors that would cover the different types of proteases such as serine, cysteine, aspatatic and

metalloproteases. However, the metalloprotease inhibitors in such cocktails might inhibit the activity of MPH, since MPH is a metalloenzyme. Therefore, the idea to engineer an MPH variant that is more resistant to proteolytic degradation was conceived. An enzyme that is more resistant against protease not only gives enzyme of better purity; it is also relevant to the application of MPH as a candidate for organophosphate bioremediation.

Based on the FTICR-MS analysis on a mixture of enzyme containing both full length and degraded MPH, the cleavage site of the proteolysis was predicted and an F88A MPH mutant was generated. However, the F88A mutant did not stop the degradation. This might be due to the mutation not performed on the protease recognition residues or the protease was tolerable to alanine at this position. F88A is the N-terminal residue from the predicted cleavage site (P1 position). Some proteases such as Asp-N endopeptidase recognise the C-terminal residue from the cleavage site (P1' position)[24]. In order for this approach to be successful, more than one mutation to the cleavage site region might be required.

Subsequently, various purification methods were attempted to eliminate the degradation product so that reasonably pure enzyme could be obtained for characterisation of WT MPH and variants isolated in **Chapters 4** and **5**. Purification workflow III that involved a SP Sepharose cationic exchange column and a Q Sepharose anionic exchange column produced acceptable purity but at the expense of yield and exposure of MPH enzyme to extreme pH. This led to the development of purification workflow IV that also used a HiTrap DEAE FF anionic exchange column and a HiTrap SP FF cationic exchange column. However, the different implementation from purification workflow III allowed the purification to be done rapidly at a neutral pH and this workflow was eventually used in the purification of the mutants isolated from the thermostability evolution (**Chapter 4**) and substrate specificity alteration evolution (**Chapter 5**). In the process of developing purification workflow III, it was also discovered that the pI of MPH appeared to be more alkaline than the calculated pI due to the limitation of ProtParam in predicting pI for highly basic proteins.[16] The purity obtained from purification workflow III was deemed sufficient for enzyme characterisation work and if the enzyme is stored in -20°C as quickly as possible, further significant degradation can be prevented. In terms of stability against

proteolysis, MPH did not suffer significant further degradation when kept in -20°C for at least a month, as demonstrated by the degradation studies for Samples I and II.

The degradation pattern shown by Sample I was different than that observed for Sample II and highlights the sensitivity of MPH to proteolytic degradation (**Figure 3.15**). The degradation in Sample I continued to progress and much smaller degradation products of various molecular weights can be observed. The degradation process also progressed more rapidly relative to Sample II. The MPH band had almost disappeared by day 30 of the degradation study. While the reason for this observation is unknown, the difference in degradation pattern might be related to the fact that Sample I was produce in *E. coli* DH5α. Compared to *E. coli* DH5α, *E. coli* BL21(DE3)$^{RecA-}$ does not express *lon* protease and *ompT* outer membrane protease.[25] While it was observed that MPH expressed in *E. coli* DH5α suffered more proteolysis (data not shown), there was no direct evidence implicating *lon and ompT* proteases to be responsible for the degradation observed here.

The different degradation profiles shown by Samples II and III shed some light on the degradation. This was because the proteins for both samples were expressed similarly and purified using the same column, the samples exhibited different degradation profiles. The only difference in preparing the samples was; Sample II was purified using syringe one-step elution, while Sample III was purified using FPLC linear gradient elution, and only one of the fractions was used for the degradation study (Fraction 19 from **Figure 3.13b**). No significant degradation was observed at the end of the experiment for Sample III while significant degradation can be seen from the SDS-PAGE profile for Sample II for the same period. This result implied that the degradation was most likely due to a protease from the expression host and was eliminated during gradient elution when preparing Sample III. For Sample II, the one-step syringe elution might not eliminate the protease since the bound MPH protein was eluted into a single fraction. As seen from the purity of purification workflow IVa (**Figure 3.14**), some background proteins were present in the purified enzyme, showing a lesser purity. As with Sample I, the proteases responsible for the degradation should only be present in very small quantities. If the proteases were present in large quantities, the degradation would happen more rapidly. Another possibility was the non-specific interaction between the proteases and MPH. The proteolysis persisted with purification workflows I (cationic exchange and gel filtration chromatography), II (cationic exchange and anionic exchange), III (anionic exchange and cationic exchange

chromatography) and IVa (Ni^{2+} affinity chromatography). It would seem unlikely that the proteases was so similar to MPH in terms of pI, molecular weight and binding capacity to Ni^{2+} charged column and co-purify with MPH in cationic exchange, gel filtration and Ni^{2+} charged columns, leading to the initial suspicion that MPH is capable of autoproteolysis.

The increasing intensity of higher molecular weight species observed for Sample II is a curious observation (**Figure 3.16**) The appearance of these higher molecular species coincided with the appearance of the degraded species and the disappearance of species with molecular weight lower than the degraded species. One hypothesis to explain this observation is the formation of soluble aggregates by the degraded species and lower molecular weight species. Insoluble aggregates were ruled out of consideration. Even though insoluble aggregates were observed for 37°C incubated samples, all samples were clarified by centrifugation before aliquots were removed for SDS-PAGE. This was consistent with the observation that although considerable degradation can be seen for the sample incubated at 37°C, lesser amount of higher molecular weight species was observed than that of 4°C and 24°C. It was however not clear as to why the soluble aggregate was not dissociated by SDS when subjected to SDS-PAGE if the hypothesis is true.

3.4. Summary

This chapter described the development and optimisation of MPH purification procedures. During the course of developing the purification protocols, it was discovered that MPH was susceptible to proteolytic degradation. The initial strategy to overcome proteolysis was to engineer MPH to be more resistant to proteolysis by site directed mutagenesis on the predicted cleavage site, based on data obtained from mass spectroscopy. However, the degradation persisted when MPH F88A was purified. One plausible reason was while the cleavage site was correctly predicted, the mutagenesis was not performed on key recognition residue by the protease or that the protease can tolerate the mutated residue.

Subsequently, a purification workflow was developed (purification workflow III) to produce purified MPH enzymes with acceptable purity and homogeneity for MPH variants characterisation in **Chapters 4** and **5.** In order to obtain more information

about the proteolysis, i.e., if the proteolysis was due to host protease or autoproteolysis, immobilised-metal affinity chromatography (IMAC) using Ni^{2+} charged column was performed. Two different elution methods were used (syringe elution and FPLC elution) and the difference observed from the degradation study of the two methods led us to conclude that host protease was the most likely culprit in MPH proteolysis. The different degradation pattern showed by MPH produced in *E. coli* DH5α highlighted the susceptibility of MPH to proteolytic degradation. Miss Nur Hafizah Azizan, in her work with investigating metal binding affinities of MPH, uses pETMCSI vector and BL21(DE3)$^{RecA-}$ for expression of MPH and should only see minimal degradation if the protein is purified and stored quickly.

3.5. References

1. Maurizi, M. R. Proteases and protein degradation in *Escherichia coli*. *Experientia* **48,** 178–201 (1992).
2. Murby, M., Uhlen, M. & Stahl, S. Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. *Protein Expression and Purification* **136,** 129–136 (1996).
3. Swamyt, K. H. S. & Goldberg, A. L. Subcellular distribution of various proteases in *Escherichia coli*. *Journal of Bacterology* **149,** 1027–1033 (1982).
4. Bagos, P. G., Nikolaou, E. P., Liakopoulos, T. D. & Tsirigos, K. D. Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics* **26,** 2811–2817 (2010).
5. Fu, G., Cui, Z., Huang, T. & Li, S. Expression, purification, and characterization of a novel methyl parathion hydrolase. *Protein Expression and Purification* **36,** 170–176 (2004).
6. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26,** 1608–1615 (2010).
7. Sun, L. *et al.* Crystallization and preliminary X-ray studies of methyl parathion hydrolase from *Pseudomonas* sp. WBC-3. *Acta Crystallographica. Section D, Biological crystallography* **D60,** 954–956 (2004).
8. Chakraborty, S., Minda, R., Salaye, L., Bhattacharjee, S. K. & Rao, B. J. Active site detection by spatial conformity and electrostatic analysis-unravelling a proteolytic function in shrimp alkaline phosphatase. *PloS One* **6,** e28470 (2011).
9. Yang, W. *et al.* Application of methyl parathion hydrolase (MPH) as a labeling enzyme. *Analytical and Bioanalytical Chemistry* **390,** 2133–2140 (2008).
10. Ali, M. H. & Imperiali, B. Protein oligomerization: how and why. *Bioorganic & Medicinal Chemistry* **13,** 5013–5020 (2005).
11. Xiao, Q., Zhang, F., Nacev, B. a, Liu, J. O. & Pei, D. Protein N-terminal processing: substrate specificity of *Escherichia coli* and human methionine aminopeptidases. *Biochemistry* **49,** 5588–5599 (2010).
12. Nelson, C. A. The binding of detergents to proteins: I. The maximum amount of dodecyl sulphate bound to proteins and the resistance to binding of several proteins. *Journal of Biological Chemistry* **246,** 2895–3901 (1971).
13. Shi, Y. *et al.* Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein Science* **21,** 1197–1209 (2012).
14. Wyss, M. *et al.* Biophysical characterization of fungal phytases (*myo*-inositol hexakisphosphate phosphohydrolases): molecular size , glycosylation pattern , and engineering of proteolytic resistance. *Applied and Environmental Microbiology* **65,** 359–366 (1999).
15. Markert, Y., Köditz, J., Mansfeld, J., Arnold, U. & Ulbrich-Hofmann, R. Increased proteolytic resistance of ribonuclease A by protein engineering. *Protein Engineering* **14,** 791–796 (2001).
16. Wilkins, M. R. *et al.* in *The Proteomic Protocols Handbook* (Walker, J. .) **112,** 571–607 (Humana Press Inc., 1999).
17. Bjellgvist, B. *et al.* The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14,** 1023–1031 (1993).

18. Bjellgvist, B., Basse, B., Olsen, E. & Celis, J. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **15,** 529–539 (1994).
19. Bornhorst, J. A. & Falke, J. J. in *Methods in Enzymology, Volume 326* **326,** 245–254 (Academic Press, 2000).
20. Sabaty, M. *et al.* Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. *BMC Biochemistry* **14,** 28 (2013).
21. Beckham, K. S. H., Byron, O., Roe, A. J. & Gabrielsen, M. The structure of an orthorhombic crystal form of a "forced reduced" thiol peroxidase reveals lattice formation aided by the presence of the affinity tag. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* **F68,** 522–526 (2012).
22. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254,** 1374–1377 (1991).
23. Varshavsky, A. The N-end rule pathway of protein degradation. *Genes to Cells* **2,** 13–28 (1997).
24. Keil, B. *Specificity of Proteolysis*. 336 (Springer-Verlag, 1992).
25. Ratelade, J. *et al.* Production of recombinant proteins in the *lon*-deficient BL21(DE3) strain of *Escherichia coli* in the absence of the DnaK chaperone. *Applied and Environmental Microbiology* **75,** 3803–3807 (2009).

DIRECTED EVOLUTION OF METHYL PARATHION HYDROLASE FOR ENHANCED THERMAL STABILITY

4. Directed Evolution of Methyl Parathion Hydrolase for Enhanced Thermal Stability

4.1. Preamble

As mentioned in **Chapter 1**, organophosphate-degrading enzymes have received considerable attention due to their potential applications in bioremediation of organophosphate (OP) contamination, biosensor for OP detection and therapies for OP poisoning.[1] Methyl parathion hydrolase (MPH), as a newer OP degrading enzyme, has aroused considerable interest due to its potential to be used as an alternative to OPH and OPDA from *Pseudomonas diminuta* and *Agrobacterium radiobacter* respectively.

Stability is one of the most desired characteristics when considering an enzyme's suitability to be used in commercial or industrial application. Despite the impressive catalytic properties of many enzymes, they are frequently not very stable, limiting their applicability in industry. In order for MPH to be an effective bioremediator, the enzyme has to be stable – showing good thermal and chemical stabilities, as well as the possibility of repeated use to keep the production cost low. Unfortunately, MPH does not have exceptional thermostability and could benefit from extra thermostability. One approach is to raise the overall stability with directed evolution. While there are many successful examples of enzyme evolution to improve thermal properties, there are no clear and universal rules to improve protein thermal stability.[2] Yana and Poulos compiled a detailed review of important factors for enhancing protein stability. They surveyed the literature and illustrated the many possible ways by which thermostability enhancement could be achieved.[3] They also concluded that there are many routes to increase thermal stability, and directed evolution can offer alternative routes to those used by natural evolution.

It has been suggested that enhanced stability promotes the evolvability of the protein.[4–6] By "evolvable", I mean that the function of a protein can be altered with mutations and the altered phenotypes are heritable. One major factor that is thought to contribute to evolvability is mutational robustness. Wagner defined mutationally robust biological system as a system that exhibits persistence in its structure or function after mutations of its parts.[7] In order for a protein to function, it must fold into its native

structure, and retention of that native structure is a prerequisite to obtain new or modified functions.[4] Bloom and co-workers proposed that enhanced thermodynamic stability can contribute to mutational robustness of a protein by demonstrating that a stabilised TEM-1 mutant, when used as the template for error-prone PCR, showed higher mutational tolerance compared to the wild type template.[6] Enhanced thermostability increases the robustness of a protein by increasing the fraction of mutants that possess the minimal stability required to fold.[5] Bloom and co-workers further illustrate the link between thermostabiliy and evolvability by using thermostable P450 variants and marginally stable variants as starting points for directed evolution toward a new range of P450 substrates.[4] The thermostable variants produced more variants with improved catalytic activity as only the thermostable variants can accommodate the highly destabilising but beneficial mutations to attain new functions. In addition, active site residues are inherently destabilising, as these residues must often satisfy functional constraints.[4,8] Accumulation of destabilising mutations was also found to be negatively epistatic, causing the protein fitness to exponentially decline as more mutations are introduced in protein evolution.[5,9] Therefore by generating thermostable MPH variants, these variants can be used as starting points in future evolution experiments for obtaining new functions through mutations that are too destabilising for wild type MPH to accommodate.

The objective of the work described in this chapter was to enhance the thermal stability of MPH by employing directed evolution techniques such as error-prone PCR and DNA shuffling. During the course of this work, the zinc content of MPH, as well as the effects of additives on MPH was also investigated.

4.2. Directed evolution experimental procedures

A schematic flowchart of the steps taken to evolve for thermostability is given in **Figure 4.1**. One round of evolution typically consists of library construction (**section 4.2.1**), library growth (**section 4.2.2**) and library screening (**section 4.2.3**). The screening procedures and strategies used in this section were adopted and modified from various references on directed evolution and thermostability evolution.[10–12]

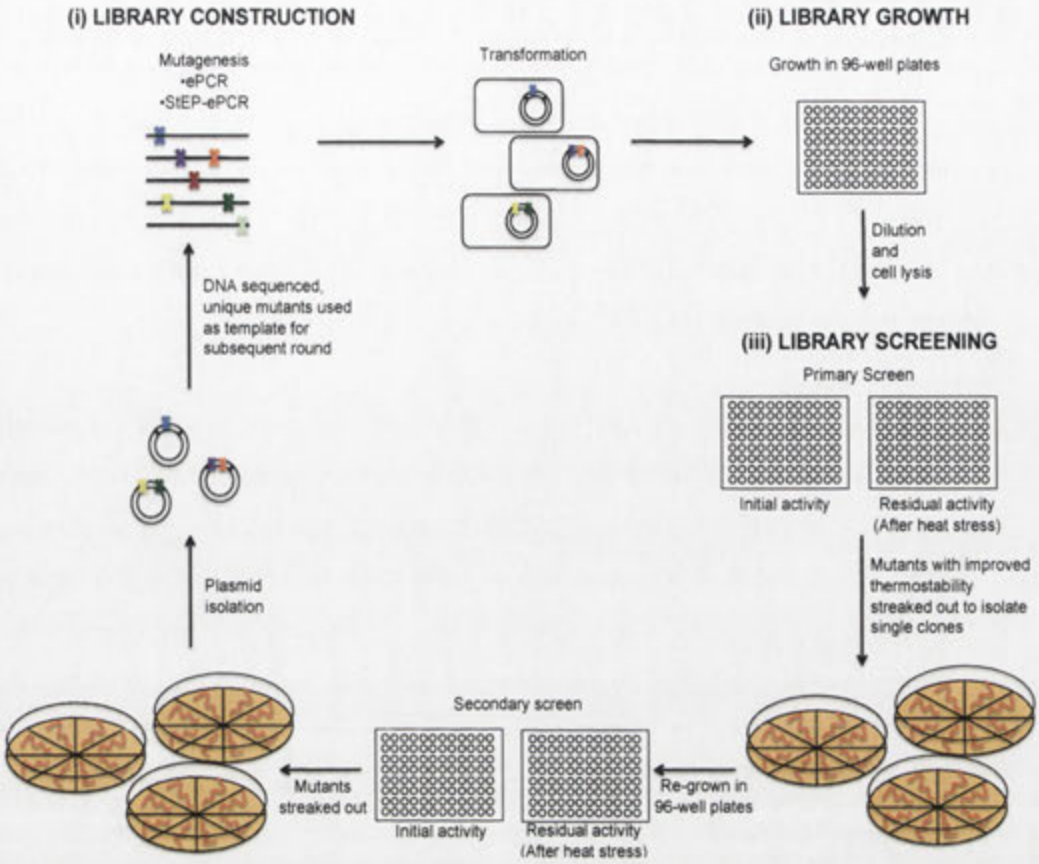


Figure 4.1: Thermostability screening overview.
Schematic of the screening process used in directed evolution of MPH thermostability. The mutants were selected based on their thermostability index, calculated using Equation 4.1.

4.2.1. Library construction

For the work described in this chapter, the first 34 amino acid residues of the intact MPH sequence that constitute the signal sequence were removed and designated M1T for the reasons described in **section 3.2** (**MPH-1-Trucncation**). One factor that had to be considered during library construction was the choice between a constitutive vector (pCY76) and an inducible vector (pETMCSI). pCY76 vector was chosen as the vector due to three main reasons:

i. Electrocompetency of *E. coli* BL21(DE3)$^{RecA-}$ was found to be at least 10-fold less than *E. coli* DH5α[13]. pETMCSI is a modified T7 pET vector that requires *E. coli* BL21 (DE3)$^{RecA-}$ for heterologous protein expression to work, while pCY76 protein expression can work in *E. coli* DH5α. Using pETMCSI and *E. coli* BL21 (DE3)$^{RecA-}$ for library generation could result in cumbersome steps in order to generate large libraries.

ii. The yield of pETMCSI obtained during plasmid extraction is found to be much lower that that of pCY76 (data not shown). Bigger volumes of cell culture are often required for plasmid isolation, when using pETMCSI vector to obtain sufficient amounts of plasmid for DNA sequencing or DNA digestion. While this is not be a problem when isolating plasmid for only a few samples, directed evolution work often involves plasmid isolation of a large number of samples and this would lead to very laborious steps in plasmid isolation.

iii. MPH is a highly efficient enzyme and very little enzyme is required for activity measurement. When an inducible protein expression system is used, large amount of enzyme would be produced. Using cell lysate from this preparation, even in very small volumes, would result in a reaction rate that is too rapid to measure. Multiple cell dilution steps would be required to obtain measurable initial activity and this could result in an even more laborious screening procedure. Constitutive protein expression system gives lower level of protein expression and can reduce the number of dilution steps required, simplifying the screening workflow.

The pCY76_M1T_WT construct was constructed similarly to pETMCSI_M1T_WT and was described in **section 3.2.** Error-prone PCR (ePCR) and staggered extension process error-prone PCR (StEP-ePCR) were used to introduce mutations into M1T. ePCR was used in the first round of evolution and StEP-ePCR was used in rounds two to four to shuffle selected mutations, in addition to introducing new mutations. The ePCR and StEP-ePCR steps were performed as described in **sections 2.5.1 and 2.5.3.** The mutagenic PCR products were digested with *Ase*I and *Eco*RI and ligated into pCY76 digested with *Nde*I and *Eco*RI as described in **sections 2.3.6 and 2.3.8.** The ligated plasmid was transformed into *E. coli* DH5α and the library was analysed with enzymatic assay based on procedures described in **sections 2.3.9 and 2.5.5.**

4.2.2. Library growth

The library master plates were grown in round bottom 96-well microtitire plates at 37°C for 24 hours in an orbital shaking incubator. The plates were prepared either by picking single colonies or by culture dilution method developed by Dr. Bradley

Stevenson[13], and both methods were described in **section 2.5.6.** During the course of this work, it was observed that the cultures in wells located at the edges of the plates underwent evaporation, causing inconsistencies in cell density. To minimise this problem, the plates were sealed by petroleum jelly and masking tape prior to incubation.

4.2.3. Library screening

After the plates were removed from the incubator, the cells were resuspended manually with multi channel pipettes to ensure that the cultures were homogenous. 50 µL of cells was transferred from the growth plates (master plate) to round bottom 96-well microtitre plates (lysis plates) and the lysis plates were kept in -80°C until required. 50 µL of 50% (v/v) glycerol (final concentration 12.5%) was added to the remaining culture in master plates and kept in -80°C for storage. To lyse the cells, the lysis plates were removed from -80°C and were left to stand at room temperature to thaw. Freeze thawing is part of the protocol for cell lysis with rLysozyme™ (Novagen). 150 µL of 50 mM Tris-HCl, 100 mM NaCl, 0.1% Tween 20, pH 7.6 containing 0.24 U/µL rLysozyme™ was added to the lysis plates to initiate lysis and the plates were incubated on a plate shaker for 30 minutes at 24°C. After lysis, the plates were centrifuged at 3700 rpm for 15 minutes to pellet the cell debris. 50 µL of supernatant was transferred to 96-well PCR plates (heat treatment plate), and 20 µL of supernatant was transferred from the heat treatment plates to 96-well flat bottom microtitre plates (initial activity plate). The heat treatment plates were heated with a thermocycler and were left to stand at room temperature for at least 30 minutes. The conditions of the heat treatment are discussed in **section 4.2.3.2.** 20 µL of lysate was then transferred to 96-well flat bottom microtitre plates for activity assay (residual activity plate).

4.2.3.1. Choosing the screening substrate

While the most active substrate for MPH is methyl parathion (MPS), using MPS as the screening substrate in this work has several disadvantages. Due to the high activity MPH has on MPS, obtaining the initial velocity with low variability was difficult. The initial velocity cannot be measured accurately if too much enzyme is used, as the reaction goes to completion before measurements can be made. When using such active substrates for screening, one can use very small amount of lysate (1 – 2 µL), or diluting the lysate to a more dilute concentrations, so that the initial velocity can be

measured. However, pipetting very small volumes would introduce pipetting errors that affect the variability of the measurements, while introducing dilution steps into the already complicated screening workflow would also affect the accuracy and throughput of the screen. With these considerations in mind, ethyl parathion (EPS) was used as the screening substrate in this work. As discussed in **Chapter 5**, MPH was approximately 10-fold less active toward EPS and that would decrease the number of dilution steps needed to achieve measurable initial velocity. The libraries were assayed with 119 μM EPS in 50 mM Tris-HCl, 100 mM NaCl, 0.1% Tween 20, pH 7.6 and the reactions were monitored by the liberation of *p*-nitrophenolate at 405 nm at 30°C.

4.2.3.2. Choosing the screening condition

Prior to screening the mutant libraries, the thermostability of WT MPH was assessed by constructing a melting curve for WT MPH. A plate containing WT MPH was prepared as described in **section 4.2.2** and was assayed as described in **section 4.2.3,** except that 25 μL of clarified lysate was transferred from the lysis plate to PCR tubes for different heat treatments. Aliquots left at 4°C served as the controls without heat treatment (initial activity, A_i) and samples were heat treated from 20°C to 100°C for 15 minutes to obtain the residual activity after heat treatment. Three wells were picked randomly for the assay to represent the technical replicates for this experiment. For this work, the stability (s) of the protein is defined by the percentage of residual activity (A_r) to initial activity (A_i), as given by **Equation 4.1.**[12] The WT MPH thermostability profile is presented in **Figure 4.2.**

$$s = \frac{A_r}{A_i} \times 100\% \qquad (4.1)$$

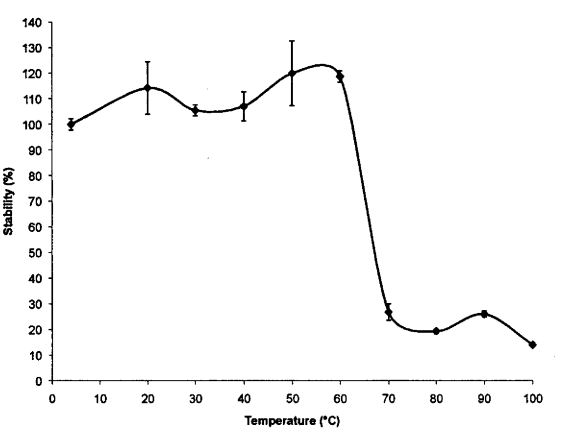


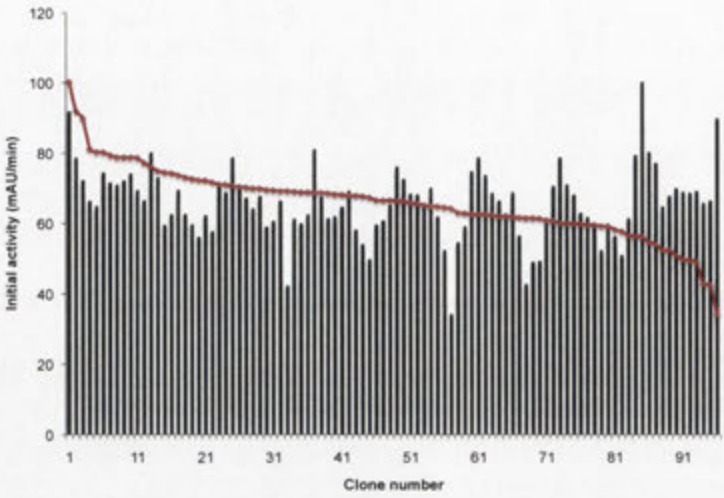
Figure 4.2: The thermostability profile of wild type MPH obtained with crude lysate.
The error bars represent the standard deviation of the sample in triplicates.

From the melting curve, it was observed that the stability decreases rapidly from 60°C to 70°C. Thus, 70°C was chosen as the heat treatment condition for library screening. Note that the heat treatment did not result in total inactivation as low levels of activity were observed even for higher temperatures.

After all the parameters have been considered, a screening validation assay was performed to assess the variability of the screen.[10] It is important to quantify the quality of a screening method as screening variability can influence the detection of real improved variants. Repeated measurements of a single variant library (i.e., wild type MPH) was used to quantify the reproducibility and uncertainty of a screen. The quality of a screening method was evaluated by calculating the coefficient of variation (CV) and plotting the activity distribution of the single variant library. The coefficient of variation is given by the percentage of standard deviation to the mean of the activity measurements (CV = [standard deviation / mean] x 100%). The ideal CV value would be less than 10%[10]. However, it is difficult to achieve such low CV values without automated high throughput screening robots. With that said, the screening workflow was optimised as much as possible to achieve the lowest CV value with the existing laboratory setup.

To assess the variability of the screen, WT MPH was grown in a 96-well microtitre plate, and was assayed as described in **sections 4.2.2** and **4.2.3**, identical to the screening workflow for mutant library. The heat treatment was 70°C for 15 minutes and was done as described in **section 4.2.3.**

(a)



(b)

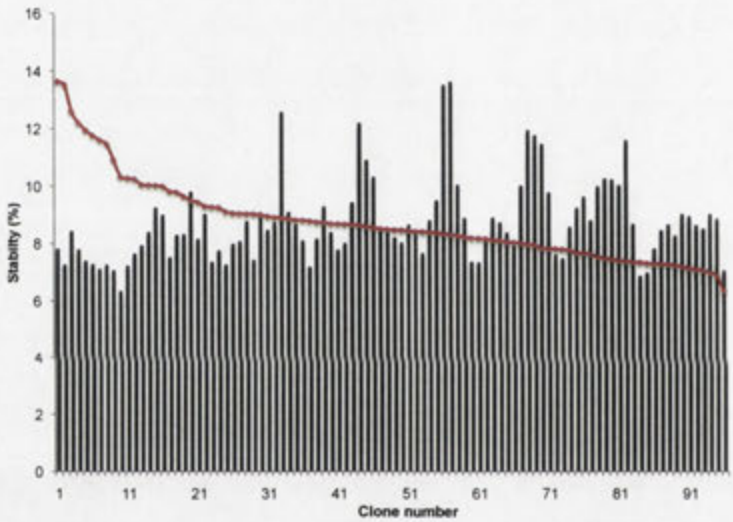


Figure 4.3: Evaluating the screening method with a plate of 96 wild type MPH samples. (a) shows the initial activity distribution of WT MPH and (b) shows the stability distribution of WT MPH.
In both (a) and (b) plot, the bar charts represent the initial activity and stability of WT MPH with respect to their positions in the 96-well plate. The line graphs in both (a) and (b) show the same initial activity and stability data sorted in descending order.

The CV values for the WT MPH initial activity and stability were 14.4% and 15.4% respectively. The initial activity and stability distributions of the single sequence library are shown in **Figure 4.3.** By plotting the WT MPH stability distribution data for the assay method, the sensitivity of the screen can be obtained by dividing the highest (13.6%) and lowest (6.3%) WT MPH stability indices obtained from the validation screen. This method of assessing the quality of the screening method was described by Salazar and Sun in a textbook.[10] **Figure 4.3,** a CV value of 15.4% allowed us to identify

mutants with at least 2.2-fold stability improvement. A more sensitive screen (smaller CV values) would be required if smaller improvements are to be detected.

4.3. Results

4.3.1. Directed evolution of MPH

Four rounds of directed evolution, consisting of one round of ePCR and three rounds of StEP-ePCR, were performed on MPH to isolate variants with improved thermostability. The results for the directed evolution are discussed in this section of the chapter.

4.3.1.1. Screening and selection

In all rounds of the evolution, mutants with less than 60% wild type MPH activity at 30°C (initial activity) were removed from further consideration. This was to prevent the selection of mutants with improved thermostability but hugely compromised initial activity. Mutants with low initial activity would also generate artificially high thermostability values contributing to false positives.[12]

Examples of the microplate screening plots are shown in **Figure 4.4**. A summary of the library generation methods, growth methods, library sizes, error rates and selection criteria for the evolution in each round can be found in **Table 4.1.** Unique mutants selected at the end of round four are displayed in **Table 4.2** and unique mutants selected from all four rounds can be found in **Appendix F.** In each round, the libraries were subjected to at least a primary and a secondary screen to identify mutants with enhanced thermostability. In rounds two and three, a tertiary screen was introduced to further eliminate false positive clones. The unique mutants, isolated from each round based on the criteria listed in **Table 4.1,** were used as the parent for the subsequent round of evolution. In round four, the heat treatment pressure was increased in an attempt to isolate mutants with further thermostability enhancements. After four rounds of evolution, all mutants carry the L256Q and K316E double mutations, suggesting that the evolution had converged under the experimental conditions that were used.

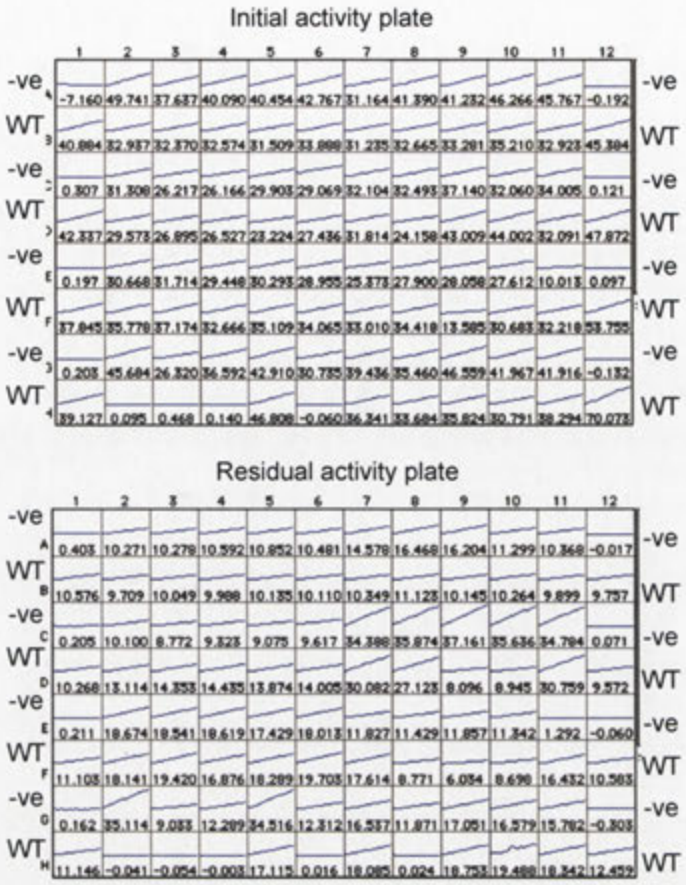


Figure 4.4: Examples of screening plots from microplate screens.
In this example of a primary screen, the initial and residual activities of the same master plate were measured and the stability value of every variant was calculated using Equation 4.1. The stability of the mutants were then compared to that of wild type and those with better stability, while maintaining activity at lower temperatures, were selected for secondary screening. The wild type and vector only controls were located in columns 1 and 12 of the plates in an alternate pattern, as labelled in the figure.

Table 4.1: Summary of the library generation methods, growth methods, library sizes, error rates and selection criteria for MPH thermostability evolution.
The error rate for rounds two to four show the introduction of new mutations in addition to DNA shuffling.

				Error rate per gene		Selection criteria		
Round	Mutagenesis method	Primary screen growth method	Library size	Nucleotide	Amino Acid	Heat treatment	Initial activity	Stability
1	ePCR	Culture dilution	3480	3.22 ± 3.63	2.13 ± 2.8	70°C, 15 mins	> 70% WT	> 100% WT
2	StEP-ePCR	Culture dilution	2880	1.11 ± 1.27	0.67 ± 0.71	70°C, 15 mins	> 60% WT	> 100% WT
3	StEP-ePCR	Culture dilution	2880	0.44 ± 1.01	0.44± 1.01	70°C, 15 mins	> 80% WT	> 200% WT
4	StEP-ePCR	Single colony	1600	1.45 ± 1.29	1.09 ± 0.99	80°C, 20 mins	> 80% WT	> 400% WT

Table 4.2: Mutants isolated after round four of thermostability evolution.
The numbers and alphabets at the top of the table indicate the position of the mutations and the wild type residues at the corresponding position. The stability indices of the mutants were normalised to that of WT MPH.

	90	113	164	202	211	237	256	275	278	281	300	316	Relative stability (%)
WT	**K**	**T**	**F**	**A**	**G**	**G**	**L**	**S**	**K**	**A**	**A**	**K**	**100**
Round 4													
1A4	E						Q					E	571
1D5							Q				G	E	588
1 E6		A					Q			V		E	644
5H6							Q					E	553
6A10							Q				G	E	485
6B3							Q					E	483
6C11							Q					E	523
8A6		A			D	D	Q					E	535
8C8							Q					E	494
8H6							Q					E	486
9B8				S			Q	I				E	582
11A3			Y				Q					E	558
10H2							Q					E	530
10G8							Q					E	555
13D4							Q					E	497
13D11	E						Q					E	515
13G6				S			Q			T		E	536
13G9							Q		N			E	515
14A7							Q					E	479
14H9							Q				G	E	496
15 E2							Q					E	513
15F3							Q			T		E	564
15H5							Q					E	477
17F11							Q			V		E	677
16A4							Q					E	636
20A6	E						Q					E	619
16H7							Q		N			E	582
17 E10							Q				G	E	603
20H9							Q					E	660

4.3.1.2. Thermal activation of MPH variants

During the course of library screening, it was observed that mutants carrying L256Q and K316E double mutations consistently exhibit thermal activation after heat treatment **(Figure 4.5).** From the round four secondary screening data, R413D4 that carries the L256Q and K316E double mutation gave 83.9% activity increase after heating at 80°C for 20 minutes, exhibiting thermal activation.

The thermostability profile of WT MPH was subsequently investigated with purified enzyme. The enzyme was expressed in *E. coli* BL21(DE3)$^{RecA-}$ harbouring pETMCSI_M1T_WT as described in **section 2.6.1** and was purified with 5 mL HiTrap DEAE FF and 5 mL HiTrap SP FF columns as described in **section 3.2.2.4**. To obtain the thermostability profile using purified enzyme, the enzyme was heat treated from 40°C to 100°C for 15 minutes prior to activity measurement. The assay mixture

contains 134 nM WT MPH enzyme and 119 μM ethyl parathion in 50 mM HEPES, 100 mM NaCl, pH 7.6. The thermostability profile of the purified enzyme is shown in **Figure 4.6.** From the figure, the WT MPH activity increased after heat treatment at temperatures 40°C to 60°C before declining rapidly. Similar to the melting curve constructed with crude lysate **(Figure 4.2)**, MPH did not undergo total inactivation at higher temperatures.

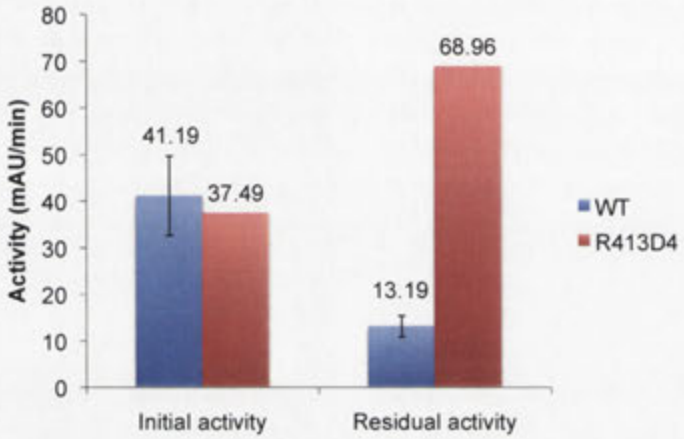


Figure 4.5: Initial and residual activities comparison between R413D4 that carries the L256Q and K316E double mutations and wild type MPH in the round four secondary screen.
The residual activities were obtained by heating the lysate at 80°C for 20 mins. The error bars for WT MPH represent the MPH WT replicates on the same screening plate.

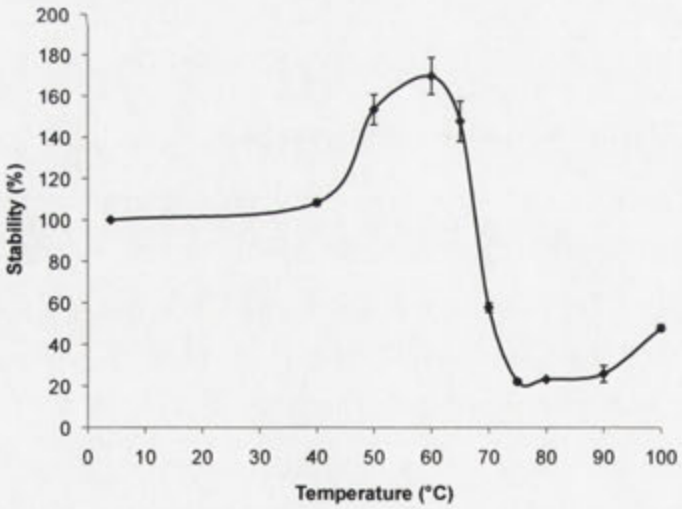


Figure 4.6: The thermostability profile of wild type MPH obtained with purified enzyme.
The error bars represent the standard deviation of the sample in triplicates.

The observation of thermal activation in WT MPH prompted a search of literature for similar observation. Thermal activation is a feature found in various thermophilic enzymes. Catalase I from *Bacillus stearothermophilus* was found to be

irreversibly converted to high-activity form upon heating.[14] Cyclodextrin glucanotransferase from *Thermococcus* sp. B1001 underwent changes in activity profile due to conformational change upon heating.[15] The cyclisation activity of the enzyme increases while the hydrolysis activity decreases after *in vitro* heat treatment leading to suggestions that thermal environment is important for the folding of the thermophilic enzyme.

4.3.2. Characterisation of MPH variants

In addition to R413D4 (**L256Q**, **K316E**), the stability indices of R31A10 (**L256Q**), R32C11 (T113A, **K316E**), R33D5 (K90E, **L256Q**), R31E3 (**L256Q**, A300G, and **K316E**), R33H8 (K90E, **L256Q**, and **K316E**) at 70°C and 80°C were reassessed to the shortlisted variants for purification and characterisation. From the library screening data, it was observed that variants carrying L256Q and K316E mutations gave the most improvement over WT MPH. R31E3 and R33H8 were selected to investigate if K90E and A300G would contribute to any further thermostability enhancement on top of L256Q and K316E mutations. R31A10, R32C11 and R33D5 were selected to reaffirm the observation that both L256Q and K316E are required for significant thermostability enhancement. R413D4 was the only mutant selected for purification and subsequent characterisation since the thermostabilities of R31E3 and R33H8 were found to be similar to R413D4 (data not shown).

In order to further investigate the properties of the isolated double site mutant, R413D4 was subcloned into pETMCSI vector for protein expression, purification and characterisation. This was done to maintain consistency as WT MPH purified enzyme used for thermostability characterisation was expressed in BL21(DE3)$^{RecA-}$. The subcloning and protein expression were done as described in **section 2.6.1** and purification was done as described in **section 3.2.2.4.**

4.3.2.1. Activity variation between batches of purified MPH

During the course of protein purification, it was discovered that WT MPH enzyme of different purification batches exhibit variation in activity (**Figure 4.7**). The activity assays were performed as described in **section 4.3.1.2.** This would affect the thermostability analysis of the enzyme since the thermostability value would depend on the initial activity of the enzyme. Enzyme concentrations were measured with a

Nanodrop spectrophotometer in triplicates as described in **section 2.6.4.** The triplicate measurements showed good agreement, indicating that the measurements were accurate, while repeating the dilutions and assays also showed similar trend, suggesting that the dilutions were done properly and the activity variation was not due to dilution error (data not shown). An SDS-PAGE profile was used to visually assess any concentration variation that might be present for enzyme used for the assay **(Figure 4.8)** The SDS-PAGE was loaded with the same amount of enzyme used in the assay in triplicates. While SDS-PAGE is not suitable for quantitative protein measurement, the differences in activity between some of the purification batches should be large enough to be observed on a gel. Batch 1 MPH WT is approximately 3-fold and 2-fold less active than Batch 2 and Batch 3 respectively. From the SDS-PAGE profile (**Figure 4.8**), no apparent difference in terms of enzyme amount can be observed, further suggesting that the enzyme loads used in measuring the activity were the same and the activity variation was caused by a different factor.

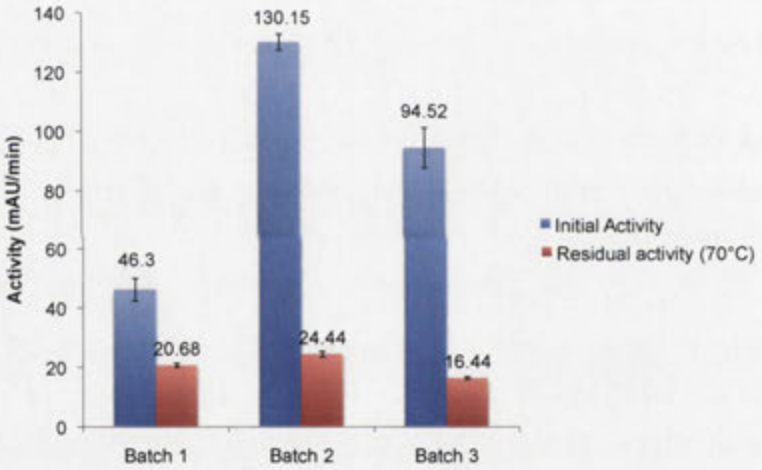


Figure 4.7: Initial and residual activities comparison between three batches of similarly purified MPH WT enzyme.
The blue columns show the initial activity of MPH WT enzyme without heat treatment while the red columns show the residual activity after heat treatment at 70°C for 15 minutes. The error bars represent the standard deviation of the quadruplicate measurements. Activity against EPS was measured as described **section 4.3.2.1.**

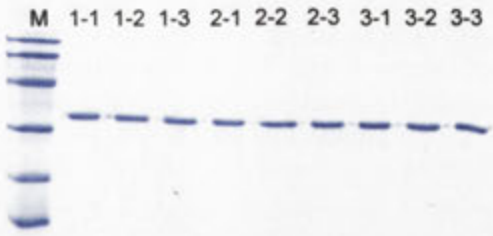


Figure 4.8: SDS-PAGE profile of the enzyme used for activity assay.
Three batches of purified WT MPH were each diluted to 2.68 μM and 10 μL of each of the enzyme was loaded onto the gel. 1-1, 1-2, and 1-3 are the replicates from one preparation; 2-1, 2-2 and 2-3 are the replicates from one preparation and so on. The molecular weights of protein in the molecular marker (M), from top to bottom are 97.3, 66.2, 45.0, 31.0, 21.5, 14.4 kDa.

4.3.2.2. Selected variant does not display the same enhancement observed in screening

To obtain a simple thermostability profile of the MPH variants, the WT MPH and R413D4 purified enzyme were incubated at 70°C and 80°C for 15 minutes prior to activity assay. The samples incubated at 4°C serve as the control without heat treatment and the data are presented in **Figure 4.9.** The assay was performed as described in **section 4.3.1.2.** While R413D4, which carries the L256Q and K316E double mutations, is more stable than WT MPH, the variant did not display the thermal activation characteristic seen during library screening. The stability obtained with purified enzyme for R413D4 was also much less compared to the stability obtained during screening. R413D4 was only 1.31 and 1.74 folds more stable than WT MPH after heat treatment at 70°C and 80°C respectively. In terms of activity, R413D4 retained 91.30% of WT activity **(Figure 4.9 inset).**

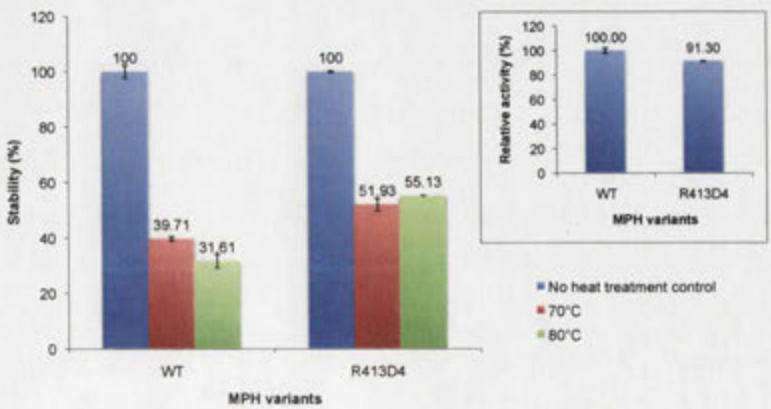


Figure 4.9: Stability comparison between purified WT MPH and R413D4.
The main panel shows the stability of WT MPH and R413D4 after heat treatment at 70°C (red) and 80°C (green). The samples incubated at 4°C (blue) were used as the non-heat treated control. The inset shows the activity comparison without heat treatment between WT MPH and R413D4, with WT MPH being 100%. The error bars represent the standard deviation of duplicate measurements in an experiment.

4.3.2.3. Enzyme pre-incubation with Zn^{2+} enhances activity and thermostabilty

In an attempt to explain the variation between results obtained with crude lysate and those obtained with purified enzyme, as well as the activity variation between batches of purified enzyme, it was decided that the effect of Zn^{2+} on MPH would be investigated. To illustrate this, the thermostability profile study of WT MPH and R413D4 was repeated with 0.1 mM $ZnCl_2$ pre-incubation prior to heat treatment. 0.1 mM $ZnCl_2$ was the chosen concentration as Zn^{2+} will form $Zn(OH)_2$, a white precipitate, in the presence of alkali. The assay buffer, 50 mM Tris-HCl or HEPES, contains alkali, as the buffer was pH adjusted to 7.6 with 10 M NaOH during preparation. A white precipitate could be observed when the Zn^{2+} concentration was higher than 0.2 mM. One way of overcoming the zinc metal concentration restriction was to avoid using hydroxides to titrate the buffers. The samples were pre-incubated with Zn^{2+} at 4°C for two hours prior to heat treatment. This pre-incubation period was chosen, as there was no considerable activity change after two hours of incubation (data not shown). The presence of 0.1% Tween 20, a component of the assay buffer for library screening, was also found to reduce the solubility of Zn^{2+} and was excluded from the assay buffer recipe in further assays. Tween 20 was included in the assay buffer to help solubilise phosphorothioate substrates that have low solubility in aqueous solution. As a result, methyl chlorpyrifos oxon that is more soluble was used as the substrate in subsequent thermostability characterisation assays. The concentration of enzyme used in assays was reduced to 67 μM, as MPH is more active towards methyl chlorpyrifos oxon. 0.1mM $ZnCl_2$ was also included in the assay buffer for metal pre-incubated samples to ensure that the metal is always available in excess.

The activity and stability comparisons between WT MPH, $ZnCl_2$ pre-incubated WT MPH, R413D4 and $ZnCl_2$ pre-incubated R413D4 are shown in **Figure 4.10**. The stabilities and activities of WT MPH and R413D4 were similar to those discussed in **section 4.3.2.2** for WT MPH and R413D4. R413D4 is 1.29 and 2.06 folds more stable than WT MPH after heat treatment at 70°C and 80°C respectively. R413D4 retained 87.59% WT MPH initial activity, consistent with the results shown in the previous section. However, upon pre-incubation with 0.1 mM $ZnCl_2$, the initial activity for WT MPH increased to 371.15%. The stabilities for metal pre-incubated WT MPH after heat treatment at 70°C and 80°C were 419.17% and 3.72%, a drastic difference from the WT

MPH enzymes assayed without additional metals. The stability of $ZnCl_2$ pre-incubated WT MPH after heat treatment at 80°C was also notably less than WT MPH without addition metals.

While R413D4 was as active as WT MPH before supplementing metals, the initial activity of R413D4 increased to 272.42% when it was pre-incubated with $ZnCl_2$, a smaller increase compared to that of WT MPH. The stabilities of $ZnCl_2$ pre-incubated R413D4 after heat treatments at 70°C and 80°C were 4.94% and 0.70% respectively, much lower than that of R413D4 without additional $ZnCl_2$.

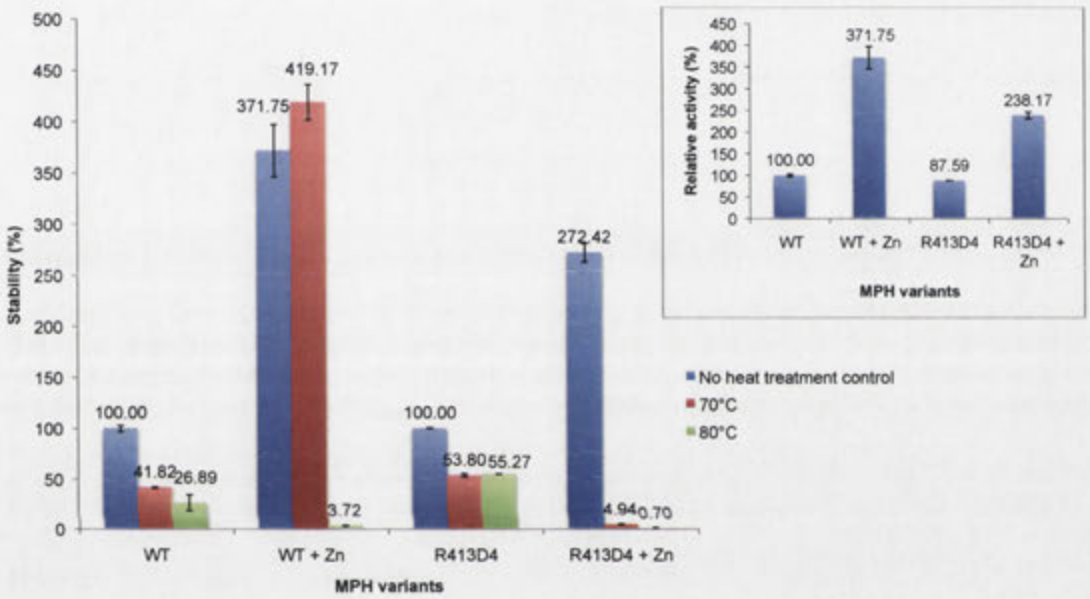


Figure 4.10: Stability comparison between purified WT MPH, WT MPH pre-incubated with 0.1 mM $ZnCl_2$. R413D4 and R413D4 pre-incubated with 0.1 mM $ZnCl_2$.
The main panel shows the stability of WT MPH, $ZnCl_2$ pre-incubated WT MPH, R413D4 and $ZnCl_2$ pre-incubated R413D4 after heat treatment at 70°C (red) and 80°C (green). The samples incubated at 4°C (blue) were used as the non-heat treated control. The inset shows the activity comparison without heat treatment between WT MPH, $ZnCl_2$ pre-incubated WT MPH, R413D4 and $ZnCl_2$ pre-incubated R413D4 with WT MPH being 100%. The error bars represent the pooled standard deviation of two independent experiments.

It is worth noting that for both WT MPH and R413D4 variants; samples without additional metals did not undergo total heat inactivation even as the samples were heated at higher temperatures. However, for $ZnCl_2$ pre-incubated samples, they were heat inactivated at higher temperatures. This is illustrated in **Figure 4.11** with the thermostability profiles of WT MPH and $ZnCl_2$ pre-incubated WT MPH. After being heat-treated at 80°C, metal pre-incubated WT MPH was inactivated while WT MPH without additional metal retained low level of activity. In addition, the low level of

activity for WT MPH without additional metal was observed even after 90°C and 100°C heat treatment. This phenomenon was observed despite $ZnCl_2$ pre-incubated WT MPH is more active and more stable up to 70°C.

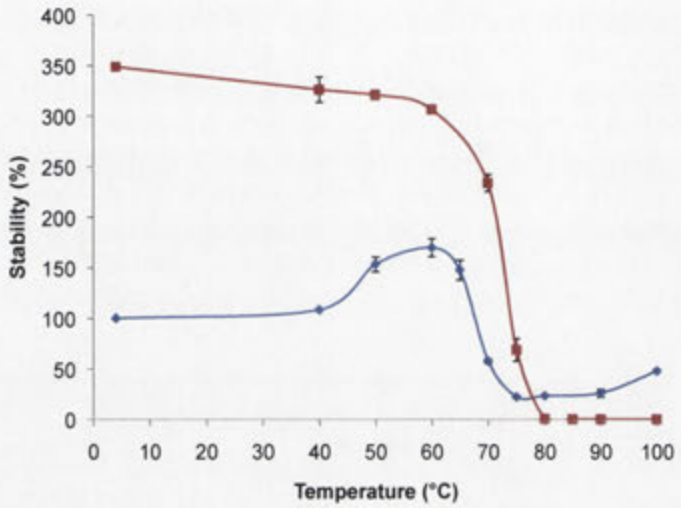


Figure 4.11: The thermostability profiles of WT MPH and $ZnCl_2$ pre-incubated WT MPH purified enzymes.
The blue line indicates the stability of WT MPH and the red line indicates the stability of $ZnCl_2$ pre-incubated WT MPH. The stability values were calculated using WT MPH activity incubated at 4°C without metal supplementation as the initial activity. The samples were heated for 15 minutes at various temperatures to obtain the profile The error bars represent the standard deviation of the samples in triplicates.

4.3.2.4. Zinc composition of wild type MPH

Due to the effects of excess Zn^{2+} on the activity and thermostability of MPH variants, the zinc content of WT MPH was investigated. WT MPH enzyme was diluted from concentrated stock to 20 μM with 50 mM HEPES, 100 mM NaCl, 0.1 mM $ZnCl_2$, pH 7.6 and was incubated at 4°C for 16 hours. The enzyme concentration was carefully chosen with two considerations: a concentration high enough that it will still be measurable after dilution, and a concentration low enough that the amount of Zn^{2+} would be in excess of the enzyme. This preparation would give a metal: enzyme ratio of 5:1. Another WT MPH sample was similarly diluted with the exception that the buffer did not contain $ZnCl_2$. This was done to assess the metal content of WT MPH as purified. The excess $ZnCl_2$ was removed by 5 mL HiTrap Desalting column and the samples were prepared for metal content measurement as described in **section 2.6.6.** The zinc content of MPH was measured by inductively coupled plasma-optical emission spectroscopy (ICP-OES) and was performed by Ms. Nur Hafizah Azizan. The data obtained is summarised in **Table 4.3.** The oxidation state of zinc was not verified

but zinc is believed to be stable in its divalent state. MPH enzyme expressed and purified as it is, was found to contain 1.07 Zn^{2+} per MPH monomer while the enzyme was found to contain 1.38 Zn^{2+} per MPH monomer after overnight incubation with 0.1 mM $ZnCl_2$. One MPH monomer is expected to contain two metal ions.

Table 4.3: The zinc metal content of MPH enzyme as purified and MPH enzyme with metal pre-incubation.
MPH was produced in BL21(DE3)$^{RecA-}$harbouring pETMCSI_M1T_WT.

Sample	Zn (mol/MPH)
MPH as purified (without signal sequence)	1.07
MPH pre-incubated with 0.1 mM $ZnCl_2$	1.38

4.3.2.5. Effect of buffer additives on WT MPH

Another aspect that was investigated as an attempt to explain the results obtained in **section 4.3.2.2** was the effect of buffer additives on the activity of WT MPH. As mentioned earlier, the initial assay condition adapted from the literature contained Tween 80 to solubilise OP substrates.[16] OP substrates are hydrophobic molecules that have low solubility in aqueous solution and require additives to remain soluble at high concentrations. This is particularly important in kinetic characterisation where very high concentrations of substrates are used. However, sometimes these additives might have detrimental effects on the activity of the enzyme. To investigate how these buffer additives affect the activity, MPH was incubated for 30 minutes with 0.1% to 1% of additive. 0.1% or 1% of the additive was then included in the assay buffer and the assays were performed as described in **section 4.3.1.2.** The assay recipe contained 67 μM purified WT MPH and 119 μM methyl chlorpyrifos oxon in 50 mM HEPES, 100 mM NaCl, 0.1 or 1% additives, 0.1 mM $ZnCl_2$, pH 7.6. The additives investigated were mainly organic solvents, such as ethanol, methanol, acetone and acetonitrile, and some other additives that are commonly used together with protein, such as glycerol, ethylene glycol and Tween 20. The activities obtained were then normalised to that of WT MPH without additives and are shown in **Figure 4.12.** It was observed that most additives had a detrimental effect on MPH activity, with the exception of 2-methyl-2,4-pentanediol, DMSO, glycerol and PEG 8000 (all showed above 90% relative activity). These detrimental effects would be even more severe when the additives are used at their typical working concentrations of 10 – 20%.[17–20]

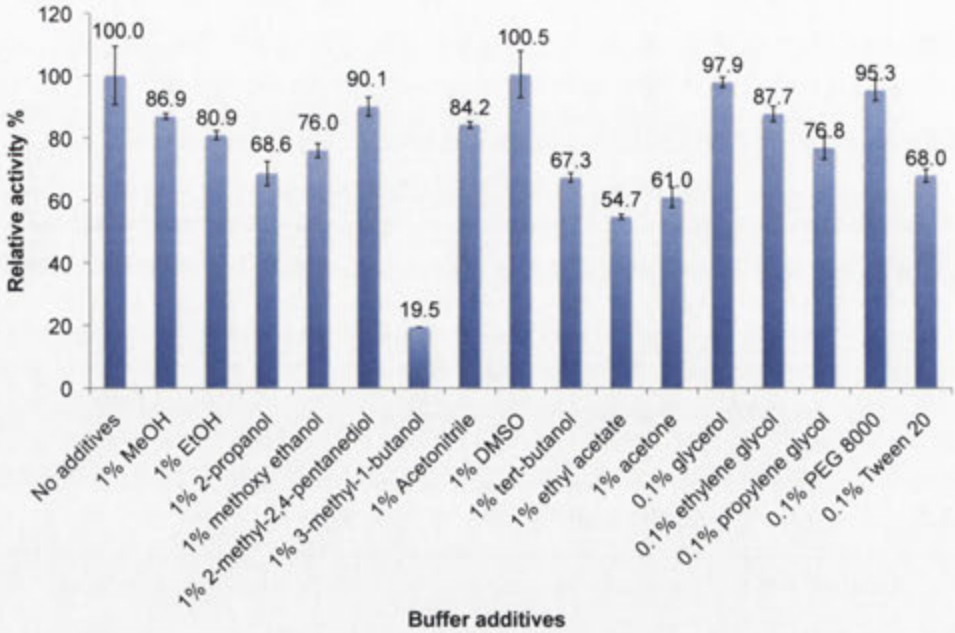


Figure 4.12: The effect of various additives on WT MPH activity.
1% was used for organic solvents while 0.1% was used for other additives. The error bars represent the standard deviations obtained for duplicate or triplicate measurements. The activities obtained for assay mixtures containing additives were normalised to the activity obtained without additives.

Methanol is one of the more commonly used cosolvent when dissolving OPs. Ely and co-workers found that at a final concentration of 10% (v/v) methanol, the activity of OPDA towards ethyl paroxon decreased 2.4-fold.[19] To further investigate the effect of methanol on the activity of MPH, the hydrolysis of methyl parathion was done in assay mixtures containing increasing concentrations of methanol. The assay mixtures contained 20 µL of 10x diluted cell lysate and 119 µM methyl parathion in 50 mM HEPES, 100 mM NaCl, 0 – 20% methanol, 0.1 mM $ZnCl_2$, pH 7.6. The cell lysate was pre-incubated with 90 µL of the buffer containing methanol for 30 minutes and another 90 µL of the buffer containing methanol and the substrate to initiate the reaction. This was done to ensure that the enzyme was adequately exposed to the solvent before activity measurements were taken. The relative activity obtained is shown in **Figure 4.13**. The MPH activity decreased, as the methanol concentration was increased. By 20% methanol, only 18.89% of activity was retained.

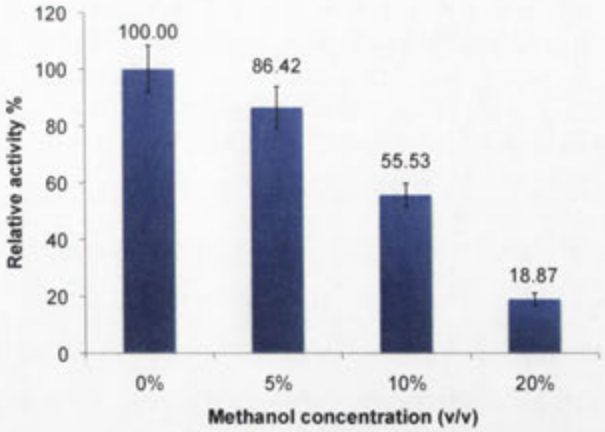


Figure 4.13: The effect of increasing methanol concentration on WT MPH activity.
The error bars represent the standard deviations obtained for 24-replicate measurements on a 96-well microtitre plate. The activities obtained for assay mixtures containing additives were normalised to the activity obtained without additives.

4.4. Discussion

4.4.1. Stability discrepancies between library screening and purified enzyme

As noted in the **section 4.3.2.2,** there were stability discrepancies between the stability values observed during screening and stability values observed during characterisation with purified R413D4 enzyme. One possibility for such observation was due to the different *E. coli* strains used for protein expression. *E. coli* DH5α was used during library screening while *E. coli* BL21(DE3)$^{RecA-}$ was used for large scale protein expression. It was suggested that different *E. coli* strains might have different metal uptake and regulation, leading to different metal or different metal content in the expressed enzyme.[20]

The possibility of molecular crowding was also considered due to reports of stability enhancement. Molecular crowding is usually used to compare the environment in the cell and the *in vitro* environment during activity assay with purified enzyme. Molecular crowding is found to enhance native state stability of proteins due to excluded volume effects.[21] A detailed review of the effects of molecular crowding on protein stability can be found in a review written by Despa *et al.*[22] One of the effects discussed in the review was thermostability enhancement due to molecular crowding. Egger and Valentine mimicked the effects of molecular crowding and confinement by sol-gel encapsulation of several proteins and observed thermostability enhancement.[23]

However, this possibility was ruled out, as it was unlikely that the environment of crude cell lysate assay would be similar to the environment described for molecular crowding. The more likely explanation would be the non-specific interactions introduced by the protein and other macromolecules present in the soluble lysate during library screening, affecting its thermostability. It was estimated that the cytoplasm of exponential or stationary phase *E. coli* contains 0.3 to 0.4 g/mL macromolecules[24], while the concentration of purified enzyme used in the assays was 2.11 to 4.22 μg/mL (67 to 134 nM). The difference between assay environments in library screening (soluble lysate) and enzyme characterisation (purified enzyme) due to the presence of host macromolecules might affect the behaviour of the enzyme. It is unclear at this point, which cellular component interacted with the MPH mutant or if the mutations in R413D4 promoted interactions of the enzyme with host macromolecules since such stability discrepancy was not observed for WT MPH.

4.4.2. Implications of MPH zinc composition

Metal content variations in metalloenzymes are not uncommon and indeed various examples of metal content variation in metallo-enzymes were found in a literature search. Dr. Jee-Loon Foo, in his PhD work, found a correlation between cobalt content and hydrolytic activity when OPDA was expressed and purified in media and buffer supplemented with $CoCl_2$.[20] Dr Foo also found that OPDA expressed in *E. coli* DH5α contain a lesser proportion of Co^{2+} compared to OPDA expressed in BL21(DE3)$^{RecA-}$, despite Co^{2+} supplementation in the expression and purification in both cases, suggesting different metal regulatory mechanisms in different *E. coli* strains. Dr. Foo also noted that none of the OPDA enzymes expressed without supplementing any metal had full metal occupancy. Omburo *et al.* found that supplementing growth medium with divalent ions improves OPDA expression in *E. coli* and OPDA purified in buffers containing divalent ions increases the proportion of the supplemented ion in OPDA. Limphong and colleagues characterised human glyoxalase II enzymes that were expressed in rich medium and minimal medium containing Zn^{2+}, Fe^{2+} and Co^{2+}.[25] They found that glyoxalase II expressed in LB has a partially occupied metal centre and that the metal content of the enzyme can be influenced by the metal included in minimal media. Human glyoxalase II is a metallo-β-lactamase superfamily enzyme that contain an active site metal centre that is identical to MPH.[26]

While MPH expressed in BL21(DE3)$^{RecA-}$ with LB media contained 1.07 equivalence of Zn^{2+} and incubation of the enzyme with Zn^{2+} post-purification did not increase the Zn^{2+} content to two equivalence, it is possible that a population of the metal binding site was occupied by other divalent metals due to depletion of Zn^{2+} during overexpression, preventing the formation of a bi-zinc metal centre during post-purification metal incubation. This phenomenon was observed in work done by Dr. Jee-Loon Foo, where OPDA expressed in BL21(DE3)$^{RecA-}$ with LB media contained 0.631 and 0.319 equivalences of Fe^{2+} and Zn^{2+} with no traces of Co^{2+} detected. When the same enzyme was expressed in BL21(DE3)$^{RecA-}$ with LB media supplemented with 1 mM $CoCl_2$, the enzyme was found to contain 1.855 equivalence of Co^{2+}. If the enzyme already contained 0.631 and 0.391 equivalences of Fe^{2+} and Zn^{2+}, it is unclear if Co^{2+} will be able to displace these metals to obtain a full Co^{2+} occupied metal centre if the enzyme is incubated with Co^{2+} post-purification. OPDA is a binuclear metalloenzyme. Another possibility for the failure to obtain two equivalence of Zn^{2+} would be due to the different binding affinities of the metal sites. If one of the metal binding sites has a weaker affinity towards Zn^{2+}, the metal occupying said site might be lost when passing the enzyme through HiTrap Desalting column. It was suggested that the more solvent exposed metal (β-metal) binds less tightly than the more buried metal (α-metal), resulting in the displacement of Zn^{2+} by Cd^{2+} at the β-metal position during protein crystallisation, as the crystallisation solution contained 0.1 M $CdCl_2$.[27] Therefore, the reason for activity variation of purified MPH discussed in section 4.5.1 is most likely due to the metal content variation of the expressed MPH since the MPH enzymes prepared were not expressed and purified in media and buffers supplemented with Zn^{2+} metals.

The reason for the metal content variation in MPH was not clear-cut and more studies are required to investigate the metal binding properties of MPH. The metal binding properties of MPH were not previously explored in the various MPH publications that have appeared to date.[16,28,29] The metal binding studies of MPH is currently being undertaken by Ms. Nur Hafizah Azizan, using site-directed mutagenesis on the first and second shell metal coordinating ligands. Foo *et al.* demonstrated that mutations of outer-shell metal coordinating ligands significantly weaken the metal coordinating strength of OPDA, providing evidence for metal binding affinity modulation by outer-shell residues.[30] Campos-Bermudez and co-workers attributed the

metal promiscuity of *Salmonella typhimurium* glyoxalase II to the presence of an aspartate residue in the second shell ligands, increasing the hardness of a primary shell coordinating histidine.[31]

Other than having catalytic roles, metal ions are shown to have structural roles in enzymes. Yeast alcohol dehydrogenase contains two zinc ions, one catalytic and one structural whereby removal of the structural zinc does not affect the activity, but compromises the stability of the protein.[32] Rochu and co-workers investigated the effects of different catalytic metal ions (Zn^{2+}, Co^{2+} and Cd^{2+}) on the thermostability of OPH and found Zn^{2+}-OPH to be the most thermostable while Cd^{2+}-OPH to be the least thermostable.[33] They also noted in their study that the stability of OPH is correlated to the substrate affinity of OPH with the most stable Zn^{2+}-OPH displaying the lowest K_m and Cd^{2+}-OPH displaying the highest K_m. Epting *et al.* compared the effects of divalent ions on the thermostability and thermal inactivation of mesophilic, thermophilic and hypothermophilic class II xylose-isomerases, found evidence suggesting *Thermotoga neapolitana* xylose isomerase requires one metal for stability and two metals for activity.[34] *Thermotoga neapolitana* xylose isomerase is a bi-nuclear metalloenzyme with a TIM-barrel fold (PDB ID: 1A0E). While MPH contains a binuclear metal centre at its active site, it is unclear if the enzyme requires two Zn^{2+} for activity or if the metals have any roles in maintaining stability. However, the difference in melting curves of MPH and Zn^{2+} pre-incubated MPH shown in **Figure 4.11** suggests that the metals do provide stability benefits to the enzyme.

The stability of MPH and Zn^{2+} pre-incubated MPH at higher temperatures (above 80°C) also showed unexpected but interesting observations **(Figure 4.11)**. Zn^{2+} pre-incubated MPH was fully inactivated at temperatures higher than 80°C while MPH without additional metals retained low levels of activities at the same temperatures. This was observed despite MPH that was pre-incubated with Zn^{2+} showed higher thermostability values at temperatures below 70°C. In other words, the Zn^{2+} pre-incubated WT MPH enzyme exhibited better thermostability and a thermostability profile that is expected from a typical enzyme, while WT MPH enzyme without Zn^{2+} supplementation showed anomalous thermostability profile. The activation characteristic observed from WT MPH without Zn^{2+} supplementation was not observed in the thermostability profile for Zn^{2+} pre-incubated MPH suggesting the possibility that the activation phenomenon was an artefact introduced by the incomplete metal centre.

Work done by several groups to engineer more thermostable MPH enzymes also showed low-level activity retention at higher temperatures, indicating the possibility that the enzymes expressed had lower than expected metal content since no metals were supplemented during expression, purification and activity measurement of the enzymes.[35–37] It is also not clear at this point, if WT MPH without a completely occupied metal centre follows a different unfolding kinetics from WT MPH with a completely occupied metal centre.

While far from conclusive, it is plausible that the thermostability evolution described in this chapter conferred a higher stability to the "metal-free" enzyme. The stability value of R413D4 without metal supplementation was indeed higher than MPH without metal supplementation when heat-treated at 70°C and 80°C **(Figure 4.10),** suggesting a more stable "metal-free" enzyme. The catalytic activity of R413D4 increased when the enzyme was pre-incubated with $ZnCl_2$ suggesting that the mutant too was expressed without a fully occupied metal centre. Roodveldt and Tawfik inadvertently stabilised the metal-free state of OPH when evolving the enzyme towards a promiscuous substrate.[38] The result is an OPH variant with a higher functional expression of OPH.

4.4.3. Structural implications of mutations

During screening, the best improvement was seen when L256Q and K316E were present in a variant, suggesting cooperativity between the two mutations. This was further supported by the much lower stability improvement seen in L256Q and K316E single mutants (data not shown). However, the two residues are not located near each other as shown in **Figure 4.14**.

The mutations were modelled onto the b-subunit of WT MPH structure (PDB ID: 1P9E-B) using SWISS-MODEL homology modelling server.[39–41] L256Q is located near the binuclear metal centre. D255 is the metal bridging aspartate residue. The L256Q mutation could introduce a hydrogen bonding between the side chain of Q256 and the backbone of E283, possibly increasing the rigidity of the overall structure **(Figure 4.15).** E283 is located on a helix strand, next to the loop where L256Q is located.

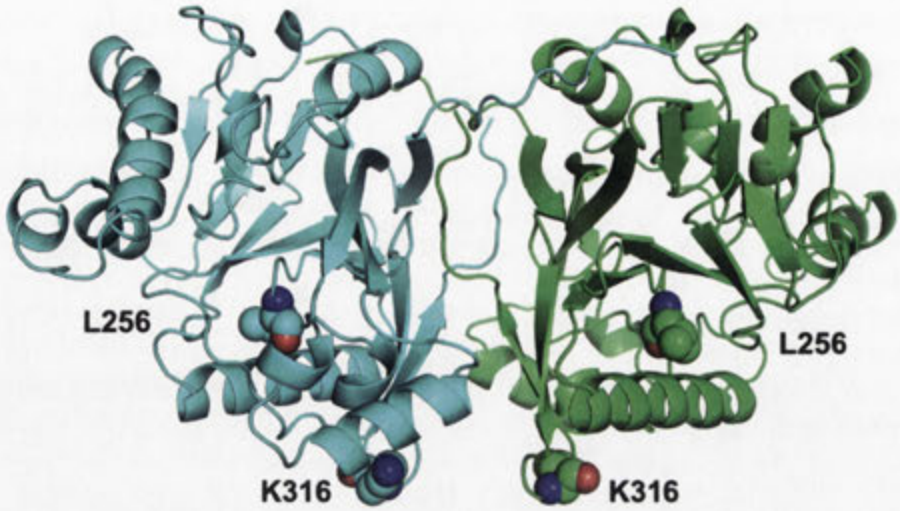


Figure 4.14: The positions of L256 and K316 in the dimeric MPH structure.

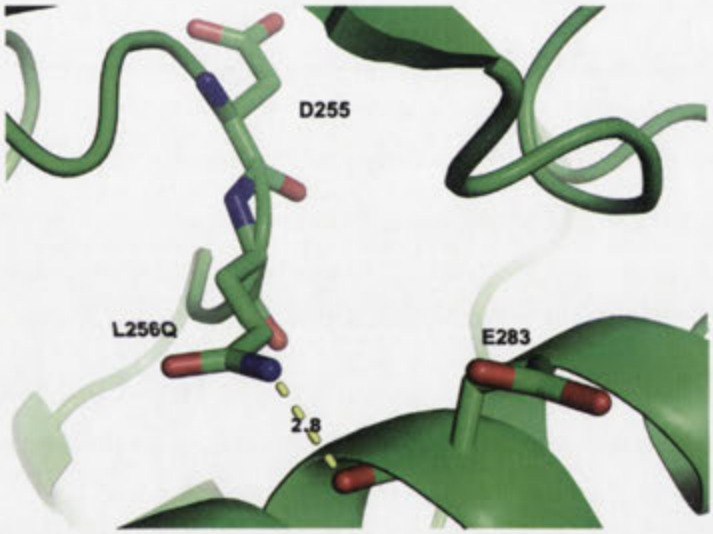


Figure 4.15: The hydrogen bonding between L256Q and E283.
The dashed line and the number show the hydrogen bond and the distance between the side chains of L256Q and E283.

K316 is a surface residue and the effect of K316E is unclear at this point. The mutation from lysine, a positively charged residue, to glutamate, a negatively charged residue resulted in an isoelectrical point (pI) shift due to the large difference in the pK_a values of lysine and glutamate. The calculated pI decreased from 7.87 to 6.71. The K316 residue in WT MPH does not seem to form interaction with any residues **(Figure 4.16a)**. From the modelled mutant structure, the side chain of K316E is located near the side chain of D265, potentially introducing repulsion (**Figure 4.16b**). This might affect

the interaction introduced by L256Q since D265 and Q256 are located on the same strand (**Figure 4.16c**).

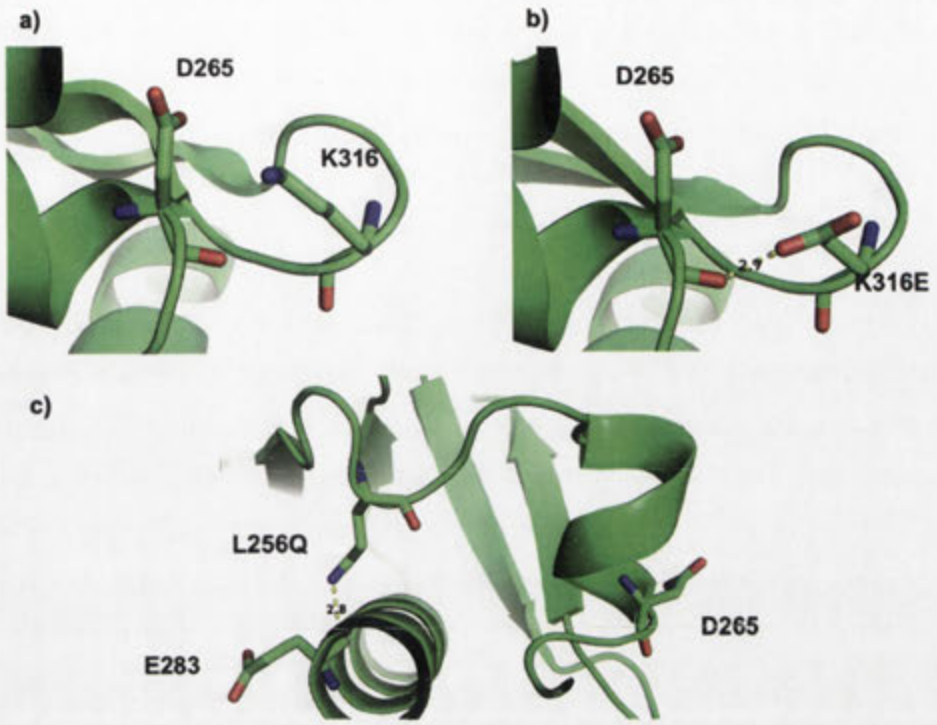


Figure 4.16: (a) The location of K316 in WT MPH. (b) The possible repulsion introduced by K316E mutation and (c) L256Q and D265 are located on the same strand.
The dashed line and number in (b) shows the distance between K316E side chain and D265 backbone. The dashed line and number in (c) shows hydrogen bonding between L256Q and E283.

4.4.4. Implications of MPH susceptibility to buffer additives

The investigation on the effects various buffer additives have on MPH revealed the susceptibility of the enzyme toward them. Due to the very low solubility of OP compounds in aqueous solutions, additives are required to solubilise these compounds when used in high concentrations. It is also undesirable for MPH to be negatively affected by various additives, since the enzyme is considered for deployment in OP bioremediation applications. MPH activity decreased by 1.8 and 5.3 folds at 10% and 20% methanol respectively. This presented us with another aspect of MPH that required improvement. Directed evolution can be applied on MPH to generate variants that can better tolerate organic solvents, thus increasing the overall stability of the enzyme.

4.4.5. Ideas to improve the thermostability evolution and future directed evolution of MPH

From the various observations noted during the library screening and enzyme characterisation experiments, several ideas to improve the thermostability and other future directed evolution experiments on MPH were conceived.

From the metal content of WT MPH without metal supplementation and other reports on metalloenzymes in the literature, it is obvious that overexpression of metalloenzymes in *E. coli* strains requires metal supplementation during expression and purification to obtain enzymes with fully occupied metal centre. The results obtained in MPH characterisation implied that the active site metal is important both catalytically and structurally. The evolution experiment performed in the absence of additional metals seemed to favour the stabilisation of the metal free enzyme, following a different pathway in enzyme stabilisation. Therefore, it is crucial in future thermostability and other directed evolution experiments of MPH that Zn^{2+} ion are supplemented to the growth media and assays buffers to maintain a fully occupied metal centre. The purification buffers should also be supplemented with Zn^{2+} ions for the same reason. This will ensure that the evolution will not select for metal free enzymes that are more stable.

In this thermostability evolution experiment, only two mutations were found to be responsible for thermostability enhancement. The difficulty in obtaining new mutations that confer substantial thermostability improvement might suggest that the pathway to attain stability enhancement was restricted by the criteria imposed during processing of selected enhanced variants. Giver and co-workers, in their work in evolving a thermostable esterase, found that the evolutionary pathway to better thermostability was restricted by selecting only highly active mutants at ambient temperature as parental template for subsequent evolution.[12] They suggested that by relaxing the requirement for high activity at ambient temperature, more evolution pathways to thermostability would be available and more mutants with upward shift in optimum temperature would be found. Gumulya and Reetz in evolving epoxide hydrolase, found that the best variants were obtained from parents consisting of neutral or inferior mutants and suggested using parents consisting of unimproved mutants as a technique to escape from local minima ("dead ends") in evolution.[42] In this experiment,

mutants with less than 60% WT MPH initial activity were removed from consideration. Relaxing this requirement in future thermostability evolution experiments might open up more evolutionary pathways and allow the selection of mutants with upward shift in activity profile.

Another possible improvement to the thermostability evolution experiment would be to adopt a high throughput activity-independent biophysical screen. One such method is the Hot-CoFi agar screening method developed by Asial and co-workers.[43] The method allows for the isolation of better performing mutants at the colony level. This method worked on the premise that only soluble enzymes will be able to pass though a filter membrane and insoluble aggregate would be eliminated. By exposing the colonies to a temperature range, a varying degree of enzyme aggregation can be obtained to construct a biophysical thermostabilty profile. A secondary activity screen can then be performed on isolated mutant to ensure that the mutants are still active and not dead because of mutations that confer enhanced stability at the expense of activity.

4.5. Summary

This chapter described the directed evolution of MPH for better thermostability. Four rounds of directed evolution and variants with enhanced thermostability were selected. A single variant, R413D4, containing L256Q and K316E double mutations gave the most thermostability improvement. Unexpectedly, the stability of the purified R413D4 was much lower than anticipated. Although there was an enhancement, the mutant showed only two-fold stability improvement, unlike the marked improvement seen during screening. Various possibilities for the issue were discussed and the most likely explanation is the increase of non-specific interaction between R413D4 and some cellular component introduced by the mutations, thus stabilising the enzyme. During the course of investigating the reason behind the discrepancy, it was also discovered that MPH is expressed in *E. coli* without additional metal had a lower than expected Zn^{2+} content, leading to compromised stability. Therefore, it is very plausible that we had generated a mutant that was more stable in its "metal free" form and stability enhancement seen in this work was metal independent. The mutant, expressed and purified without metal supplementation was more stable than the WT variant, expressed and purified without metal supplementation.

Investigation on the effect of additives on MPH had exposed its susceptibility toward organic solvents. This revelation is important, as this can potentially be a major weakness of the enzyme when deployed in the field for bioremediation. Either protein engineering or protein entrapment methods can be used to improve and protect the protein against solvent-induced denaturation.

The experiment was terminated at this point due to time constraint. However, suggestions on how the experiment could be improved had been given and discussed. The uncovering of the metal requirement during growth and expression subsequently contributed to a successful evolution experiment in altering the substrate specificity of MPH to various organophosphates (see **Chapter 5**).

4.6. References

1. Singh, B. K. Organophosphorus-degrading bacteria: ecology and industrial applications. *Nature Reviews Microbiology* **7,** 156–164 (2009).
2. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. & Engel, P. C. Protein thermostability in extremophiles. *Biochimie* **80,** 933–941 (1998).
3. Yano, J. K. & Poulos, T. L. New understandings of thermostable and peizostable enzymes. *Current Opinion in Biotechnology* **14,** 360–365 (2003).
4. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 5869–5874 (2006).
5. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Current Opinion in Structural biology* **19,** 596–604 (2009).
6. Bloom, J. D. *et al.* Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 606–611 (2005).
7. Wagner, A. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences* **275,** 91–100 (2008).
8. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Computational Biology* **4,** e1000002 (2008).
9. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444,** 929–932 (2006).
10. Salazar, O. & Sun, L. in *Methods in Molecular Biology. Volume 230: Directed Enzyme Evolution: Screening and Selection Methods* (Arnold, F. H. & Georgiou, G.) 85–97 (Human Press Inc).
11. Cirino, P. C. & Georgescu, R. in *Methods in Molecular Biology. Volume 230: Directed Enzyme Evolution: Screening and Selection Methods* (Arnold, F. H. & Georgiou, G.) 117–125 (Human Press Inc).
12. Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 12809–12813 (1998).
13. Stevenson, B. Directed evolution of pyruvate decarboxylase for *in vitro* glycolysis. (2006).
14. Kobayashi, C. *et al.* Thermal conversion from low- to high-activity forms of catalase I from *Bacillus stearothermophilus*. *Biochemistry* **272,** 23011–23016 (1997).
15. Yamamoto, T. *et al. In vitro* heat effect on functional and conformational changes of cyclodextrin glucanotransferase from hyperthermophilic archaea. *Biochemical and Biophysical Research Communications* **265,** 57–61 (1999).
16. Fu, G., Cui, Z., Huang, T. & Li, S. Expression, purification, and characterization of a novel methyl parathion hydrolase. *Protein Expression and Purification* **36,** 170–176 (2004).
17. Yang, H. *et al.* Evolution of an organophosphate-degrading enzyme: a comparison of natural and directed evolution. *Protein Engineering* **16,** 135–145 (2003).
18. Jackson, C. J. *et al.* Anomalous scattering analysis of Agrobacterium radiobacter phosphotriesterase: the prominent role of iron in the heterobinuclear active site. *Biochemical Journal* **397,** 501–508 (2006).

19. Ely, F. *et al.* The organophosphate-degrading enzyme from *Agrobacterium radiobacter* displays mechanistic flexibility for catalysis. *The Biochemical Journal* **432,** 565–573 (2010).
20. Foo, J.-L. Mechanistic study of enzymes involved in phosphate ester hydrolysis. (2009).
21. Cheung, M. S., Klimov, D. & Thirumalai, D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 4753–4758 (2005).
22. Despa, F., Orgill, D. P. & Lee, R. C. Molecular crowding effects on protein stability. *Annals of the New York Academy of Sciences* **1066,** 54–66 (2005).
23. Eggers, D. K. & Valentine, J. S. Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Science* **10,** 250–261 (2001).
24. Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli. Journal of Molecular Biology* **222,** 599–620 (1991).
25. Limphong, P. *et al.* Human glyoxalase II contains an Fe(II)Zn(II) center but is active as a mononuclear Zn(II) enzyme. *Biochemistry* **48,** 5426–5434 (2009).
26. Cameron, a D., Ridderström, M., Olin, B. & Mannervik, B. Crystal structure of human glyoxalase II and its complex with a glutathione thiolester substrate analogue. *Structure* **7,** 1067–1078 (1999).
27. Dong, Y.-J. *et al.* Crystal structure of methyl parathion hydrolase from *Pseudomonas sp.* WBC-3. *Journal of Molecular Biology* **353,** 655–663 (2005).
28. Ekkhunnatham, A., Jongsareejit, B., Yamkunthong, W. & Wichitwechkarn, J. Purification and characterization of methyl parathion hydrolase from *Burkholderia cepacia* capable of degrading organophosphate insecticides. *World Journal of Microbiology & Biotechnology* **28,** 1739–1746 (2012).
29. Yang, J., Yang, C., Jiang, H. & Qiao, C. Overexpression of methyl parathion hydrolase and its application in detoxification of organophosphates. *Biodegradation* **19,** 831–839 (2008).
30. Foo, J.-L. *et al.* Mutation of outer-shell residues modulates metal ion co-ordination strength in a metalloenzyme. *The Biochemical Journal* **429,** 313–321 (2010).
31. Campos-Bermudez, V. a *et al.* Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: new insights into metal ion selectivity. *Biochemistry* **46,** 11069–11079 (2007).
32. Magonet, E., Hayen, P., Delforge, D., Delaive, E. & Remacle, J. Importance of the structural zinc atom for the stability of yeast alcohol dehydrogenase. *The Biochemical Journal* **287,** 361–365 (1992).
33. Rochu, D. *et al.* Contribution of the active-site metal cation to the catalytic activity and to the conformational stability of phosphotriesterase: temperature- and pH-dependence. *The Biochemical Journal* **380,** 627–633 (2004).
34. Epting, K. L., Vieille, C., Zeikus, J. G. & Kelly, R. M. Influence of divalent cations on the structural thermostability and thermal inactivation kinetics of class II xylose isomerases. *The FEBS Journal* **272,** 1454–1464 (2005).
35. Tian, J. *et al.* Enhanced thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 by rational engineering of a glycine to proline mutation. *The FEBS Journal* **277,** 4901–4908 (2010).
36. Su, Y. *et al.* Improving the thermostability of a methyl parathion hydrolase by adding the ionic bond on protein surface. *Applied Biochemistry and Biotechnology* **165,** 989–997 (2011).

37. Tian, J. *et al.* Improving the thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 using a computationally aided method. *Applied Microbiology and Biotechnology* **97,** 2997–3006 (2013).
38. Roodveldt, C. & Tawfik, D. S. Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *Protein Engineering, Design & Selection* **18,** 51–58 (2005).
39. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* **31,** 3381–3385 (2003).
40. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).
41. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).
42. Gumulya, Y. & Reetz, M. T. Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *Chembiochem* **12,** 2502–2510 (2011).
43. Asial, I. *et al.* Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nature Communications* **4,** 2901 (2013).

ALTERING THE SUBSTRATE SPECIFICITY OF METHYL PARATHION HYDROLASE

5. Altering the Substrate Specificity of Methyl Parathion Hydrolase

5.1. Preamble

As noted in Chapter 1, methyl parathion hydrolase (MPH) has a pronounced substrate preference for methyl parathion (MPS). Altering any of the phosphorous substituent of the substrate causes a significant drop in activity – this stands in stark contrast to the substrate specificities of OPH and OPDA that show a preference for some substrates, but still exhibit a significant level of activity for a broad range of substrate. MPH, like OPDA and OPH, is a binuclear metalloenzyme – the metal centre of MPH differs in detail to those observed in OPDA and OPH, but one would not expect these differences to result in a narrowing of substrate specificity. Furthermore, the overall structure of MPH is similar to a variety of other enzymes – MPH is part of a structural superfamily that has diverse range of substrates.[1,2] These latter observations suggest that it should be possible to alter the substrate specificity of MPH. The objective of the work described in this chapter was to alter the substrate specificity of MPH, to examine the activity of selected mutants against various substrates and to examine the mechanistic implications of the resulting mutations that resulted in the altered specificity. The overall workflow of the work described in this chapter is given in **Figure 5.1.**

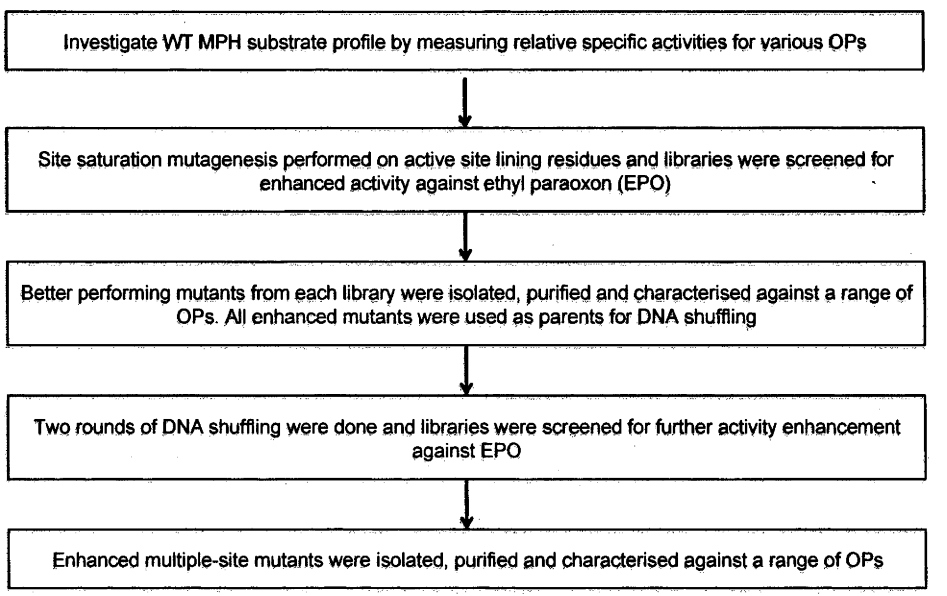


Figure 5.1: Workflow overview in altering MPH substrate specificity.

5.2. Approach – site saturation mutagenesis of active site residues

There were several approaches that could have been taken to alter the substrate specificity of MPH. Rational approaches require a detailed knowledge of the protein and how it interacts with the substrate – this is usually obtained from an examination on the structure of the protein with a bound substrate mimic or an inhibitor. The structure of MPH was known[3], but the structure of a suitable complex was not known at the time that this work commenced – so rational approaches were not feasible. Model complexes could have been made, but these were not deemed accurate enough to allow the redesign of MPH active site.

Other approaches to altering the substrate specificity of MPH were considered. Evolutionary approach involving error prone PCR (ePCR) of the MPH gene was one of the other approaches considered. However, these approaches, as with evolutionary approach, did not concentrate mutagenesis on the active site residues of the enzyme – the residues most likely to be responsible for substrate specificity. Furthermore, high throughput screening for increased stability of MPH (as described in **Chapter 4**) proved to be subject to considerable experimental error. Given these considerations, it felt best to use site saturation of active site residues to alter the substrate specificity of MPH – these residues being the most likely to affect substrate specificity. Furthermore, this approach requires screening of relatively small libraries. Whole gene random mutagenesis can be applied to further optimise the enzyme once the active site has been optimised for a particular substrate.

5.3. Choosing substrate for primary and secondary screens

The aim of our experiments was to alter substrate specificity, however some consideration was given to the possible changes that could be brought about by site saturation mutagenesis. First, it was deemed an uphill task to convert MPH to a different type of enzyme – the possible substrates would be restricted to those that could undergo hydrolysis reaction. Secondly, it was thought that it would be possible to work with a substrate for which the native enzyme has a low level of activity. In other words, it was thought possible to enhance the activity of the enzyme for a specific substrate, but it was not thought feasible to generate activity with a completely new substrate.

Therefore, the activities of MPH towards various substrates were measured and the results are shown in **Figure 5.2**. The activity measurements were performed as described in **Section 2.7.1**.

From the various OP substrates investigated, MPH showed preference towards dimethyl-substituted substrates such as MPS, methyl paraoxon (MPO) and methyl chlorpyrifos oxon (MCO) over diethyl-substituted analogues such as ethyl parathion (EPS), EPO and ethyl chorpyrifos oxon (ECO). The enzyme also showed preference towards substrates having 3,5,6-trichloro-2-pyridinol (TCPy) leaving group compared to the *p*-nitrophenol (pNP) analogues (e.g. MPO vs MCO).

It was thought that it was reasonable to enhance the activity of MPH towards ethyl paraoxon (EPO) as the native enzyme displayed weak, but detectable activity towards this substrate – about 1% of the activity of the enzyme towards MPS. In addition, EPO possessed two features that were common to other poor substrates of MPH – it had ethyl substituents, as opposed to the methyl substituents of MPS. Furthermore, the phosphoryl group (P=O) of EPO was also found in other substrates for which MPH exhibited low activity. It was thought that an enzyme with enhanced activity towards EPO would also show increased activity towards a variety of other substrates.

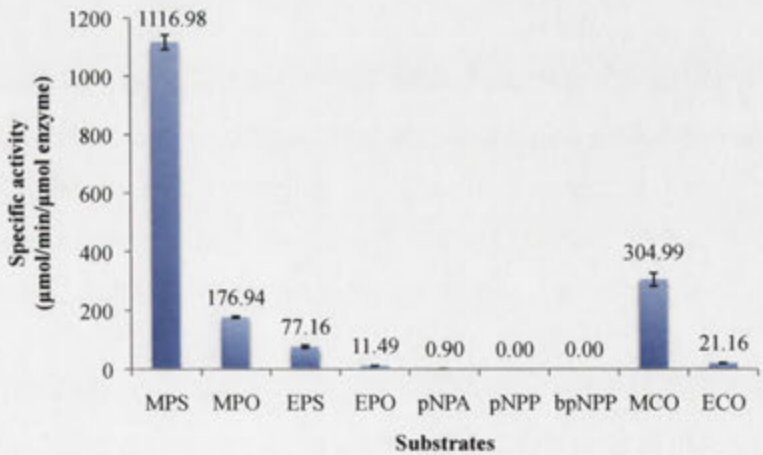


Figure 5.2: MPH activity towards various substrates.
The activities towards various substrates were measured with purified MPH. 119 µM of each substrate was assayed in 50 mM HEPES, 100 mM NaCl, 0.1 mM $ZnCl_2$. The Y-axis error bars represent the standard deviation between replicates in an experiment. The substrates investigated were as follows: methyl parathion (MPS), methyl paraoxon (MPO), ethyl parathion (EPS), ethyl paraoxon (EPO), *p*-nitrophenyl acetate (pNPA), *p*-nitrophenyl phosphate (pNPP), *bis*-(*p*-nitrophenyl) phosphate (bpNPP), methyl chlorpyrifos oxon (MCO) and ethyl chlorpyrifos oxon (ECO).

5.4. Choosing residues to mutate

There were many residues lining the active site – too many to mutate them all in the time permitted for thesis work. Residues to be mutated were selected based on sequence and structure comparisons with related enzymes – enzymes with similar structures and catalytic activities, but with different types of substrates. During the course of this work, differences were observed in the conformations of residues lining the active sites of the two crystallographic subunits. These differences and their implications on the catalytic mechanism of MPH were not mentioned in previous publications and will be described in more detail below. The residues that had different conformations in the two subunits were also marked for mutagenesis.

5.4.1. Sequence and structure comparison with related enzymes

The residues around the metal centre that form an essential section of the substrate-binding pocket were identified in the report of the crystal structure for MPH and the coordinates are available from the Protein Data Bank (PDB ID: 1P9E). The authors of the structure report noted that hydrophobic residues including Leu65, Leu67, Phe119, Trp179, Phe196, Leu258 and Leu273 line the putative substrate-binding pocket and investigated the roles of aromatic amino acid residues present in the substrate-binding pocket, namely Phe119, Trp179 and Phe196 by site directed mutagenesis.[3] They noted that converting any of these residues to alanine caused a reduction of catalytic activity towards MPS and attributed the reduction to the disruption of substrate binding. However, closer inspection of the provided kinetic values revealed that while W179A and F196A mutations resulted in K_m increase, F119A mutation did not show significant K_m increase. While the latter result implicated some residues in substrate binding, it did not give clues as to how the substrate specificity could be altered.

It has been noted that MPH is part of a structural superfamily – it is said to have a metallo-β-lactamase or αββα fold. Structures of numerous members of the superfamily are available from the Protein Data Bank (PDB - http://www.rcsb.org/pdb/home/home.do). Seven structures were downloaded for comparison – these included examples of B1, B2 and B3 β-lactamases for which a number of similar structures are available. PyMOL was used to visualize the structures and superimpose the coordinates onto subunit B of MPH that contains two Zn^{2+} ions.

The superimposition was done in two stages. First, PyMOL's Cealign command was used was for the superimposition. Cealign aligns protein structures based on structural similarities and is useful for comparing structurally similar proteins with low sequence similarity. Minor adjustments to the superimposition were then done with PyMOL's pair fitting function by using the two active site metals, the bridging water and the metal coordinating residues as points of alignment, and the RMSD for the alignments can be found in **Table 5.1**. It was noted that the metals and metal-binding ligands superimposed with good agreement – as is evident in **Figures 5.3**, **5.5** and **5.7**. It was also noted that the MPH active site was the most crowded of all the structures compared – all the other active sites were more open than that of MPH. This difference can be seen clearly in **Figures 5.4**, **5.6** and **5.8** in which MPH structure was displayed in cartoon mode with side chains displayed in lines, while the other protein structures were displayed in the surface representation mode. Identifying the residues that contribute to the active site confinement could be the key in altering the substrate preference for MPH. Examples of the individual structural comparisons are given below.

Table 5.1: List of MBL superfamily members used in superimposition study.
The table lists down the MBL superfamily enzyme used, their enzyme type, PDB ID, the organism the enzyme originated from, the ligands used for pair fitting and RMSD values of the pair fitting.

Protein name	Enzyme class	PDB ID	Organism of origin	Ligands used for pair fitting	RMSD (Å)
MPH	MBL superfamily phosphotriesterase	1P9E[3]	*Pseudomonas sp.* WBC-3	H147, H149, D151, H152, H234, D255 H302, Zn_1, Zn_2, bridging OH^-	N/A
AIM-1	B3 metallo-β-lactamase	4AWY[4]	*Pseudomonas aeruginosa*	H116, H118, D120, H121, H196, H263, Zn_1, Zn_2, bridging OH^-	0.387
BcII	B1 metallo-β-lactamase	1BVT[5]	*Bacillus cereus*	H86, H88, D90, H149, C168, H210, Zn_1, Zn_2	0.761
CphA	B2 metallo-β-lactamase	1X8G[6]	*Aeromonas hydrophila*	N116, H118, D120, H196, C221, H263, Zn_1	0.782
L1	B3 metallo-β-lactamase	1SML[7]	*Stenotrophomonas maltophilia*	H84, H86, D88, H89, H160, H225, Zn_1, Zn_2, bridging OH^-	0.393
Glx2	MBL superfamily glyoxlase	1QH5[8]	*Homo sapiens*	H54, H56, D58, H59, H110, D134, H173, Zn_1, Zn_2, bridging OH^-	0.262
GloB	MBL superfamily glyoxalase	2QED[9]	*Salmonella typhimurium*	H53, H55, D57, H58, H110, D127, His165, Fe_1, Fe_2, bridging OH^-	0.271
AHL lactonase	MBL superfamily lactonase	2BR6[10]	*Bacillus thuringiensis*	H104, H106, D108, H109, H169, D191, H235, Zn_1, Zn_2, bridging OH^-	0.454
ZiPD	MBL superfamily phosphodiesterase	2CBN[11]	*Escherichia coli*	H64, H66, D68, H69, H141, D212, H270, Zn_1, Zn_2,	0.337

5.4.1.1. MPH/AIM-1 superimposition

In general, AIM-1 has a more open active site when compared to MPH – it has extra helices and sheets as is evident in the superimposition shown in **Figure 5.4**. Both proteins have the same metal-binding ligands with the exception of Asp255 that bridges the two metal ions in MPH - Ser207 in AIM-1 that is equivalent to Asp255 in MPH **(Figure 5.3).**

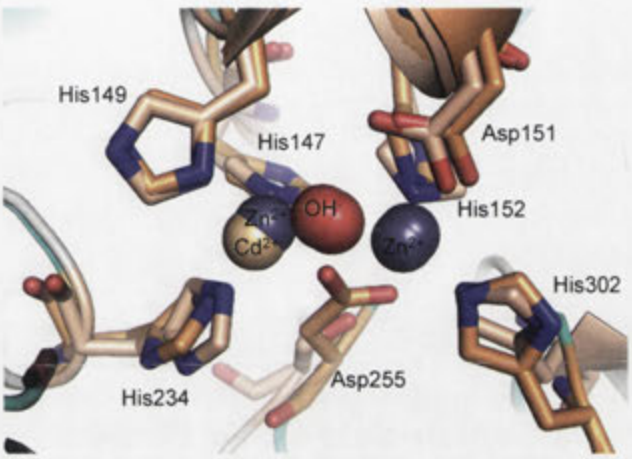


Figure 5.3: Structural superimposition of MPH and AIM-1 metal-binding ligands.
The amino acid numbering scheme from MPH was used in labelling the metal-binding ligands. The MPH ligands were coloured yellow while AIM-1 ligands were coloured wheat. The metal-binding ligands of the two enzymes are similar with the exception of Asp255, which its equivalent is replaced by a non-bridging serine in AIM-1.

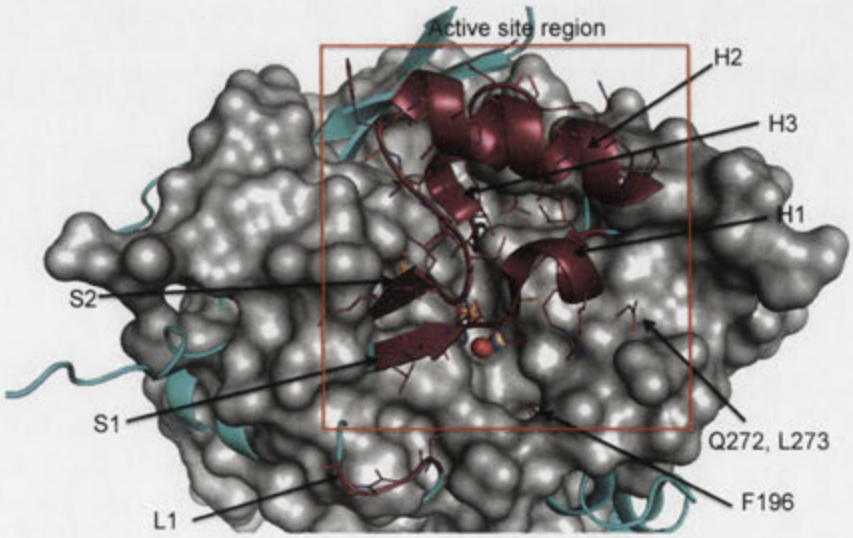


Figure 5.4: Structural superimposition of MPH and AIM-1.
Figure shows the structural superimposition of MPH and A1M-1, where MPH is coloured cyan and AIM-1 coloured grey in surface representation mode. The red box indicates the active site region. The maroon coloured elements are MPH elements that are not present in AIM-1. L: Loop; S: Sheet; H: Helix.

5.4.1.2. MPH/GloB superimposition

As is evident in the superimposition shown in **Figure 5.6**, GloB has a more open active site cavity. The residues Leu67, Pro68, Arg72, Phe119, Pro150, Phe196 and Leu258 of MPH protrude into what would be the active site of GloB. Both the human glyoxalase II (Glx2) and *Salmonella thyphimurium* glyoxalase II (GloB) share identical metal-binding ligands with MPH (**Figure 5.5**). However, when comparing the two glyoxalases, GloB has a bi-iron center whereas human Glx2 has a bi-zinc centre (figure not shown). MPH appears to function with two Zn^{2+} ions, but the protein was crystallised from a solution containing Cd^{2+} and this ion has displaced one of the Zn^{2+} ions in the crystalline protein – one subunit has a Cd^{2+}-Zn^{2+} centre while the other has two Zn^{2+} ions.

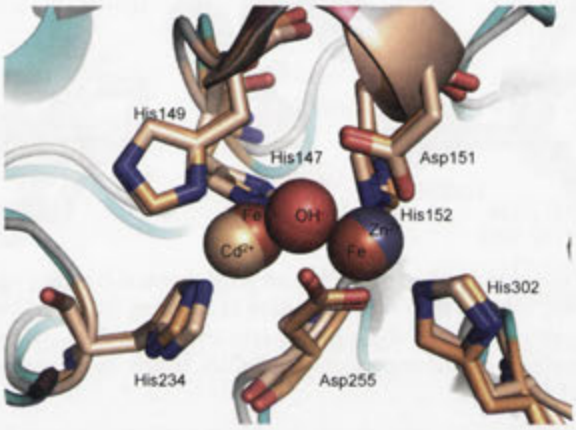


Figure 5.5: Structural superimposition of MPH and GloB metal-binding ligands.
The amino acid numbering scheme from MPH was used in labelling the metal-binding ligands. The MPH ligands are coloured yellow while GloB ligands are coloured wheat. The metals are Zn in MPH and Fe in GloB.

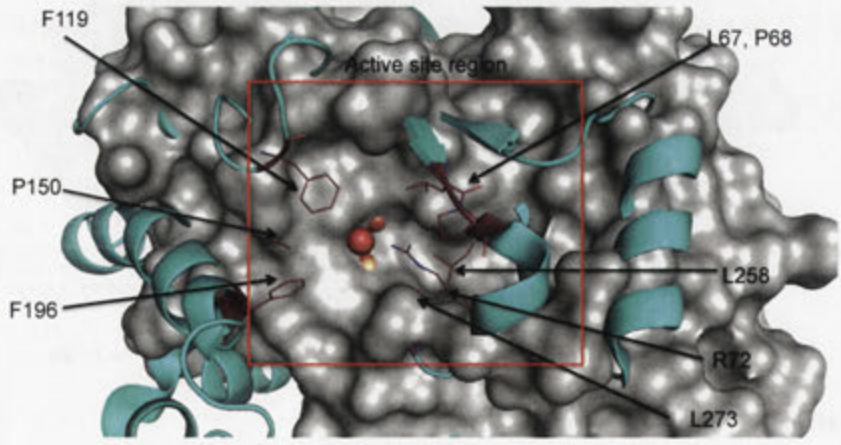


Figure 5.6: Structural superimposition of MPH and GloB.
Figure shows the structural superimposition of MPH and glyoxalase, where MPH is coloured cyan and GloB coloured grey in surface representation mode. The red box indicates the active site region. The MPH residues that line the active site are different from GloB and are coloured maroon.

5.4.1.3. MPH/ZiPD superimposition

Similar to GloB, ZiPD has a much more open active site compared to MPH. The MPH residues that protrude into the active site of GloB also protrude into the active site of ZiPD (**Figure 5.8**). The metal-binding ligands are the same in MPH and ZiPD (**Figure 5.7**). The physiological function ZiPD appears to be in tRNA processing, but its exact role is not clear.[11] ZiPD is also found to hydrolyse the phosphodiesters bpNPP and thymidine-5'-*p*-nitrophenyl phosphate.[12]

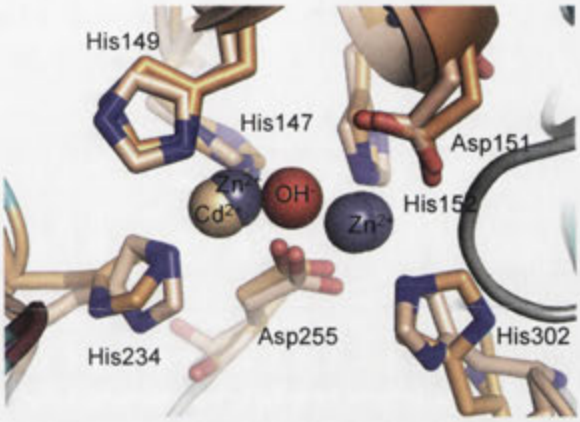


Figure 5.7: Structural superimposition of MPH and ZiPD metal-binding ligands.
The amino acid numbering scheme from MPH was used in labelling the metal-binding ligands. The MPH ligands were coloured yellow while ZiPD ligands were coloured wheat. The bridging hydroxide is not present in the ZiPD crystal structure.

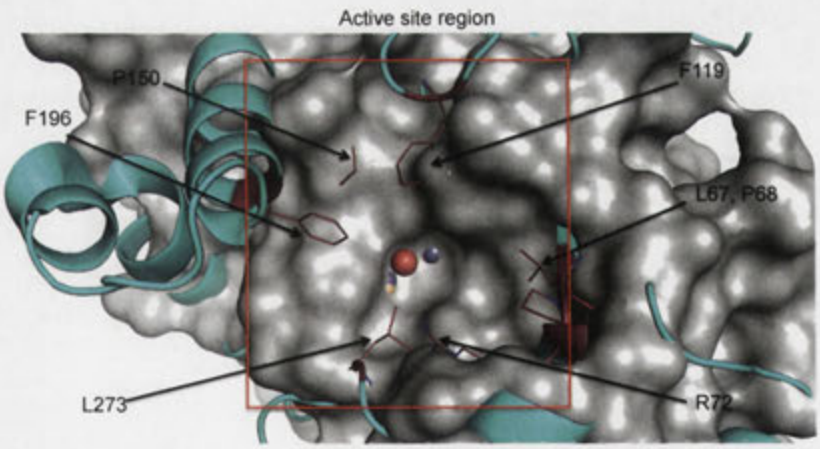


Figure 5.8: Structural superimposition of MPH and ZiPD.
Figure shows the structural superimposition of MPH and ZiPD where MPH is colored cyan and ZiPD colored grey in surface representation mode. The maroon colored fragments are the extra residues around MPH active site when compared to ZiPD. The red box indicates the active site region.

5.4.1.4. Summary of MPH superimpositions

As mentioned above, MPH was compared to seven other proteins in the β-lactamase superfamily. The results for AIM-1, GloB and ZiPD were very similar to each other and to the other members of the superfamily. A summary of the observations made from all the superimposition studies is given below in **Table 5.2**. From these comparisons, it was evident that much of the MPH active site overlapped with those of many other enzymes suggesting that these residues fulfilled a structural role and were not important for discriminating between substrates. The residues that did not appear to overlap with the active site residues of the other proteins were taken to be those that were important for determining the substrate specificity of MPH – these residues are marked with a red highlighter in **Table 5.2**.

The program ConSurf[13] was also used to identify residues that might be important for substrate binding based on sequence conservation. This program accesses protein sequence databases and finds the 150 most similar sequences to an input sequence – MPH in this case. Most of the sequences that were found were part of the β-lactamase superfamily and had a diverse range of substrates. Despite the differences in function, a number of conserved residues were found in all the proteins. For example, a number of metal-binding residues were found in all the sequences. ConSurf also quantifies the degree of conservation for each residue – a score of nine indicates very highly conserved residue while a score of one indicates a variable residue. With one exception, the residues identified in the previous section as being important for substrate binding had conservation scores that indicated that they were not highly conserved. The exception Leu258, although highly conserved, did show some variability – four substitutions had previously been observed. The residues listed in **Table 5.2** and highlighted in red were selected for site saturation mutagenesis.

Table 5.2: Summary of MPH superimposition with MBL superfamily enzymes.
Residues that contribute to MPH active site confinement compared to other structure are marked with asterisks and sites selected for SSM are shaded red. The ConSurf conservation scores for the selected sites are also listed in the table.

Residue	AIM-1	BcII	CphA	L1	GloB	Glx2	AHL lactonase	ZiPD	ConSurf rating
V65	*	*				*			
A66	*	*		*		*			
L67	*	*	*	*	*	*		*	5
P68	*	*	*	*	*	*		*	5
V69	*			*		*			
D70	*			*		*			
K71	*			*		*			
R72	*	*	*	*	*	*	*	*	6
L73	*	*		*		*			
N74		*		*					
Q75		*		*					
P76	*	*		*					
A77	*	*		*					
P78	*	*		*					
K79	*	*		*					
T80	*	*		*					
Q81	*	*		*					
S82	*	*		*					
A83	*	*		*					
L84	*	*		*					
A85	*	*		*					
K86	*	*		*					
S87	*	*		*					
F88	*	*		*					
Q89	*	*		*					
K90	*	*		*		*			
A91	*	*		*		*			
P92	*	*		*		*			
L93	*	*		*		*			
E94	*	*		*		*			
T95	*			*		*			
S96	*								
F119	*	*	*		*	*	*	*	5
G120	*								
P121	*								
P150					*	*		*	6
D190							*		
S193		*							
F196	*	*			*	*	*	*	3
I257	*								
L258	*	*	*	*	*	*			9
V259	*	*		*		*			
A260	*			*					
A261	*			*					
V262	*			*					
Q263	*			*		*			
D265	*								
D266	*								
Q272	*	*		*			*		
L273	*	*	*	*	*	*	*	*	7

5.5. Site saturation mutagenesis

Having decided which residues would be altered, there were still a number of issues to be considered before libraries could be generated. The bottleneck of directed evolution studies is the library screening throughput. First, if the residues were altered simultaneously, the number of possible variants was enormous. Consider 20 residues at each of eight sites would give 256 x 10^{10} variants - only a very small fraction of such a library could be screened. Simultaneously mutating three residues would produce about 8000 variants – this is feasible.

A second point to consider is that there are duplications when making random mutations. Therefore, a much larger library needs to be generated to compensate for duplications and to ensure that a reasonable fraction of the possible mutants are generated – this is referred to as oversampling. There are published formulae to calculate the number of mutants that should be generated to give a specified coverage of a randomly generated library consisting of a given number of possible variants. **Appendix H** provides a review of this literature and the formulae used in the calculation required to produce **Tables 5.3** and **5.4**. The program CASTer 2.0[14] was used to calculate the oversampling necessary for the libraries used in this work.

Some consideration should be given to the codons used to generate the mutant library – the third consideration. An amino acid is specified by a codon, consisted three nucleotides, generally represented as NNN where N=A/T/G/C. Normally, site saturation mutagenesis is done by specifying the amino acids with NNK codons – where K = G/T. All 20 amino acids are generated with the NNK specification and combinations of codons are required. Alternatively, Reetz and colleagues have suggested the use of an NDT library[14] – where D=AGT and T=T (Thiamine). This gives 12 codon combinations and 12 amino acids (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly) – which Reetz describes as a balanced mix of polar and non-polar, aliphatic and aromatic, and negatively and positively charged representatives while excluding mostly structurally similar amino acid residues. **Table 5.3** lists the library size required to give 95% coverage of the NNK and NDT libraries for different numbers of mutation sites.

Table 5.3: Oversampling required in achieving 95% coverage for NNK & NDT.

No of sites	NNK		NDT	
	Codons	Transfromants needed	Codons	Transformants needed
1	32	94	12	34
2	1,024	3,066	144	430
3	32,768	98,163	1,728	5,175
4	1,048,576	3,141,251	20,736	62,188
5	33,554,432	100,520,093	248,832	745,433

For the work described in this chapter, the gene for WT MPH was cloned into pJWL1030 vector without the signal sequence (designated pJWL1030_M1T_WT). MPH was amplified from pCY76_M1T_WT with F-M1T-pJWL1030flank_Ase and R-M1T-pJWL1030flank_Pst primers and cloned into *Nde*I and *Pst*I sites in pJWL1030. The construction of pCY76_M1T_WT can be found in **Chapter 4.** Site saturation mutagenesis (SSM) was performed on sites listed in **Table 5.2** using the megaprimed ligase-free SSM method on pJWL1030_M1T_WT, as described in **section 2.5.4.3.**

5.5.1. Library size

It was decided that the amino acids Leu67, Pro68 and Arg72 would be mutated simultaneously using NDT randomization codons. It was thought that Arg72 and Trp179 could be important for activity so they were mutated separately with NNK randomisation. Other amino acids were mutated with NDT libraries as shown in **Table 5.4** – this table also gives the number of variants that must be screened to obtain at least 95% of the theoretically possible library and the number of variants screened, as described in the next section. Any mutants displaying increased activity would be shuffled with the aim of producing recombined mutants with further enhanced activity.

5.5.2. Library screening with ethyl paraoxon

As described in **section 2.5.6**, primary and secondary screenings involved the use of 96-well microtitre plates with the reaction monitored by measuring increased absorption due to the formation of *p*-nitrophenolate ion. 10 μL of cells were used in the activity screens. The results of these screens are shown in **Table 5.5**. A small number of mutants with deteriorated activity were also sampled to examine the mutations that caused the decrease in activity. Only changes to Phe119, Phe196, Leu258 and Leu273 gave enhanced activity.

Table 5.4: Summary of libraries generated, randomisation used, and number of mutants screened.
*The number of possible variants and oversampling required for 95% coverage was calculated using CASTer 2.0.

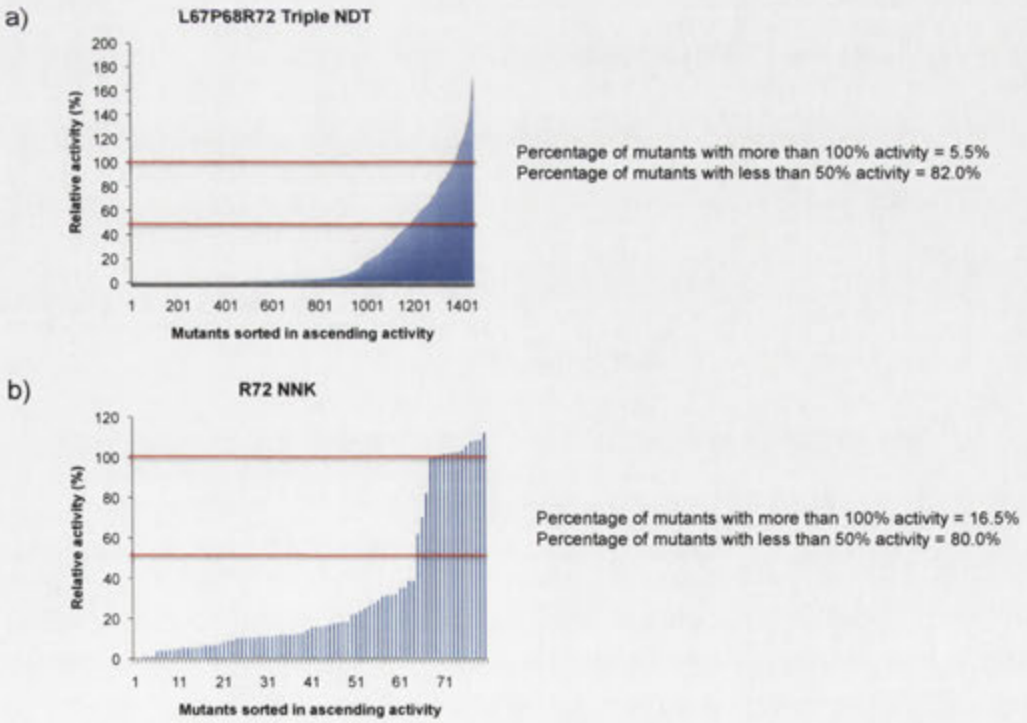
SSM Library	Codon randomisation	Number of possible variants*	Library size to achieve 95% coverage*	Number of mutants screened
L67P68R72	NDT-NDT-NDT	1278	5175	5800
R72	NNK	32	94	80
F119	NDT	12	34	40
P150	NDT	12	34	40
W179	NNK	32	94	80
F196	NDT	12	34	40
L258	NDT	12	34	40
L273	NDT	12	34	40

Table 5.5: Mutants isolated from the SSM libraries.
Table shows the mutants with improved activity and mutants with activity loss for all the libraries screened. The numbers in bracket indicate the relative activity of the mutant to WT MPH on EPO.

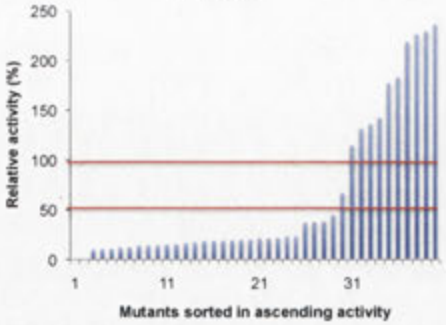
Library	Improved mutant	Deteriorating mutation
L67P68R72 Triple NDT	None	Did not sample
R72 NNK	None	A (~36%), T (~14%), G (~11%)
F119 NDT	Y (~188%)	G (~19%), N (~18%)
P150 NDT	None	Y (~27%), H (~8%), G (~65%)
W179 NNK	None	S (~4%), H (~10%), Y (~17%), M (~20%)
F196 NDT	I (~142%), L (~151%), Y (113%)	R (~15%), D (~9%)
L258 NDT	I (~124%), S (~198%), N (198%), H (~479%)	R (5%)
L273 NDT	I (~158%), V (~155%)	C (~9%), D (~6%)

5.5.3. Implications for mutational tolerance

While analysing the mutants from the various site saturation libraries, it was obvious that mutations to some sites gave a higher number of detrimental mutants than others. The plots shown in **Figure 5.9** show the activity for each mutant plotted. While these plots do not represent a thorough mutational tolerance study on MPH, they give an indication of the mutational tolerance for each site. By looking at the distribution of mutants retaining less than 50% WT MPH activity on EPO, the mutational tolerance of the chosen sites can be classified into intolerable (>70% mutants), moderately tolerable (40 – 70% mutants) and tolerable (<40% mutants). Mutation intolerable sites include Arg72, Phe119 and Trp179; moderately tolerable sites include Pro150, Phe196 and Leu273; and L258 is the only highly tolerable site. For example, most of the changes to Trp179 resulted in a substantial loss of activity (**Figure 5.9e**) while most of the changes to Phe196 resulted in some loss of activity, although very few of the changes caused a total loss of activity (**Figure 5.9f**). Analysing the mutational tolerance can draw some hypotheses regarding the possible roles of these sites. For example, Phe119 and Phe196 exhibited different mutational tolerance and could be an indication to which of these phenylalanine residues is essential for MPH activity. The highly tolerant L258 residue could suggest that L258 is the key to enzyme promiscuity on MPH.

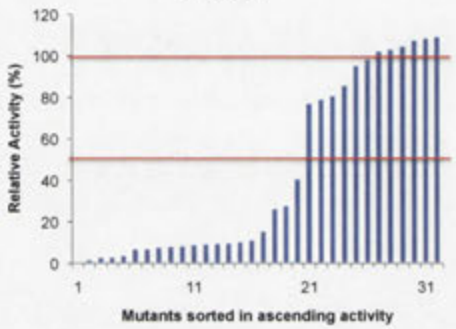


c)



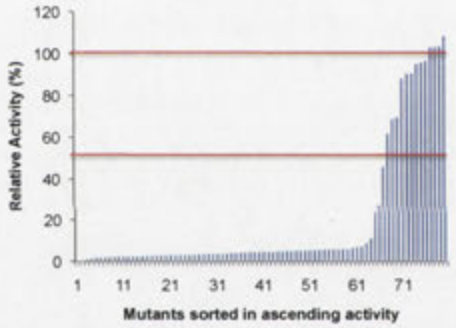
Percentage of mutants with more than 100% activity = 25.0%
Percentage of mutants with less than 50% activity = 72.5%

d)



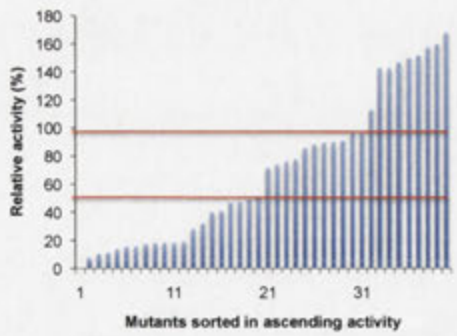
Percentage of mutants with more than 100% activity = 18.75%
Percentage of mutants with less than 50% activity = 50.5%

e)



Percentage of mutants with more than 100% activity = 5.0%
Percentage of mutants with less than 50% activity = 85.0%

f)



Percentage of mutants with more than 100% activity = 22.5%
Percentage of mutants with less than 50% activity = 47.5%

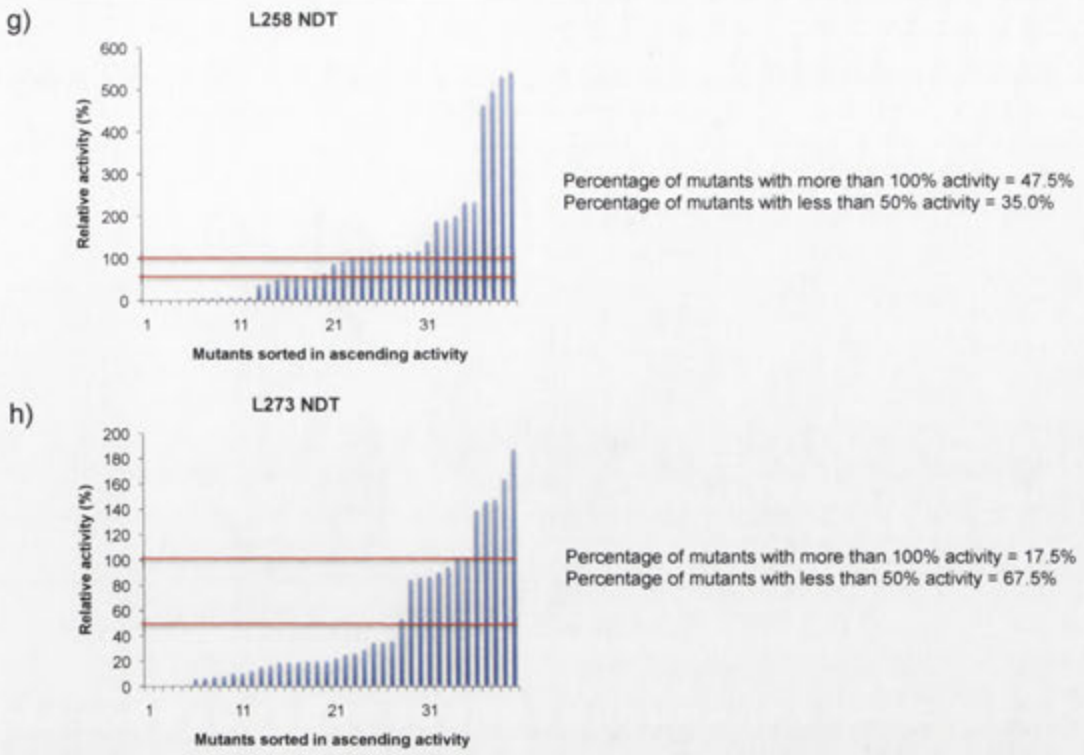


Figure 5.9: Mutant activity for all the libraries sorted in ascending order.
The activities for all the individual mutants were normalised to WT MPH. The red line indicates where the WT MPH activity was and the numbers in x-axis indicate the designated mutant number. The first red line indicates 100% relative activity to WT and the second red line indicated 50% relative activity to WT.

5.5.4. Explanation for loss of activity at positions 72, 150 and 179

In R72 NNK library, it was found that mutations R72T, R72A and R72G resulted in activity loss with 14%, 36% and 11% relative activities retained respectively. While the Arg72 side chain is not interacting with other residues in the structure, mutation to uncharged amino acid (Thr) and abolishing the side chain (Ala, Gly) seemed to cause the activity loss, indicating that the charged side chain is important in maintaining activity. However, when comparing the two MPH crystallographic subunits, some anomalies involving Arg72 were observed and are discussed in **section 5.8.**

In P150 NDT library, P150Y and P150H mutations were found to result in activity drop with the mutants having 27% and 8% relative activities retained respectively. This was probably due to steric effect introduced by the relatively bigger sides chains. P150 is located in between the metal coordinating His149 and Asp151. From the wild type crystal structure, Pro150 is surrounded by Phe119, Leu118 and Phe196 (**Figure 5.10**). Mutating Pro150 into residues with larger side chain (His, Tyr)

would introduce clash with the aforementioned residues, causing a loss in activity. Phe119 and Phe196 are key substrate binding residues and will be discussed in later sections.

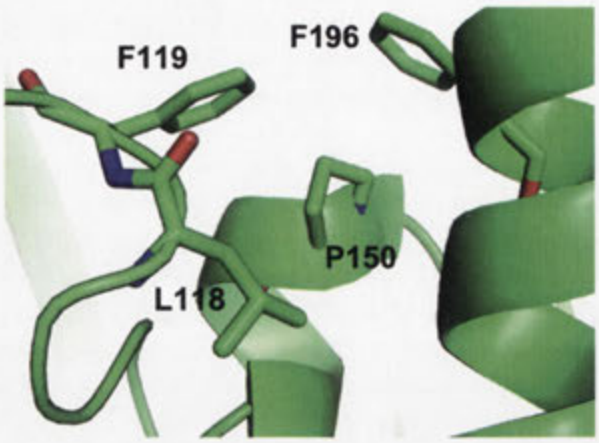


Figure 5.10: The residues in the vicinity of Pro150.
Leu118, Phe119 and Phe196 are located in the vicinity of Pro150. Phe119 and Phe196 are thought to be involved in substrate binding. The surrounding residues with bulky side chains restrict Pro150 to only tolerate residues with small side chain. This was supported by the importance of Phe119 and Phe196 that will be discussed in later sections.

In W179 NNK library, four deleterious mutations sampled were W179S, W179H, W179Y and W179M and their respective relative activities were 4%, 10%, 17%, and 20%. This activity deterioration could be due to the disruption of hydrogen bonding between the side chains of Trp179 and Glu175. Glu175 seems to act as a bridging residue between the side chain of Trp179 and the backbone of M148 (**Figure 5.11**). Replacing W179 with other amino acid residue would result in the disruption of the interaction that might have negative structural repercussions. Interestingly, W179H mutation also showed a detrimental effect to the activity. W179H seemed to be able to form a hydrogen bond with E175 due to similarity of the imidazole ring in histidine with the pyrole ring in tryptophan (**Figure 5.12a**). However, histidine can exist in two tautomeric structures (**Figure 5.12b**), possibly affecting the interaction with E175. Dong *et al.* found that W179F mutant retains WT-like k_{cat} and K_m while W179A shows diminished k_{cat} and increased K_m against MPS.[3] While the observation for W179A is consistent with the hypothesis that W179 interaction is required for stability, observation for W179F suggests a different role for W179.

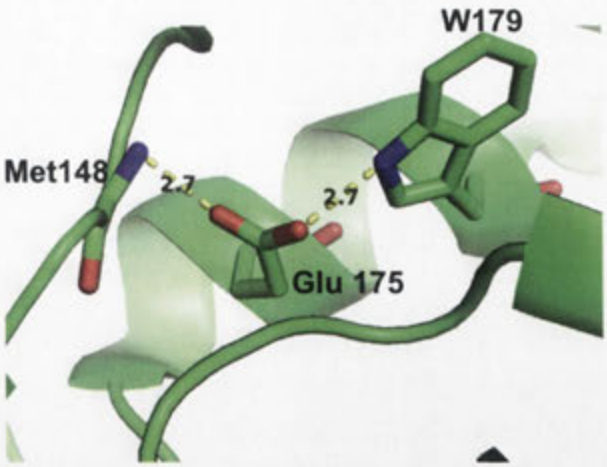


Figure 5.11: The interaction between Trp179, Glu175 and Met148.
The numbers and dashed lines show the distances of the interactions between Trp179, Glu175 and Met148. Glu175 is acting as a bridging residue for Trp179 and Met148, possibly providing structural stability.

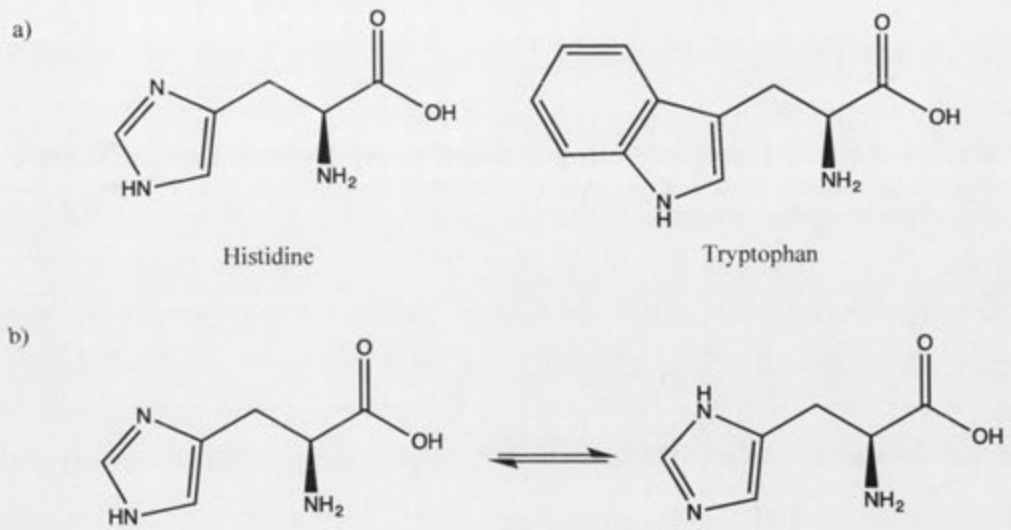


Figure 5.12: The structural comparison of histidine, tryptophan and the tautomeric forms of histidine.
a) The histidine imidazole ring is similar to the pyrole ring in tryptophan, allowing hydrogen bond to form. b) Histidine can exist in two tautomeric forms affecting where the hydrogen bond would be formed.

5.6. Shuffling of single site mutants

A brief relative activity validation test was done as described in **section 2.7.1** on the mutants that showed improvement against EPO during screening to shortlist mutants for purification and kinetic studies (data not shown). Ten mutants, namely F119Y, F196L, F196I, F196Y, L258S, L258I, L258H, L258N, L273I and L273V were picked for purification and kinetic studies. The expression and purification were done

as described in **Chapter 3.** However, the scales of expression and purification were smaller. The MPH variants in this chapter were expressed in 30 mL LBK supplemented with 0.1 mM $ZnCl_2$ and purified with Column 2 (5 mL HiTrap Sp FF) of purification workflow III (**section 3.2.2.4**). $ZnCl_2$ was maintained at 0.1 mM during purification to ensure sufficient Zn^{2+} ions were available to the enzyme. Kinetic parameters for methyl parathion (MPS), methyl paraoxon (MPO), ethyl parathion (EPS), ethyl paraoxon (EPO), methyl chlorpyrifos oxon (MCO), ethyl chlorpyrifos oxon (ECO) were obtained for all the MPH variants and are discussed in **section 5.7**. The substrates used cover the variable substituents of the substrate – dimethyl vs diethyl, P=S vs P=O moieties and *p*-nitrophenolate vs 3,5,6-trichloro-2-pyridinol leaving groups.

All ten MPH mutants and wild type variants were used as the parents to shuffle the single site mutations. Including the wild type variant in the shuffling procedure would allow us to use it as a control for the screening process – the confidence of the screen is higher if none of the selected variants returns as wild type even though it is included in the parent mix. The shuffling was done using staggered extension process (StEP) as described in **section 2.5.2.**[15] Error-prone StEP was not used at this stage to avoid the introduction of mutations remote from the active site and interfere with mechanistic assessment of MPH – we would like to confine any activity changes to only active site residue mutations. The shuffled mutants were screened similarly as the single site saturation library – for further enhancement in EPO activity. Further activity enhancement was deemed possible via shuffling as all the selected single site mutants showed enhanced activity against EPO. The ten single site mutants represent the possible amino acid variations at four different sites for activity improvement against EPO. Two rounds of StEP shuffling were carried out, with multiple site variants from the first shuffling being used as the parent for the second shuffling and the results for these screens are summarised in **Table 5.6.** After two rounds of StEP, only double and triple-site mutants were found. This could be due to the distance between some of the mutations. For example, L258 and L273 are only separated by 45 nucleotides. The recombination efficiency decreases as the distance between mutations decreases.[15] The apparent lack of shuffling could also be due to silent crossovers in DNA shuffling. Shuffling that occured in the region without mutation would result in silent crossover.

Table 5.6: Summary of (a) StEP libraries generated and number of mutants screened and (b) improved mutants selected from the screen and the mutations found in them.

a)

	Number of mutants screened
Round 1 StEP	400
Round 2 StEP	400

b)

	Selected mutants	Mutation
Round 1 StEP	R1B2	F119Y, L258N
	R1D4	F196I, L273V
	R1A6	F119Y, L258S
	R1B6	F196I, L258N
	R2B2	F196I, L258H
	R2C2	F196I, L273I
	R2D2	F119Y, L258H
	R2F2	F196Y, L258H
Round 2 StEP	R2G4	F196L, L258H
	R2A2	F119Y, F196I, L273V
	R2E3	F119Y, F196I, L258N
	R2F3	F119Y, F196I, L258H
	R2C4	F119Y, F196Y, L258H
	R2D4	F119Y, F196I, L273I

5.7. Kinetic characterisation of MPH variants

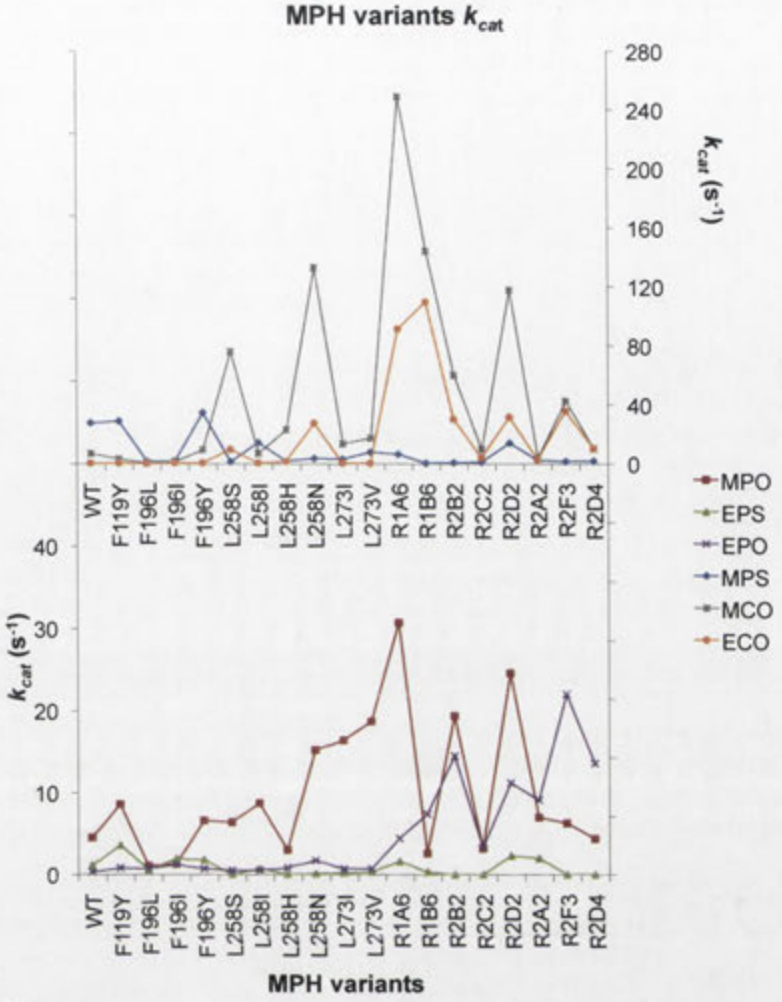
As mentioned in the previous section, all improved MPH single site mutants from the site saturation mutagenesis libraries and the wild type variant were characterised against six organophosphate substrates and the kinetic parameters can be found in **Figures 5.13a** (k_{cat}), **5.13b** (K_m) and **5.13c** (k_{cat}/K_m). In general, all of the single site mutants selected from the EPO screening showed modest enhancement in k_{cat}/K_m.

The substrate concentrations used for the kinetic measurements for all variants were based on concentrations used for wild type MPH. As a result, higher curve fitting errors were observed in some mutants with greatly increased K_m on some substrates. However, the changes in some K_m values were so large that the overall trend in substrate specificity of a given variant can still be observed even after the errors were

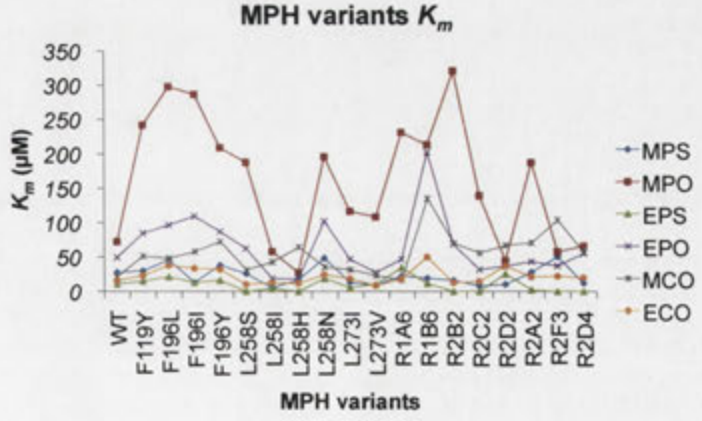
taken into consideration. For all the kinetic data presented, the errors obtained for the kinetic parameters represent the fitting error of the data to Michaelis-Menten equation.

The substrate profiles for the selected mutants (listed in **Table 5.6b**) from the shuffling libraries were investigated using crude lysate for activity validation (data not shown). The assay procedure was similar to the library screen described in **section 2.5.6** except that the growth scale was larger. R1A6, R1B6, R2B2, R2C2, R2D2, R2A2, R2F3 and R2D4 were shortlisted for purification and kinetic studies. These mutants were picked for further study based on improvement seen in the crude lysate assay. Mutants that were not shortlisted either had only moderate improvements, or were represented by other variants that exhibited similar profile. The purification and kinetic characterisation were performed similarly to the single site mutants and were described in **sections 5.6 and 2.7.2** respectively. Kinetic parameters for the selected multiple-site mutants on six OP substrates are also found in **Figures 5.13a** (k_{cat}), **5.13b** (K_m) and **5.13c** (k_{cat}/K_m). As hypothesised, all the multiple site variants showed further improvement towards EPO in terms of k_{cat} and catalytic efficiencies. Unexpectedly, improvement was also seen against ECO for all variants since not all single site mutants show improvement towards ECO. Most mutants showed lower k_{cat} and catalytic efficiencies towards MPS and EPS, probably due to a shift of specificity away from phosphorothioate. The kinetics for L258S, L258H, R2B2, R2C2, R2F3 and R2D4 towards EPS were difficult to determine due to the low k_{cat} and K_m of these mutants. Tables detailing the kinetic parameters for all MPH variants characterised are given in **Appendix G**. Detailed discussion of the substrate specificity and possible active site change in MPH variants are given in **sections 5.7.1 – 5.73**.

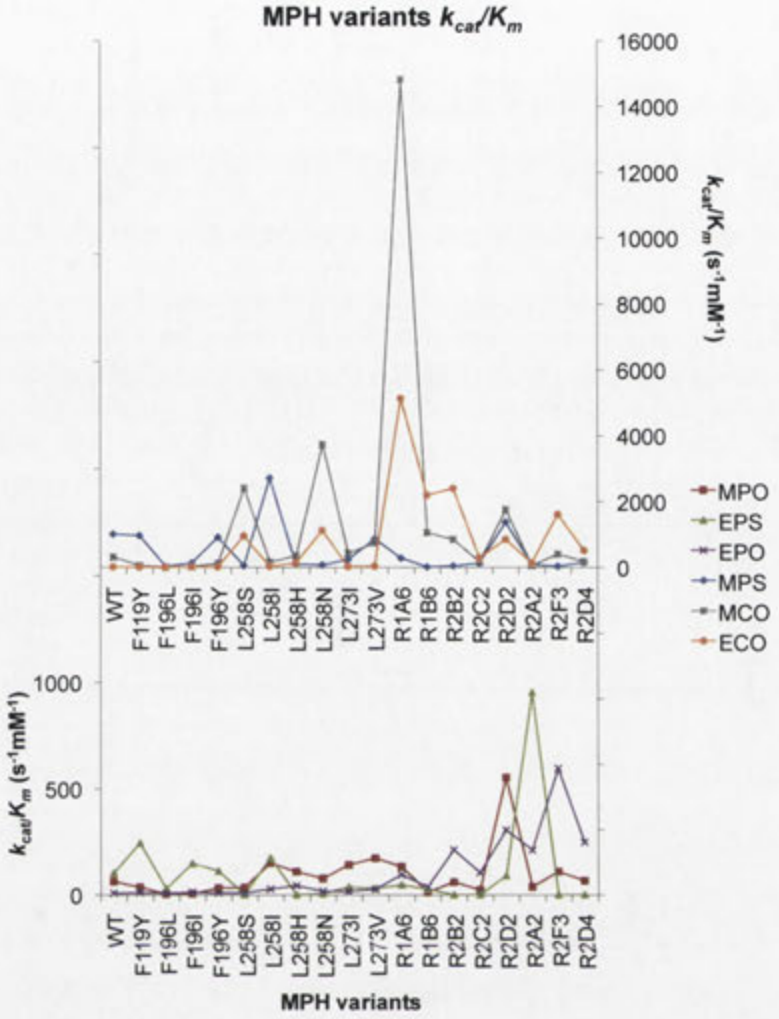
a)


MPH variants k_{cat}
k_{cat} (s^{-1})
MPH variants
MPO
EPS
EPO
MPS
MCO
ECO
WT
F119Y
F196L
F196I
F196Y
L258S
L258I
L258H
L258N
L273I
L273V
R1A6
R1B6
R2B2
R2C2
R2D2
R2A2
R2F3
R2D4

b)


MPH variants K_m
K_m (μM)
MPH variants
MPS
MPO
EPS
EPO
MCO
ECO
WT
F119Y
F196L
F196I
F196Y
L258S
L258I
L258H
L258N
L273I
L273V
R1A6
R1B6
R2B2
R2C2
R2D2
R2A2
R2F3
R2D4

c)



Variants	Mutations
R1A6	F119Y, L258S
R1B6	F196I, L258N
R2B2	F196I, L258H
R2C2	F196I, L273I
R2D2	F119Y, L258H
R2A2	F119Y, F196I, L273V
R2F3	F119Y, F196Y, L258H
R2D4	F119Y, F196I, L273I

Figure 5.13: Kinetic parameters for the hydrolysis of various OPs by MPH variants isolated in this study.

The graphs showing a) k_{cat} (s^{-1}), b) K_m (μM) and c) k_{cat}/K_m ($s^{-1}mM^{-1}$) of MPH variants. R1A6, R1B6, R2B2, R2C2, R2D2, R2A2, R2F3 and R2D4 are multiple-site mutants obtained from DNA shuffling and the mutations are specified in the table above. In graphs (a) and (c), the graphs are separated into top and bottom panels to accommodate the large value difference between MPH variants for the respective kinetic parameter.

5.7.1. Wild type MPH

By comparing the k_{cat} values of different substrates, wild type (WT) MPH is most active on MPS, followed by MCO, MPO, and EPS. The k_{cat} values for EPO and ECO were similar. This shows that the WT variant was most active towards dimethyl-substituted monothiophosphates (MPS), with the lowest activity coming from diethyl-substituted oxon substrates (EPO, ECO). It is also worth noting that WT MPH was more active towards dimethyl-substituted oxons (MPO) than diethyl-substituted thiophosphate (EPS). This shows the priority dimethyl-diethyl factor has over oxon-thion factor. Interestingly, WT MPH showed a higher activity towards 3,5,6-trichloro-2-pyridinol (TCPy) leaving group oxons (MCO, ECO) than pNP leaving group oxons (MPO, EPO) indicating the influence the leaving group has on the k_{cat} value. Due to the low substrate solubility of methyl chlorpyrifos (MCS) and ethyl chlorpyrifos (ECS), the kinetic parameters for these two substrates were not determined. Based on the trend in pNP based substrates, it could be postulated that the k_{cat} for MCS and ECS are higher than the oxon counterparts (MCO and ECO), and higher than pNP counterparts (MPS and EPS). This would require further confirmation in future.

In terms of K_m, WT MPH had higher K_m for oxon substrates when compared to thiophosphate substrates. The K_m values for diethyl substituted substrates (EPS, EPO, ECO) were lower compared to the dimethyl substituted substrates (MPS, MPO, MCO). The K_m values for TCPy oxons (MCO and ECO) were lower when compared to the pNP counterparts (MPO and ECO). The k_{cat}/K_m ratio was approaching diffusion rate for MPS at 1.01 x 10^6 $s^{-1}M^{-1}$, consistent with the kinetics reported by Dong *et al.* at 0.99 x 10^6 $s^{-1}M^{-1}$.[3] A summary of WT MPH kinetics towards various substrates can be found in **Table 5.7**.

Table 5.7: Kinetic parameters for MPH WT on various OPs.

Variant	Substrate	k_{cat} (s^{-1})	Relative k_{cat}	Km (μM)	Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
WT	MPS	27.5 ± 1.800	1.00	27.3 ± 6.0	1.00	1007.3	1.00
	MPO	4.5 ± 0.300	1.00	72.2 ± 10.2	1.00	62.3	1.00
	EPS	1.1 ± 0.004	1.00	10.8 ± 1.4	1.00	101.9	1.00
	EPO	0.3 ± 0.010	1.00	49.0 ± 5.5	1.00	6.1	1.00
	MCO	6.9 ± 0.100	1.00	21.4 ± 1.6	1.00	322.4	1.00
	ECO	0.3 ± 0.010	1.00	15.6 ± 2.5	1.00	19.2	1.00

5.7.2. Single site mutants

In order to understand what the mutations at each of the sites do and how they change the substrate binding pocket, SWISS-MODEL homology modelling server was used.[16–18] For reasons discussed in **section 5.8**, subunit B of WT MPH structure (PDB ID: 1P9E-B) was used as the template for mutation modelling.

5.7.2.1. F119 variants

Residue F119 seemed to be essential for activity. From the activity distribution plot for F119 site saturation library, as shown in **Figure 5.9c**, only the wild type residue F119 and F119Y, retained high EPO hydrolysing activity. Approximately two thirds of the screened mutants had less than 50% EPO activity compared to wild type. The only improved variant against EPO, F119Y, suggests that the π-bonds on the aromatic rings are important in substrate binding, particularly the potential to form π-stacking with the OP substrates. From the modelled F119Y structure, the mutation altered the shape of the active site opening **(Figure 5.14)** and introduced a hydroxyl group that can form hydrogen bond with the backbone of A66, possibly stabilising the residue at position 119 **(Figure 5.15).**

The k_{cat} values of F119Y for MPO, EPS and EPO increased, while remaining similar to WT for MPS and ECO. The only substrate that observed a drop in k_{cat} was MCO. There was an overall increase in K_m for all substrates examined with the exception of MPS and EPS. K_m values for MPS and EPS remained similar to WT. EPS and EPO were the only two substrates that observed an increase in k_{cat}/K_m (2.39 fold and 1.56 fold, respectively).

Dong *et al.* reported that F119W shows a decrease in k_{cat} and K_m while F119A shows a very small increase in K_m and a decrease in k_{cat} against MPS.[3] The small K_m increase reported for F119A (37.4 μM to 41.1 μM) might be within error since the error values were not reported. They proposed that F119 is part of an aromatic cluster that is involved in enzyme-substrate binding, analogous to those in OPDA. While the F119W data agreed with the proposal, F119A did not. In this work, F119G and F119N were sampled from the pool of deleterious mutations, consistent with the proposal that the π-bond on the aromatic rings in phenylalanine and tyrosine are responsible for substrate

interaction. F119G mutation abolishes the benzene ring while F119N mutation changes a hydrophobic residue to a polar residue, affecting the interaction. Since the NDT randomisation used for site saturation mutagenesis in this library does not code for alanine, F119A was not isolated in this work. A summary of F119Y kinetics towards various substrates can be found in **Table 5.8.**

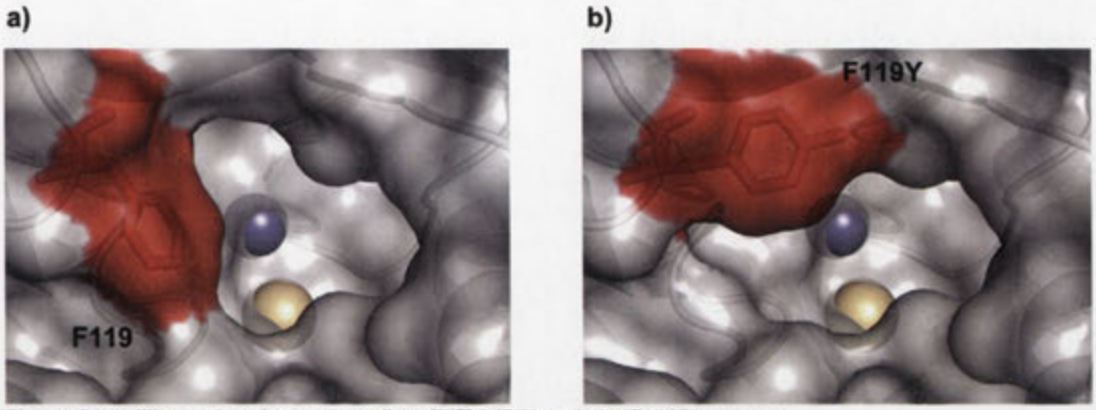


Figure 5.14: The active site pocket of (a) WT MPH and (b) F119Y mutant.
The residue at position 119 is coloured red and the grey surfaces represent the solvent excluded surfaces of WT MPH in (a) and F119Y in (b). The mutation was modelled with SWISS-MODEL using subunit B of WT MPH as template. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal.

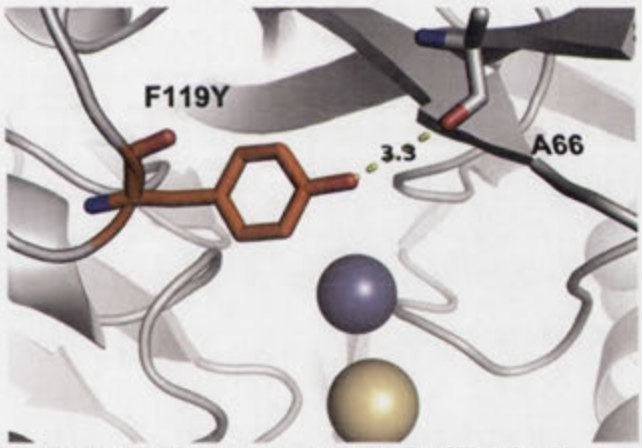


Figure 5.15: Possible hydrogen bond interaction between F119Y and A66.

Table 5.8: Kinetic parameters for F119Y on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for F119Y with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value.

Variant	Substrate	k_{cat} (s^{-1})			Relative k_{cat}	K_m (μM)			Relative K_m	k_{cat}/K_m ($s^{-1}mM^{-1}$)	Relative k_{cat}/K_m
F119Y	MPS	28.8	±	0.300	1.05	29.8	±	1.0	1.09	966.4	0.96
	MPO	8.6	±	1.500	1.91	241.9	±	65.3	3.35	35.6	0.57
	EPS	3.6	±	0.200	3.27	14.8	±	2.2	1.37	243.2	2.39
	EPO	0.8	±	0.100	2.67	84.3	±	21.1	1.72	9.5	1.56
	MCO	3.3	±	0.040	0.48	51.5	±	1.7	2.41	64.1	0.20
	ECO	0.3	±	0.002	1.00	23.1	±	0.9	1.48	13.0	0.68

5.7.2.2. F196 variants

Three variants that were more active against EPO were selected for at this site, namely F196L, F196I and F196Y. As with the F119Y mutant, F196Y gave an increase in k_{cat} values for all substrates tested. F196Y also retained the aromatic ring, as seen in F119Y. The increase in k_{cat} was accompanied by an increase in K_m for all substrates, resulting in EPS and EPO being the only substrates that observed an overall increase in k_{cat}/K_m (1.10 fold and 1.32 fold, respectively).

F196L and F196I are two interesting mutations; they replace an aromatic side chain with smaller hydrophobic side chains. The difference between leucine and isoleucine is a rearrangement of the methyl groups in the side chain to accommodate bulkier OPs. F196L gave a decrease in k_{cat} for all substrates except EPO, the screening substrate, while F196I gave a decrease in k_{cat} for dimethyl-substituted OPs (MPS, MPO, MCO). The k_{cat} values for F196I against diethyl-substituted substrates EPS, EPO and ECO increase 1.73, 4 and 2 folds respectively. The K_m values for both F196L and F196I increased for all substrates, except F196I for MPS where the K_m decreases. F196L sees an overall decrease in k_{cat}/K_m except EPO, where the k_{cat}/K_m increased 1.19 fold. F196I gave a decrease in k_{cat}/K_m against most substrates except EPS and EPO where the in k_{cat}/K_m increased 1.45 and 1.80 folds, respectively. It is worth noting that despite all the F196 variants showing increase in k_{cat} and k_{cat}/K_m against EPO, they were still lower than that of MPS, the fastest substrate examined. A summary of F196 variants kinetics towards various substrates can be found in **Table 5.9** and a comparison of active site changes can be found in **Figure 5.16.**

Dong *et al.* reported the kinetics of F196W and F196A against MPS to demonstrate that F196 is part of an aromatic cluster responsible for substrate binding.[3] F196W gave an increase in k_{cat} *and* K_m against MPS, leading to a small increase in k_{cat}/K_m. Meanwhile, F196A gave a large increase in K_m and a large decrease in k_{cat} against MPS, consistent with the observation made here where an aromatic ring is required to retain the high hydrolysing efficiency of MPH. In this work, F196I and F196L retained only 0.05 and 0.06 folds k_{cat} compared to WT MPH in hydrolysing MPS, while F196Y showed 1.25 fold k_{cat} improvement for the same substrate.

Jeong *et al.* found F196A increases the k_{cat} against pNPDPP by 78 fold while the K_m increases 2.7 fold leading to an overall increase in k_{cat}/K_m of 29.2 fold (2.16 x 10^4 $s^{-1}M^{-1}$) in evolving MPH against *p*-nitrophenyl diphenylphosphate (pNPDPP) (**Figure 5.17**). The authors attributed the improvement in catalytic efficiency to the reduction in side chain size from phenylalanine to alanine – enlargement of the substrate-binding pocket that eventually facilitated the substrate access.[19] Based on this information, we deduce that the lack of activity against bpNPP observed in this work is probably due to the charge on the phosphate core rather than the bulk of bpNPP, since bpNPP is smaller than pNPDPP (**Figure 5.17**). This is supported by the presence of predominantly hydrophobic residues that line the substrate-binding pocket. We can also deduce that k_{cat} improvements seen for some of the substrates in F196L and F196L were partly due to the enlargement of the substrate-binding pocket to accommodate larger substrates such as EPO.

Another possible role F196 might have is the interaction with H149, a metal coordinating ligand. The imidazole ring of histidine residues and benzene ring of phenylalanine residues are capable of forming π-interaction due to its delocalized electrons. F196 contains a benzene ring that is perpendicular to H149 at approximately 3.6 Å away (**Figure 5.18**). The decrease in catalytic efficiencies of F196I and F196L mutants on most substrates examined might be an indication of the importance of the aromatic ring.

Mutations that were found to be detrimental are F196R and F196D, consistent with the premise that hydrophobic residues are required in the substrate-binding pocket and the aromatic ring is critical for k_{cat}.

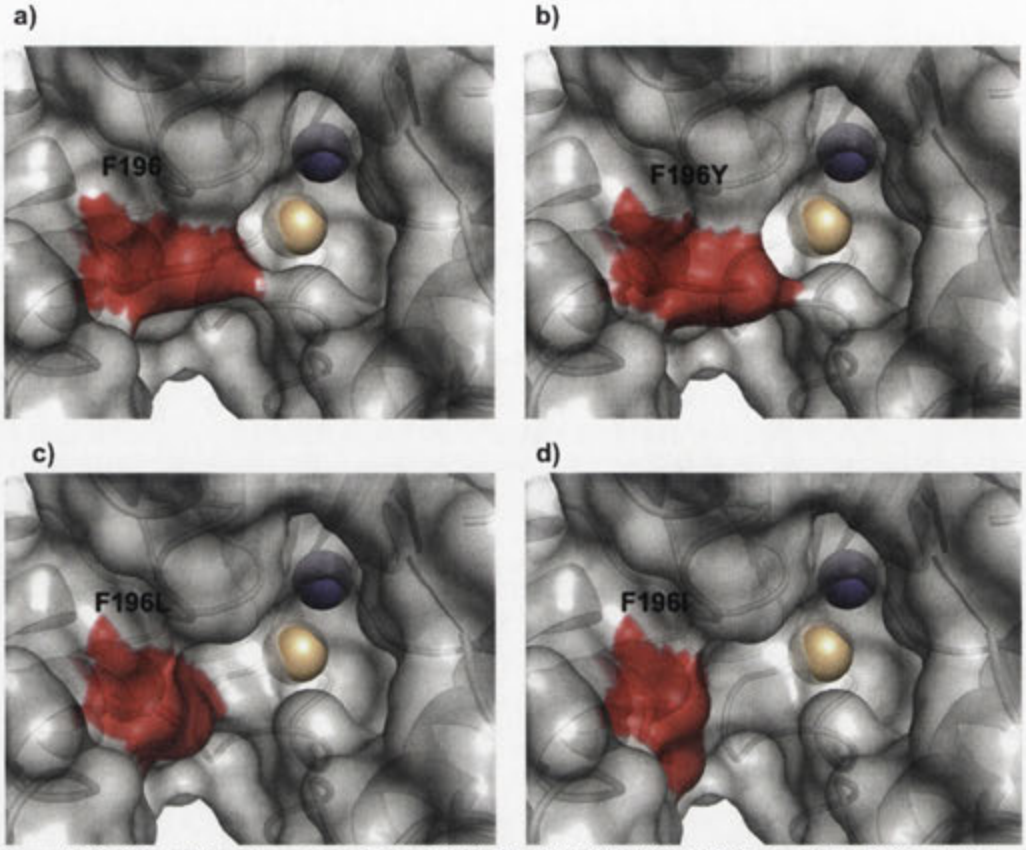
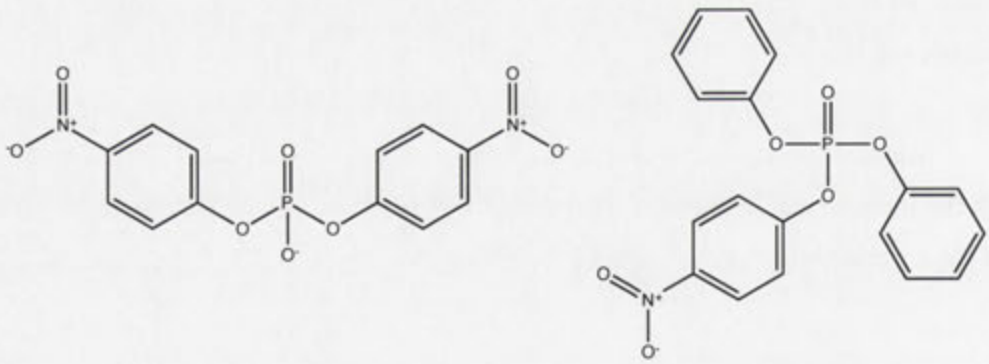


Figure 5.16: The active site pocket of (a) WT MPH, (b) F196Y, (c) F196L and (d) F196I variants. The residue at position 196 is coloured red and the grey surfaces represent the solvent excluded surfaces of WT MPH in (a), F196Y in (b), F196L in (c) and F196I in (d). The mutation was modelled with SWISS-MODEL using subunit B of WT MPH as template. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal.



bis(*p*-nitrophenyl) phosphate

p-nitrophenyl diphenylphosphate

Figure 5.17: Chemical structures of bis-*p*-nitrophenyl phosphate (bpNPP) and *p*-nitrophenyl diphenylphosphate (pNPDPP).

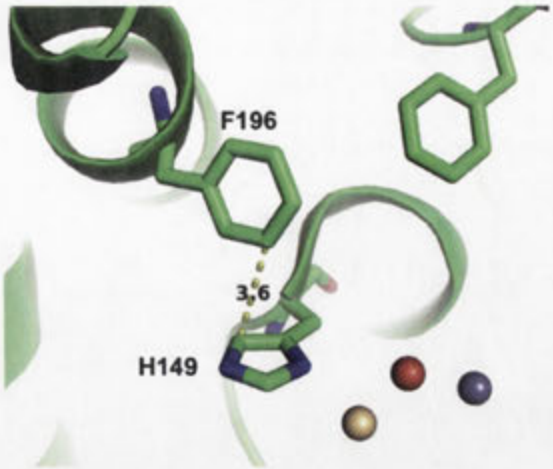


Figure 5.18: The possible interaction between F196 and H149.
The yellow dash line and number indicate the measured distance of the two residues, the wheat sphere illustrates Cd^{2+}, grey sphere illustrates Zn^{2+} and the red sphere illustrates the bridging water.

Table 5.9: Kinetic parameters for F196 variants on various OPs
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for F196 variants with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value.

Variant	Substrate	k_{cat} (s^{-1})			Relative k_{cat}	K_m (μM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
F196L	MPS	1.5	±	0.060	0.05	45.6	±	4.5	1.67	32.9	0.03
	MPO	1.1	±	0.100	0.24	297.9	±	53.1	4.13	3.7	0.06
	EPS	0.6	±	0.020	0.55	21.2	±	2.7	1.96	28.3	0.28
	EPO	0.7	±	0.030	2.33	96.3	±	9.5	1.97	7.3	1.19
	MCO	0.5	±	0.010	0.07	48.3	±	4.0	2.26	10.4	0.03
	ECO	0.2	±	0.010	0.67	37.8	±	4.4	2.42	5.3	0.28
F196I	MPS	1.7	±	0.050	0.06	12.9	±	1.7	0.47	131.8	0.13
	MPO	1.3	±	0.100	0.29	286.8	±	45.6	3.97	4.5	0.07
	EPS	1.9	±	0.100	1.73	12.9	±	1.8	1.19	147.3	1.45
	EPO	1.2	±	0.100	4.00	109.1	±	9.8	2.23	11.0	1.80
	MCO	1.5	±	0.040	0.22	57.9	±	4.1	2.71	25.9	0.08
	ECO	0.6	±	0.004	2.00	33.7	±	0.9	2.16	17.8	0.93
F196Y	MPS	34.5	±	0.700	1.25	37.8	±	2.2	1.38	912.7	0.91
	MPO	6.6	±	0.600	1.47	208.8	±	30.9	2.89	31.6	0.51
	EPS	1.8	±	0.040	1.64	16.1	±	0.8	1.49	111.8	1.10
	EPO	0.7	±	0.070	2.33	86.7	±	18.8	1.77	8.1	1.32
	MCO	9.5	±	0.400	1.38	72.5	±	7.8	3.39	131.0	0.41
	ECO	0.5	±	0.010	1.67	31.7	±	2.6	2.03	15.8	0.82

5.7.2.3. L258 variants

L258 is the most interesting site because four different mutants were found to have increased activity against EPO and about half of the library exhibited at least WT activity against EPO **(Figure 5.9g).** L258 forms part of the base of the active site pocket **(Figure 5.19b)**. The beneficial mutations can be grouped into hydrophilic (L258H, L258N, L258S) and hydrophobic mutations (L258I). From the WT crystal structure, there are a few potential residues in the vicinity, with which L258 mutants could interact with, namely D255 and T271 **(Figure 5.19a).** Interactions with T271 might

further affect residues L273 and D274 since they are close to T271. D255 is the metal bridging ligand, D274 is a secondary shell metal coordination ligand (interacting with H234, a primary shell metal ligand) and L273 is an active site lining residue. All these might contribute to the flexibility of this site.

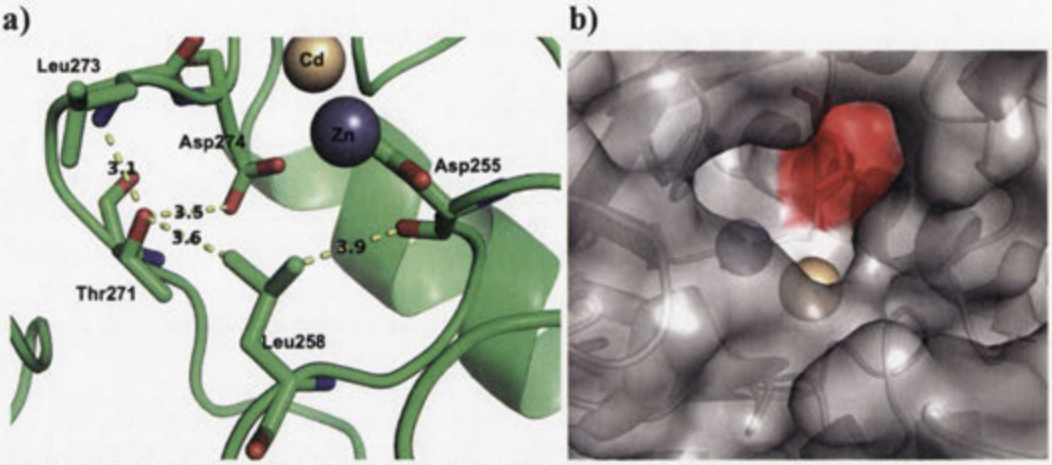


Figure 5.19: (a) L258 and nearby residues and (b) location of L258 in the active site pocket.
(a) There are a number of nearby residues that can interact with L258 variants, particularly Thr271 and Asp255. Thr271 seems to interact with Leu273, another candidate for substrate binding, and Asp274. Asp255 is the metal-bridging residue. The dashed lines represent the distance and the numbers represent the distance in Angstrom. (b) Leu258 is located at the bottom of the active site pocket and is coloured red.

L258I is a conservative mutation that fine-tunes the substrate-binding pocket for oxon substrates. The k_{cat} for all oxon substrates increased while the K_m for most substrates decreased, except MCO where the K_m increased two fold. This leads to an overall increase in k_{cat}/K_m for all substrates, except MCO where the efficiency decreased by half. **Figure 5.20** illustrates the difference in solvent excluded surface for L258 and I258 modelled using SWISS-MODEL. The L258I mutation changes the position of its side chain, altering the shape of the substrate pocket.

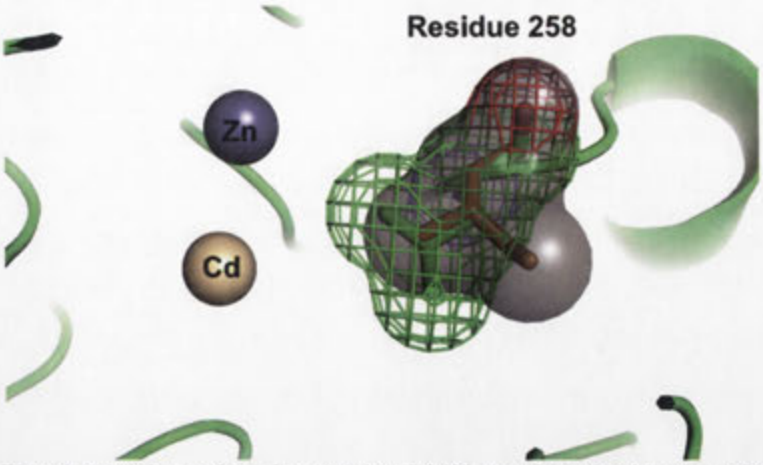


Figure 5.20: The changes in surface for position L258I possibly altering the substrate-binding pocket.
The mesh represents the surface for L258 while the grey surface represents the surface for L258I. The L258I mutant (orange) was modelled with SWISS-MODEL using WT MPH as template.

The mutations that introduce hydrophilic groups are the most interesting mutations of all the changes in libraries. The hydrophilic mutations (L258N, L258S and L258H) introduce side chain with functional groups that are capable of interacting with surrounding residues and possibly the substrate. The k_{cat} trends for L258S and L258N were similar with both variants displaying an overall increase in k_{cat} for oxon OPs and decrease in k_{cat} for thiophosphate OPs. The kinetics of L258S against EPS was not determined due to very low activity against the substrate. Both variants showed large increase in k_{cat} for TCPy leaving group substrates (MCO and ECO), with L258S showing 11.03 and 32.33 folds increase respectively while L258N showed an increase of 19.29 and 91.33 folds increase respectively. It is worth noting that the k_{cat} values of L258S and L258N against MCO were 76.1 s^{-1} and 133.1 s^{-1} respectively, approximately 2.7 and 4.8 folds better than the k_{cat} value of WT MPH against MPS (27.5 s^{-1}).

The L258S mutation gave an overall K_m increase, with the exception of MPS and ECO, where the K_m values were similar to WT. L258N saw an overall increase in K_m values for all the substrates examined. In terms of overall catalytic efficiency, L258S gave an increase with EPO, MCO and ECO while a decrease for the rest. L258N gave an increase in catalytic efficiency against MPO, EPO, MCO and ECO while a decrease against MPS and EPS. The k_{cat}/K_m of L258S against MCO and ECO were 2.42 x 10^6 $s^{-1}M^{-1}$ and 0.96 x 10^6 $s^{-1}M^{-1}$ respectively while k_{cat}/K_m of L258N against MCO and ECO were 3.74 x 10^6 $s^{-1}M^{-1}$ and 1.13 x 10^6 $s^{-1}M^{-1}$ respectively. This represents a shift in

substrate specificity from pNP leaving group substrates towards TCPy leaving group substrates.

While the cause for these improvements is not well understood, L258S introduces an alcohol group that might interact with the substrate and has a smaller side chain than leucine (**Figure 5.21a**). From the modelled L258S structure, the active site pocket has a deeper groove that might accommodate larger substrate. It is also observed that the alcohol group can form intramolecular hydrogen bond with its backbone, which in turn interacts with Gln263. **(Figure 5.21b)**. L258N side chain is similar in size to L258 and again introduces an amide group that might interact with the substrate **(Figure 5.22).**

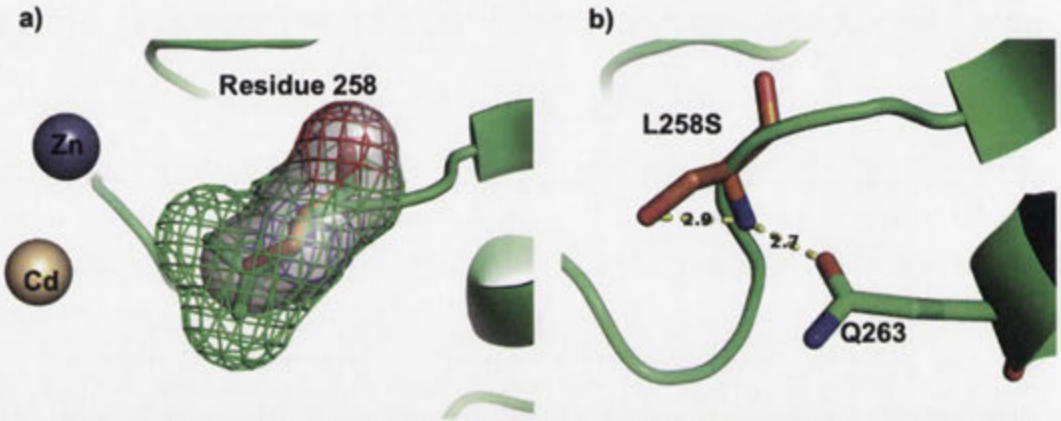


Figure 5.21: (a) The changes in surface for L258S and (b) The intramolecular hydrogen bond introduced by L258S.
a) The mesh represents the surface for L258 while the grey surface represents the surface for L258S. The L258S mutation is coloured orange and was modelled with SWISS-MODEL using WT MPH as template. L258S represents a reduction in surface while introducing an alcohol group. b) The possible intramolecular hydrogen bond due to the introduction of alcohol group. The dashed lines represent the interaction and the numbers represent the distance in Angstrom.

Figure 5.22: The changes in surface for L258N.
The mesh represents the surface for L258 while the grey surface represents the surface for L258N. The mutant is coloured orange and was modelled with SWISS-MODEL using WT MPH as template.

The k_{cat} improvements for L258H were modest compared to L258N. The k_{cat} values for EPO, MCO and ECO increased 2.67, 3.35 and 5.33 folds, respectively, while the k_{cat} values for MPS and MPO decreased. Kinetic parameters against EPS were not determined due to low activity against EPS. The K_m for most substrates decreased except MCO. With regards to catalytic efficiency, L258H gave the biggest increase out of all the single site mutants characterised for EPO, the screening substrate, by 7.05 fold improvement. L258H retained only 0.1 fold of MPS catalytic efficiency. The L258H mutation introduces an imidazole ring that can possibly interact with the backbone of D255, the metal bridging residue **(Figure 5.23b).** The imidazole ring is also within distance (2.9 Å) to form π-stacking with the imidazole ring of H234, another metal coordinating ligand.

The detrimental mutation found in L258 NDT library was L258R. The bulk and charge of the side chain seem to be the cause of the loss in activity. A summary L258 variants kinetics towards various substrates can be found in **Table 5.10**.

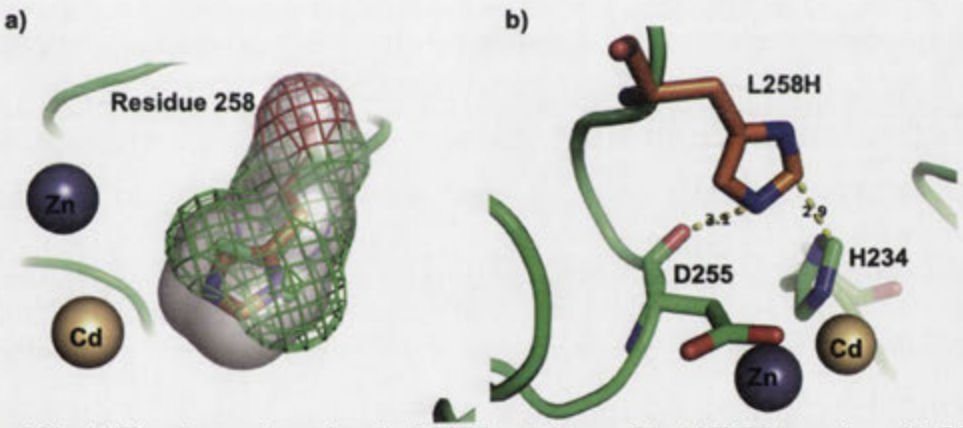


Figure 5.23: (a) The changes in surface for L258H and (b) possible L258H interaction with D255 and H234.
a) The mesh represents the surface for L258 while the grey surface represents the surface for L258H. The L258H mutations is coloured orange and was modelled with SWISS-MODEL using WT MPH as template. b) The L258H mutation introduced an imidazole ring that is capable of interacting with the backbone of D255, the metal bridging ligand. There is also a possibility L258H π-stacks with H234, another metal coordinating ligand. The dashed line and number represents the distance in Angstrom.

Table 5.10: Kinetic parameters for L258 variants on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for L258 variants with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value. ND: not determined

Variant	Substrate	k_{cat} (s^{-1})			Relative k_{cat}	K_m (μM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
L258S	MPS	1.2	±	0.050	0.04	25.8	±	3.7	0.95	46.5	0.05
	MPO	6.4	±	1.000	1.42	187.7	±	47.3	2.60	34.1	0.55
	EPS	ND			ND	ND			ND	ND	ND
	EPO	0.5	±	0.003	1.67	62.4	±	9.0	1.27	8.0	1.31
	MCO	76.1	±	2.700	11.03	31.5	±	4.0	1.47	2415.9	7.49
	ECO	9.7	±	0.200	32.33	10.1	±	1.5	0.65	960.4	50.02
L258I	MPS	14.1	±	0.500	0.51	5.2	±	0.9	0.19	2711.5	2.69
	MPO	8.7	±	0.200	1.93	57.9	±	2.8	0.80	150.3	2.41
	EPS	0.7	±	0.010	0.64	4.0	±	0.5	0.37	175.0	1.72
	EPO	0.5	±	0.010	1.67	18.6	±	1.9	0.38	26.9	4.41
	MCO	7.0	±	0.200	1.01	42.9	±	4.2	2.00	163.2	0.51
	ECO	0.4	±	0.004	1.33	12.7	±	0.7	0.81	31.5	1.64
L258H	MPS	1.7	±	0.070	0.06	16.9	±	2.7	0.62	100.6	0.10
	MPO	3.0	±	0.200	0.67	27.4	±	5.9	0.38	109.5	1.76
	EPS	ND			ND	ND			ND	ND	ND
	EPO	0.8	±	0.040	2.67	18.6	±	3.9	0.38	43.0	7.05
	MCO	23.1	±	0.200	3.35	64.5	±	1.6	3.01	358.1	1.11
	ECO	1.6	±	0.010	5.33	11.8	±	0.5	0.76	135.6	7.06
L258N	MPS	3.5	±	0.300	0.13	48.2	±	10.0	1.77	72.6	0.07
	MPO	15.2	±	0.900	3.38	195.6	±	20.0	2.71	77.7	1.25
	EPS	0.1	±	0.010	0.09	18.8	±	3.7	1.74	5.3	0.05
	EPO	1.7	±	0.040	5.67	102.0	±	5.6	2.08	16.7	2.73
	MCO	133.1	±	2.400	19.29	35.6	±	2.2	1.66	3738.8	11.60
	ECO	27.4	±	0.600	91.33	24.3	±	2.2	1.56	1127.6	58.73

5.7.2.4. L273 variants

L273I and L273V were the improved variants against EPO. L273I and L273V mutations are conservative since Ile and Val are hydrophobic and probably represent the rearrangement of side chains to allow better access to substrates. The opening of the substrate pocket seems to be enlarged that might increase substrate access (**Figures 5.24b and c**).

Both L273I and L273V showed k_{cat} improvements towards oxon OPs and diminished k_{cat} towards thiophosphate OPs. L273I showed K_m decrease towards MPS, EPS and EPO while L273V showed K_m decrease towards MPS, EPS, EPO and ECO. The overall catalytic efficiency for both variants increased for oxon OPs and reduced for thiolphosphate OPs. L273V seems to offer better improvement than L273I. A summary L273 variants kinetics towards various substrates can be found in **Table 5.11**.

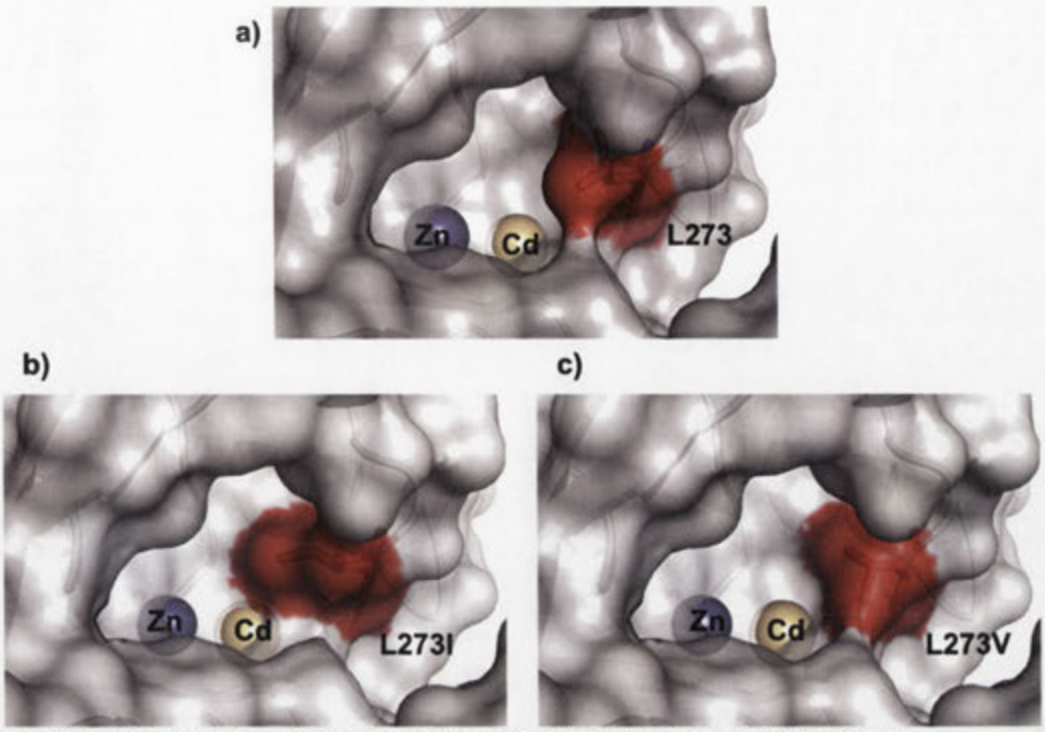


Figure 5.24: The active site pocket of (a) WT MPH, (b) L273I and (c) L273V mutants.
The residues at position 273 are coloured red and the grey surfaces represent the solvent excluded surface of WT MPH in (a), L273I in (b) and L273V in (c). The opening of the active site pocket seems to be the biggest with L273V mutation due to the smaller side chain. The mutations were modelled with SWISS-MODEL using WT MPH as template.

Table 5.11: Kinetic parameters for L273 variants on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for L273 variants with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value.

Variant	Substrate	k_{cat} (s^{-1})			Relative k_{cat}	K_m (μM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
L273I	MPS	3.0	±	0.100	0.11	12.8	±	1.8	0.47	234.4	0.23
	MPO	16.4	±	0.700	3.64	116.5	±	10.5	1.61	140.8	2.26
	EPS	0.2	±	0.003	0.18	5.5	±	0.4	0.51	36.4	0.36
	EPO	0.7	±	0.030	2.33	46.7	±	6.1	0.95	15.0	2.46
	MCO	13.6	±	0.200	1.97	31.7	±	2.0	1.48	429.0	1.33
	ECO	0.5	±	0.020	1.67	18.5	±	3.3	1.19	27.0	1.41
L273V	MPS	7.7	±	0.050	0.28	8.9	±	0.3	0.33	865.2	0.86
	MPO	18.7	±	3.000	4.16	108.1	±	35.3	1.50	173.0	2.78
	EPS	0.3	±	0.030	0.27	10.3	±	4.0	0.95	29.1	0.29
	EPO	0.7	±	0.010	2.33	27.3	±	2.5	0.56	25.6	4.20
	MCO	17.4	±	0.500	2.52	23.4	±	2.9	1.09	743.6	2.31
	ECO	0.3	±	0.006	1.00	8.5	±	1.3	0.54	35.3	1.84

The detrimental mutations found for this site were L273C and L273D. L273C is a polar mutation that might repel the hydrophobic substrate. L273D has a side chain that is similar in size the L273. The detrimental effect might be due to the negative charge on Asp. Jeong *et al.* reported that L273A mutant has lower K_m towards pNPDPP, a much larger substrate, and attributed the improvement to better substrate access due to reduction in side chain bulk.[19]

Since L273 and L258 mutations improved catalytic efficiency towards oxon substrates and the two residues are located near each other, it is possible that this part of the substrate-binding pocket determines the specificity for oxon/thiolphosphate substrate **(Figure 5.25)**. It is also worth noting that L258I did not change the specificity towards oxon substrates. The shift towards oxon OPs only happened when L258 is mutated to a polar residue.

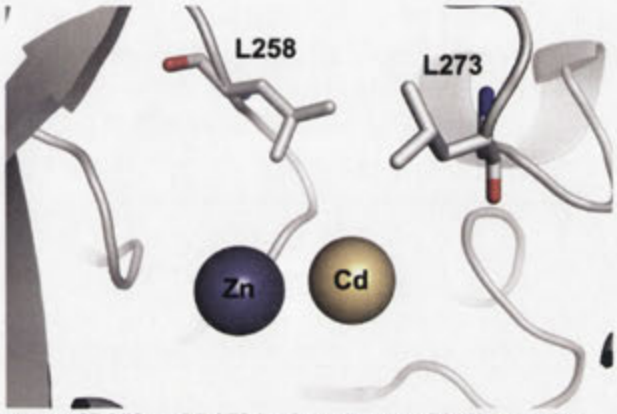


Figure 5.25: The locations of L258 and L273 in the substrate-binding pocket.

5.7.3. Multiple site mutants

5.7.3.1. R1A6

R1A6 contains F119Y and L258S mutations. R1A6 showed k_{cat} improvements for all substrates examined except MPS. The k_{cat} towards MPO, EPS, EPO, MCO and ECO improved 6.82, 1.45, 14.67, 36.1 and 304.33 folds respectively. The K_m for MPS, EPO, MCO and ECO remained similar to WT MPH while K_m for MPO and EPS increased by 3.20 fold and 3.19 fold, respectively. With the exception of MPS and EPS, the catalytic efficiencies for MPO, EPO, MCO and ECO improved. The k_{cat}/K_m for MCO and ECO were 1.47×10^7 $s^{-1}M^{-1}$ and 5.13×10^6 $s^{-1}M^{-1}$ respectively, approaching the diffusion limit. Those figures represent the catalytic efficiency improvement of 45.72 fold and 267.15 fold, respectively. The k_{cat}/K_m for MCO and ECO were also 15 and 5 folds larger than that of WT MPH for MPS, making R1A6 more efficient at turning over TCPy leaving group substrates. From the single-site variant kinetic data, it can postulated that a variant having F119Y and L258N mutations would give an even better improvement towards TCPy leaving group substrates. **Figure 5.26** shows the changes in surface with F119Y and L258S modelled onto WT MPH. The changes

introduced by F119Y and L258S altered the shape of the active site pocket to a narrower and deeper pocket. **Table 5.12** summarises the kinetic parameters for R1A6.

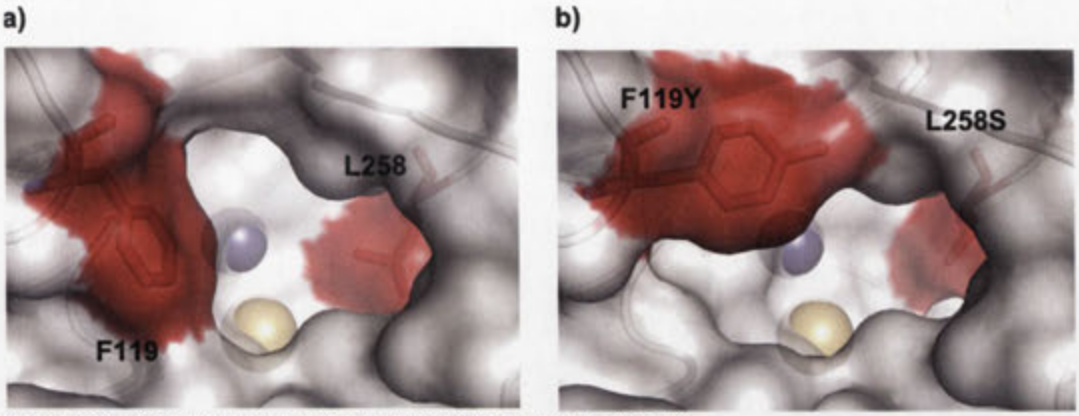


Figure 5.26: The active site pocket of (a) WT MPH and (b) R1A6.
The shape of the active site is altered by F119Y and L258S in R1A6 as compared to WT MPH. The residues at position 119 and 258 are coloured red and the grey surfaces represent the solvent excluded surfaces of WT MPH in (a) and R1A6 in (b) respectively. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal. The mutations were modelled with SWISS-MODEL using subunit B of WT MPH as template.

Table 5.12: Kinetic parameters for R1A6 on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for R1A6 with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m were colour-coded for easy reference. Green: increase in value, pink: decrease in value

Variant	Substrate	k_{cat} (s^{-1})	Relative k_{cat}	K_m (μM)	Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
R1A6	MPS	6.3 ± 0.20	0.23	22.5 ± 2.2	0.82	280.0	0.28
	MPO	30.7 ± 5.50	6.82	231.4 ± 65.2	3.20	132.7	2.13
	EPS	1.6 ± 0.10	1.45	34.5 ± 4.5	3.19	46.4	0.46
	EPO	4.4 ± 0.10	14.67	46.6 ± 2.6	0.95	94.4	15.48
	MCO	249.1 ± 10.30	36.10	16.9 ± 4.1	0.79	14739.6	45.72
	ECO	91.3 ± 5.20	304.33	17.8 ± 5.1	1.14	5129.2	267.15

5.7.3.2. R1B6 and R2B2

This group of variants contains mutations at position F196 and L258.

R1B6 contains F196I and L258N mutations. R1B6 showed decreased k_{cat} against MPS, MPO, and EPS while k_{cat} values against EPO, MCO and ECO have increased by 24.67, 20.90 and 365.67 folds, respectively. The K_m values for all substrate have increased except MPS. This led to a decrease in k_{cat}/K_m for MPS, MPO and EPS. The k_{cat}/K_m values for EPO, MPO and ECO have increased 6, 3.31 and 114.5 folds, respectively. As seen in single-variants, L258H shifts the specificity towards EPO while L258N shifts the specificity towards MCO and ECO. In this variant, the increase in K_m

for MCO due to F196I decreasing the efficiency for MCO. F196I and L258N seem to work synergistically to improve the k_{cat} for EPO and ECO.

R2B2 contains the F196I and L258H mutations. The k_{cat} values against oxon OPs increased while k_{cat} against MPS decreased. Kinetics towards EPS was not determined due to low activity. The k_{cat} against MPO, EPO, MCO, and ECO increased 4.31, 48.33, 8.72 and 101 folds, respectively. The K_m for MPS and ECO decreased, while the *Km* for MPO, EPO and MCO increased. The catalytic efficiencies for EPO, MCO and ECO increased 34.75, 2.65 and 125.25 folds, respectively. It is worth noting that while R1B6 and R2B2 shared similar catalytic efficiency against ECO, R1B6 had higher k_{cat} *and* K_m while R2B2 had lower k_{cat} *and* K_m. The L258H mutation in R2B2 seems to increase the catalytic efficiency for EPO more than L258N in R1B6. **Figure 5.27** shows the changes in the substrate-binding pocket with regards to mutations in R1B6 and R2B2. Mutation at F196I seems to be largely responsible for enlarging the pocket opening in both MPH variants. **Table 5.13** summarises the kinetic parameters for R1B6 and R2B2 on various OPs.

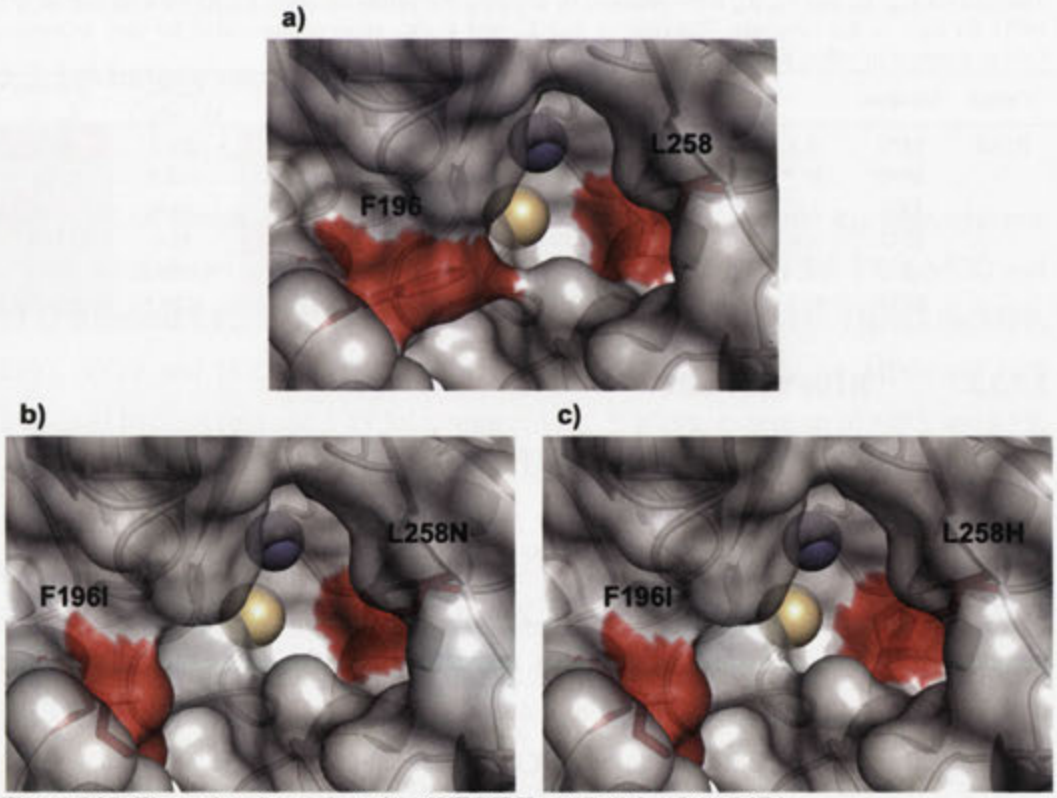


Figure 5.27: The active site pocket of (a) WT MPH, (b) R1B6 and (c) R2B2.
The shape of the active site is altered by F196I and L258N mutations in R1B6 and F196I and L258H in R2B2 as compared to WT MPH. The residues at positions 196 and 258 are coloured red and the grey surfaces represent the solvent excluded surfaces of WT MPH in (a), R1B6 in (b) and R2B2 in (c) respectively. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal. The mutations were modelled with SWISS-MODEL using subunit B of WT MPH as template.

Table 5.13: Kinetic parameters for R1B6 and R2B2 on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for R1B6 and R2B2 with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value

Variant	Substrate	k_{cat} (s^{-1})	Relative k_{cat}	K_m (μM)	Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
R1B6	MPS	0.3 ± 0.01	0.01	19.1 ± 1.5	0.70	15.7	0.02
	MPO	2.6 ± 0.40	0.58	213.3 ± 51.1	2.95	12.2	0.20
	EPS	0.3 ± 0.01	0.27	11.5 ± 0.7	1.06	26.1	0.26
	EPO	7.4 ± 0.20	24.67	202.2 ± 10.5	4.13	36.6	6.00
	MCO	144.2 ± 8.47	20.90	135.3 ± 18.5	6.32	1065.8	3.31
	ECO	109.7 ± 6.10	365.67	49.9 ± 8.0	3.20	2198.4	114.50
R2B2	MPS	0.7 ± 0.01	0.03	16.1 ± 0.8	0.59	43.5	0.04
	MPO	19.4 ± 1.60	4.31	321.0 ± 39.1	4.45	60.4	0.97
	EPS	ND	ND	ND	ND	ND	ND
	EPO	14.5 ± 0.80	48.33	68.4 ± 8.9	1.40	212.0	34.75
	MCO	60.2 ± 1.70	8.72	70.4 ± 5.9	3.29	855.1	2.65
	ECO	30.3 ± 0.90	101.00	12.6 ± 2.8	0.81	2404.8	125.25

5.7.3.3. R2C2, R2D4 and R2A2

This group of variants does not have any of the L258 mutations.

R2C2 contains F196I and L273I mutations. The k_{cat} values towards EPO, MCO and EPO increased while k_{cat} values towards MPS an MPO decreased. Kinetic parameters for EPS were not determined due to low activity towards EPS. F196I and L273I worked synergistically to increase the k_{cat} by 11 fold and 14 fold towards EPO and ECO. The K_m values towards MPO and MCO increased while the K_m values towards MPS and EPO decreased. The K_m for ECO remained similar to WT. The catalytic efficiencies for EPO and ECO have increased by 17.12 fold and 15.09 fold, respectively. The F196I and L273I mutations seem to enlarge the opening of the active site, increasing the access of larger substrates.

R2D4 contains F119Y, F196I and L273I mutations. Compared to R2C2, R2D4 has one extra mutation at F119. The inclusion of F119Y further increased the k_{cat} towards EPO to 45.33 fold. The k_{cat} towards MCO and ECO also saw improvements. The catalytic efficiencies towards EPO and ECO improved 40.61 fold and 26.83 fold, respectively. The combination of three mutations causes a bigger opening than R2C2, possibly facilitating even better substrate access.

R2A2 contains F119Y, F196I and L273V mutations. The k_{cat} towards MPO, EPS, EPO and ECO has increased 1.56, 1.82, 30.33 and 9.33 folds, respectively. The

catalytic efficiencies of EPS, EPO and ECO have increased 9.35, 34.53 and 6.91 folds, respectively.

It is worth noting that for all three variants, the stacking of single mutations with modest improvement towards EPO (1.56-4.2 fold) can result in much better improvement against EPO (17.12-40.61 fold). **Figure 5.28** shows the change in the substrate-binding pocket caused by the mutations and **Table 5.14** summarises the kinetic parameters for R2C2, R2D4 and R2A2.

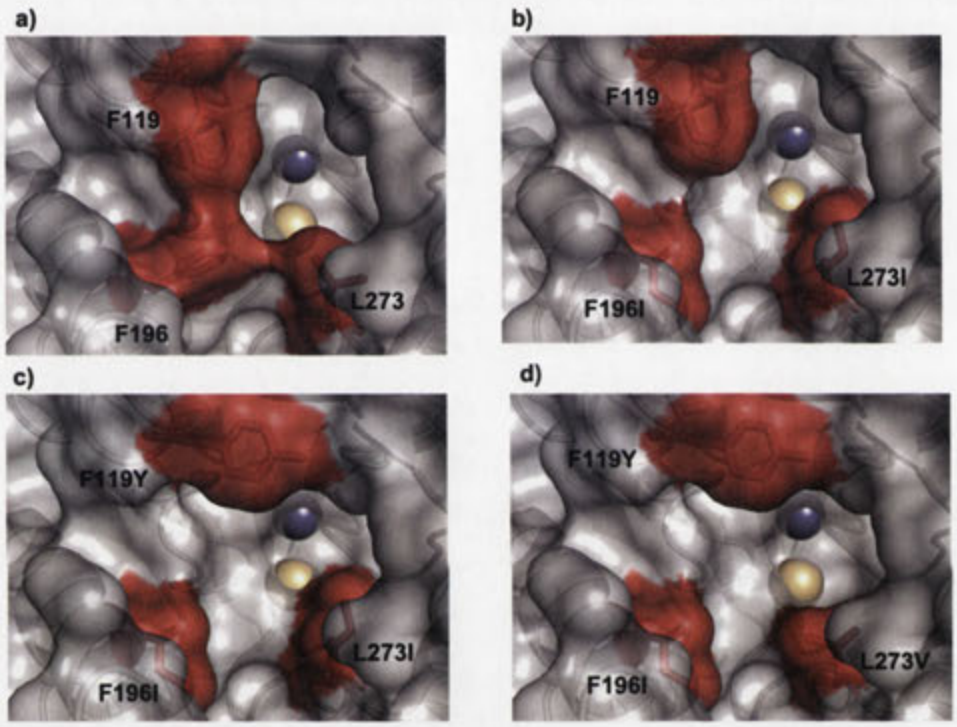


Figure 5.28: The active site pocket of (a) WT MPH, (b) R2C2, (c) R2D4 and (d) R2A2.
The shape of the active site cavity entrance is altered by mutations in R2C2, R2D4 and R2A2 as compared to WT MPH. The residues at positions 119, 196 and 273 are coloured red to highlight the changes. The grey surfaces represent the solvent excluded surfaces of WT MPH in (a), R2C2 in (b), R2D4 in (c) and R2A2 in (d) respectively. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal. The mutations were modelled with SWISS-MODEL using subunit B of WT MPH as template.

Table 5.14: Kinetic parameters for R2C2, R2D4 and R2A2 on various OPs.
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for R2C2, R2D4 and R2A2 with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value.

Variant	Substrate	k_{cat} (s^{-1})			Relative k_{cat}	K_m (μM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
R2C2	MPS	1.0	±	0.03	0.04	7.9	±	1.2	0.29	126.6	0.13
	MPO	3.2	±	0.40	0.71	139.2	±	28.6	1.93	23.0	0.37
	EPS	ND			ND	ND			ND	ND	ND
	EPO	3.3	±	0.10	11.00	31.6	±	4.2	0.64	104.4	17.12
	MCO	9.8	±	0.40	1.42	55.9	±	7.6	2.61	175.3	0.54
	ECO	4.2	±	0.10	14.00	14.5	±	1.6	0.93	289.7	15.09
R2D4	MPS	1.8	±	0.10	0.07	12.0	±	2.1	0.44	150.0	0.15
	MPO	4.4	±	0.50	0.98	66.0	±	15.2	0.91	66.7	1.07
	EPS	ND			ND	ND			ND	ND	ND
	EPO	13.6	±	0.90	45.33	54.9	±	9.3	1.12	247.7	40.61
	MCO	10.4	±	0.10	1.51	57.1	±	1.7	2.67	182.1	0.56
	ECO	10.2	±	0.10	34.00	19.8	±	1.5	1.27	515.2	26.83
R2A2	MPS	2.3	±	0.30	0.08	27.8	±	7.4	1.02	82.7	0.08
	MPO	7.0	±	0.60	1.56	187.5	±	28.2	2.60	37.3	0.60
	EPS	2.0	±	0.02	1.82	2.1	±	0.3	0.19	952.4	9.35
	EPO	9.1	±	0.30	30.33	43.2	±	4.1	0.88	210.6	34.53
	MCO	3.9	±	0.10	0.57	70.6	±	3.3	3.30	55.2	0.17
	ECO	2.8	±	0.10	9.33	21.1	±	4.3	1.35	132.7	6.91

5.7.3.4. R2D2 and R2F3

R2D2 and R2F3 both contain F119Y and L258H mutations. R2F3 contains an extra mutation of F196I.

R2D2 seemed to be a decent all-rounder with k_{cat} improvements for all substrates except MPS. The k_{cat} for MPO, EPO, MCO and ECO have improved 5.47, 2.09, 37.33, 17.07 and 105.33 folds, respectively. While the k_{cat} for MPS has decreased, the overall catalytic efficiency is similar to WT MPH due to decreased K_m. The catalytic efficiencies for MPO, EPO, MCO and ECO have increased 8.87, 50.03, 5.45 and 44.36 folds, respectively. The catalytic efficiency of EPS remained similar to WT due to an increase in K_m despite the improvement in k_{cat}. Of the six substrates examined, five of the catalytic efficiencies obtained were at least 1 x 10^5 $s^{-1}M^{-1}$ and two out of the five catalytic efficiencies were at least 1 x 10^6 $s^{-1}M^{-1}$. These are significant improvements from WT MPH that has only one substrate where its catalytic efficiency is 1 x 10^6 $s^{-1}M^{-1}$. The R2D2 reaction progression curve shows a stark difference between WT MPH and the variant. At higher substrate concentrations, R2D2 reaches stable catalytic rate slower than WT MPH. **(Figure 5.29)**. As the substrate

concentration increases, R2D2 requires longer time to reach stable catalytic rate **(Figure 5.29b)**. WT MPH does not display such behaviour **(Figure 5.29a)**.

This behaviour was observed for MPO and EPO. R2D2 did not show such behaviour against MPS and EPS, probably due to lower substrate concentrations used to determine the kinetic parameters against those substrates. Higher MPS and EPS concentrations would result in the substrates precipitating out of solution due to their hydrophobic nature. R2D2 exhibited such behaviour to a lesser degree against MCO and ECO when very high concentration was used (379.8 μM). The behaviour was only observed in R2D2 (F119Y, L258H) and was not observed in F119Y mutant, L258H mutant and R2F3 (F119Y, F196I, L258H). A similar characteristic was observed in yeast pyruvate decarboxylase, albeit for different reasons. Pyruvate decarboxylase has allosteric properties as part of its regulatory mechanism.[20] This creates a pyruvate concentration threshold such that the catalytic rate is low if the substrate concentration falls below the threshold. Stevenson *et al.*, in their protein evolution work, found a mutant that abolishes such allosteric property. They attributed the observation to reduced sensitivity to phosphate and the ease of substrate activation due to the mutations.[21] What we found in R2D2 is the opposite of what Stevenson *et al.* observed. The reason for R2D2 displaying such character is still unclear at this point and requires further investigation.

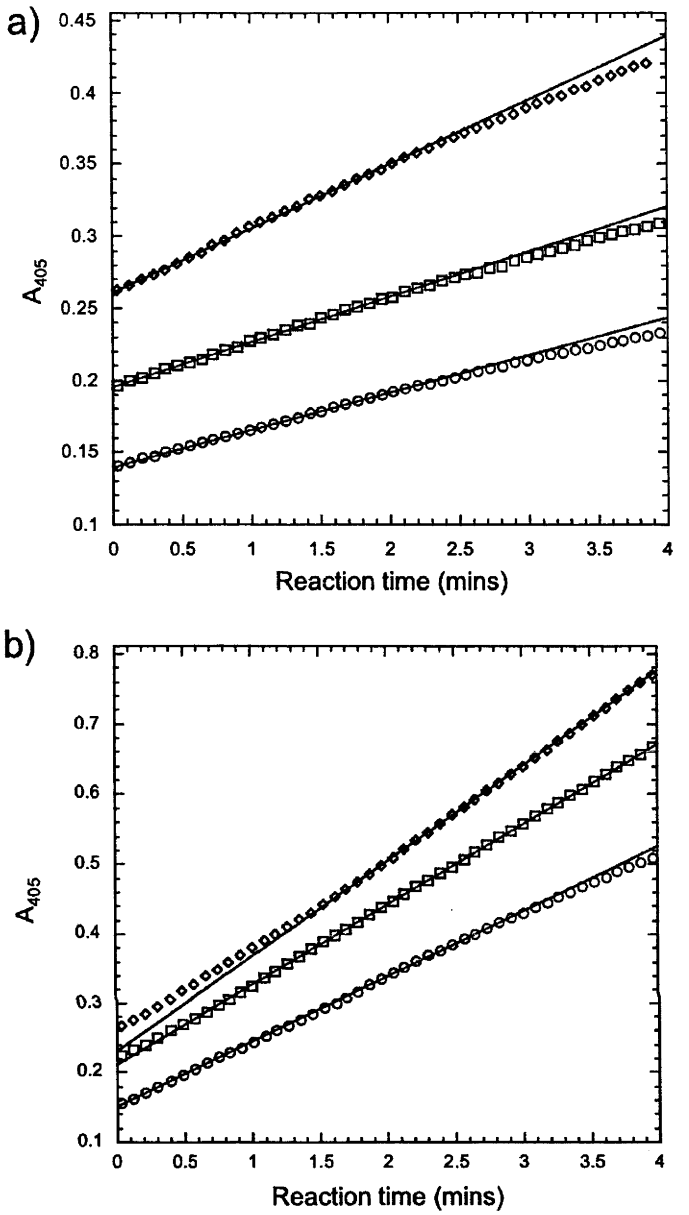


Figure 5.29: The reaction progression curve for (a) WT MPH and (b) R2D2 with MPO at pH 7.6 and 30°C.

The concentrations of MPO used were 60.4 μM (○), 119.6 μM (□) and 189.9 μM (◇). The regression line represents the stable-state reaction.

R2F3 contains F119Y, F196I and L258H mutations. R2F3 is the best mutant against EPO selected from the evolution. R2F3 showed 73.33-fold k_{cat} improvement and 97.47-fold catalytic efficiency improvement against EPO. The k_{cat}/K_m for R2F3 was 5.94 x 10^5 $s^{-1}M^{-1}$, up from 6.1 x 10^3 $s^{-1}M^{-1}$ for WT MPH. R2F3 also showed 119.33-fold k_{cat} improvement and 83.99-fold catalytic efficiency improvement against ECO, thus making R2F3 efficient at hydrolysing diethyl substituted oxon substrates. The improvement against EPO was consistent with other variants that carry the L258H mutation. L258H is the best beneficial single mutation found against EPO and almost

all variants carrying the L258H mutation showed large improvement in EPO-hydrolysing activity. R2F3 also displayed slightly improved catalytic efficiencies against MPO and MCO at 1.76 and 1.27 folds, respectively.

Figure 5.30 shows the difference in substrate binding pockets in R2D2 and R2F3 with WT MPH. **Table 5.15** summarizes the kinetic parameters of R2D2 and R2F3 against various OPs.

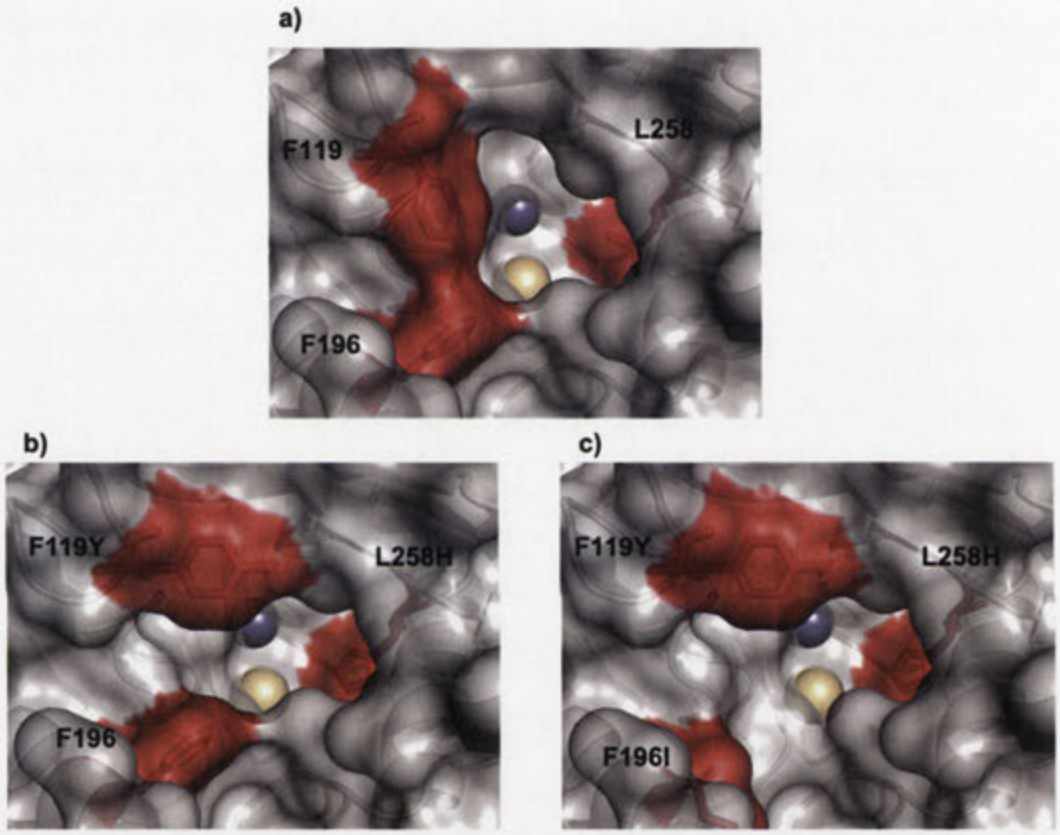


Figure 5.30: The substrate-binding pocket comparison between (a) WT MPH, (b) R2D2 and (c) R2F3.
The shape of the active site cavity is altered by mutations in R2D2 and R2F3. The residues at positions 119, 196 and 258 are coloured to highlight the changes.The grey surface represents the solvent excluded surface for WT MPH in (a), R2D2 in (b) and R2F3 in (c), respectively. The grey sphere represents Zn metal and the wheat sphere represents the Cd metal. The mutations were modelled with SWISS-MODEL using subunit B of WT MPH as template.

Table 5.15: Kinetic parameters for R2D2 and R2F3 on various OPs
The relative k_{cat}, K_m and k_{cat}/K_m were obtained by dividing the values obtained for R2D2 and R2F3 with that of WT MPH for each of the substrate. The relative k_{cat}, K_m and k_{cat}/K_m are colour-coded for easy reference. Green: increase in value, pink: decrease in value

Variant	Substrate	k_{cat} (s^{-1})	Relative k_{cat}	K_m (μM)	Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
R2D2	MPS	13.9 ± 0.40	0.51	10.1 ± 1.2	0.37	1376.2	1.37
	MPO	24.6 ± 1.30	5.47	44.5 ± 7.5	0.62	552.8	8.87
	EPS	2.3 ± 0.20	2.09	25.4 ± 6.9	2.35	90.6	0.89
	EPO	11.2 ± 0.30	37.33	36.7 ± 3.0	0.75	305.2	50.03
	MCO	117.8 ± 6.60	17.07	67.0 ± 11.4	3.13	1758.2	5.45
	ECO	31.6 ± 1.30	105.33	37.1 ± 5.9	2.38	851.8	44.36
R2F3	MPS	1.5 ± 0.10	0.05	50.4 ± 3.4	1.85	29.8	0.03
	MPO	6.3 ± 0.40	1.40	57.6 ± 8.9	0.80	109.4	1.76
	EPS	ND	ND	ND	ND	ND	ND
	EPO	22.0 ± 2.00	73.33	37.0 ± 10.5	0.76	594.6	97.47
	MCO	42.4 ± 2.90	6.14	103.9 ± 18.1	4.86	408.1	1.27
	ECO	35.8 ± 0.50	119.33	22.2 ± 1.6	1.42	1612.6	83.99

5.8. Mechanistic implications of mutations to residues surrounding the active site

The site saturation mutagenesis study highlights the "plasticity" of MPH active site. In this study, nine active site lining residues were probed for flexibility in altering substrate specificity, as well as an attempt to understand the substrate binding mode of the enzyme. Of the nine residues, four residues demonstrated flexibility to different degrees when altered for improved activity against ethyl paraxon. While some residues such as R72, F119 and W179 seemed intolerant to mutations **(Figures 5.9b, c and e)**, the deterioration in activity might stem from structural, substrate-binding or catalytic reasons. It is also worth noting that while F119 was intolerant to most mutations introduced by the NDT codon randomisation, F119Y was found to improve the k_{cat} for most substrate tested, albeit with an increase in the K_m for all substrates tested **(Table 5.8)**. This could be due to the structural similarities of phenylalanine and tyrosine, which led us to believe that the aromatic ring is essential in substrate binding.

While there are two phenylalanine residues in MPH active site, we believe F119 is the more essential phenylalanine residue based on three observations – the activity distributions of F119 and F196 mutant libraries, the number of improved variants selected from F119 and F196 libraries and the kinetic parameters of the improved mutants on various substrates when compared to wild type. As mentioned in **section 5.5.3,** F119 and F196 exhibited different mutational tolerance based on the activity

distribution plots of the mutant libraries for the two sites on EPO hydrolysing capability. From the plots **(Figures 5.9c and f)**, 48% of F196 mutants retain less than 50% relative activity, compared to 73% of F119 mutants retaining the same amount of activity. Amino acid changes at F196 are more tolerant compared to F119. Of the improved mutants selected from F119 and F196 libraries, F119Y is the only improved mutant found from the F119 library whereas F196Y, F196L and F196I were found from the F196 library. These suggest there are no other alternative residues at position 119 and aromatic residue at this position is necessary for activity. In contrast, F196L and F196I are hydrophobic alkyl residues suggesting leucine and isoleucine can alternative residues at this position. When analysing the kinetic parameters of WT MPH, F119Y, F196Y, F196L and F196I on various OP substrates, it was observed that the substrate profiles of F119Y and F196Y are largely similar to WT MPH with only minor differences. On the other hand, F196I and F196L showed decrease in catalytic efficiencies against most substrates with the exception of EPO, the screening substrate. It is worth stressing here that EPO was a poor substrate for MPH and the catalytic efficiencies of F196I and F196L against EPO were still lower than that for MPS despite the improvement. However, this seems to suggest that while F196 is important for the high catalytic efficiency in hydrolysing MPS, its role is not as important in hydrolysing bulkier substrates such as EPO and ECO. This could be due to steric hindrance of the phenylalanine when larger substrates are involved and mutations to leucine and isoleucine enlarge the substrate pocket to accommodate the substrate.

The possible roles of P150 and W179 were discussed in **Section 5.5.4** and were attributed to structural roles in the enzyme. As discussed in the previous section, L273 is believed to be one of the residues to shape the active site pocket. The two mutations, L273I and L273V were found to enhance the catalytic efficiencies of oxon substrates, suggesting that the oxygen/sulphur group of the substrate is positioned nearby. The role of L258 is unclear but it is of great interest due to its flexibility and the catalytic efficiency enhancement seen from mutations at this site. Oxon substrates seem to benefit the most from polar mutations to this site, suggesting the oxygen/sulphur group of the substrate is also nearby.

During the course of the structural comparisons described in the previous section, it became apparent that there are structural differences in the active sites of the two subunits of MPH. While the overall superimposition of the two MPH subunits gave

good agreement – RMSD of 0.5322 for 288 residues, there were some conformational differences in some substrate binding residues **(Figure 5.31).** Analysis of the by residue Cα RMSD of the superimposed MPH subunits showed good agreement, with the exception of some regions in the enzyme showing a difference of more than one angstrom. These regions correspond to flexible regions in the protein, as indicated by their high B-factor values **(Figure 5.32).** The B-factors were extracted from the MPH PDB structure using StrucTools (http://helixweb.nih.gov/structbio/). In **Figure 5.32**, the side chain B-factors of some amino acid reduce to zero due to the absence of side chain (e.g. glycine).

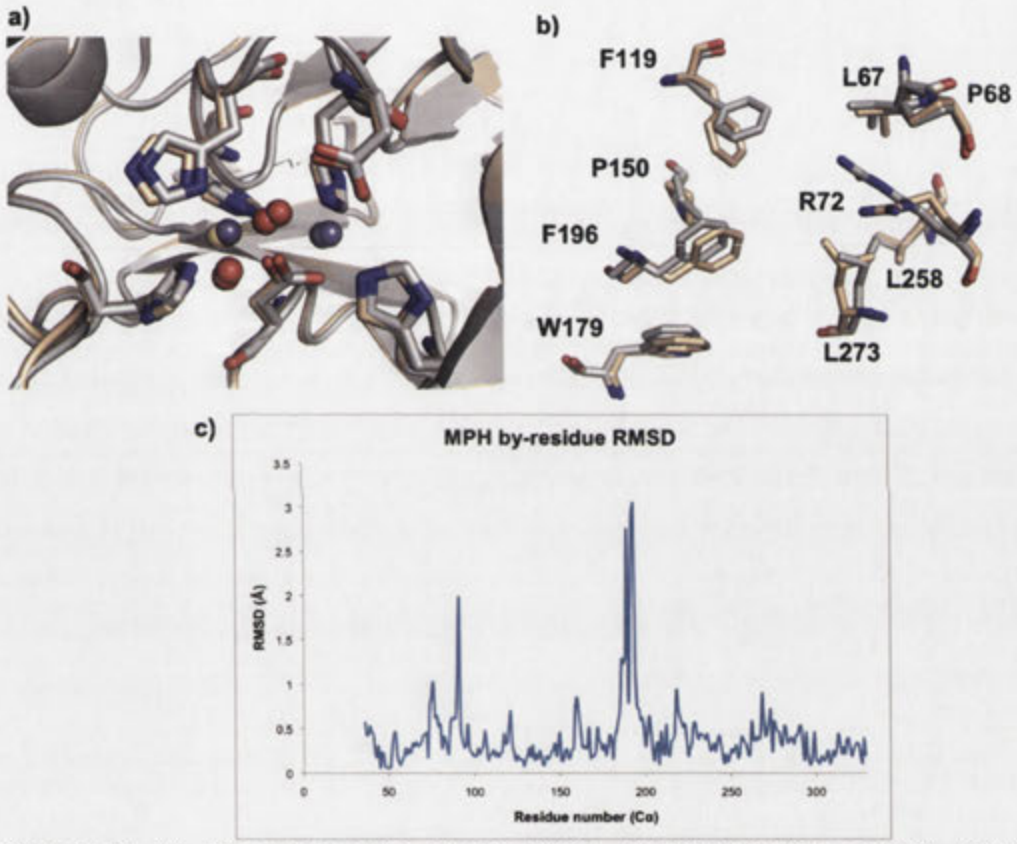


Figure 5.31: (a) The metal centre superimposition of different MPH subunits, (b) the superimpostion of active site lining residues and (c) the by-residue Cα RMSD plot of the different MPH subunits.

Subunit A of MPH is coloured grey, subunit B is coloured wheat, the Zn metals are coloured grey, Cd metal is coloured wheat and water molecules are coloured red.

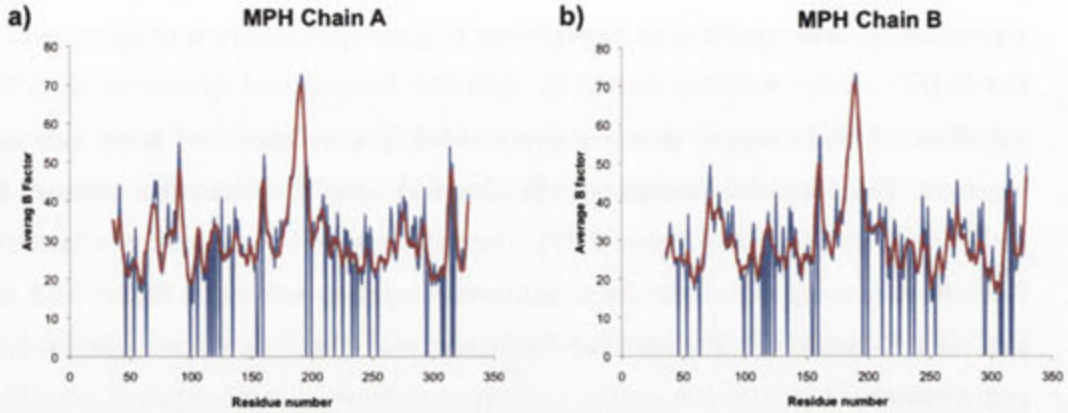


Figure 5.32: The average B-factor for different subunits of MPH.
(a) shows the average B factor for subunit A and (b) shows the average B factor for subunit B. In both plots, the B-factor of the main chain is coloured red, while the B-factor of the side chains is coloured blue.

The B-factor values and the Cα RMSD of active site lining residues for both subunits A and B are summarised in **Table 5.16.** The big difference in Cα RMSD values observed between the two subunits in W179 and F196 is probably due to the two residues being located near a highly flexible loop (residue 177 to residue 197). R72, F119 and L273 residues are located at slightly flexible regions on the enzyme. It is worth noting that the Cα RMSD values for R72 and F119 were not large despite having different conformations. While the exact reason for the flexibility of those residues is not known, the different conformations of some residues, particularly R72, F119, F196 and L273, result in different active site conformations between the two MPH subunits **(Figure 5.33).** This difference might have mechanistic repercussion, since different active site conformations will result in different hypotheses for substrate binding.

Table 5.16: The average B-factor values and Cα RMSD of MPH subunit
The B-factor values were extracted from MPH PDB structure (PDB ID: 1P9E) and the B-factor values for the overall structures were obtained by calculating the average Cα B-factor of the respective subunits. The numbers in bracket for overall structure Cα B-factor represent the standard deviation of the data set.

Residues	Chain A B-Factor		Chain B B-Factor		Cα RMSD (Å)
	Side Chain	Main Chain	Side Chain	Main Chain	
L67	35.21	36.09	24.28	27.27	0.31
P68	38.47	38.87	29.03	30.94	0.31
R72	44.78	35.28	49.15	40.89	0.45
F119	35.35	32.95	36.88	34.27	0.34
P150	27.08	26.54	25.99	25.11	0.45
W179	34.15	35.1	34.21	37.39	0.43
F196	48.04	43.32	42.02	41.17	0.72
L258	25.51	25.19	27.98	26.58	0.34
L273	35.09	34.7	39.35	37.03	0.71
Structure average		30.14 (8.74)		29.89 (8.90)	0.53

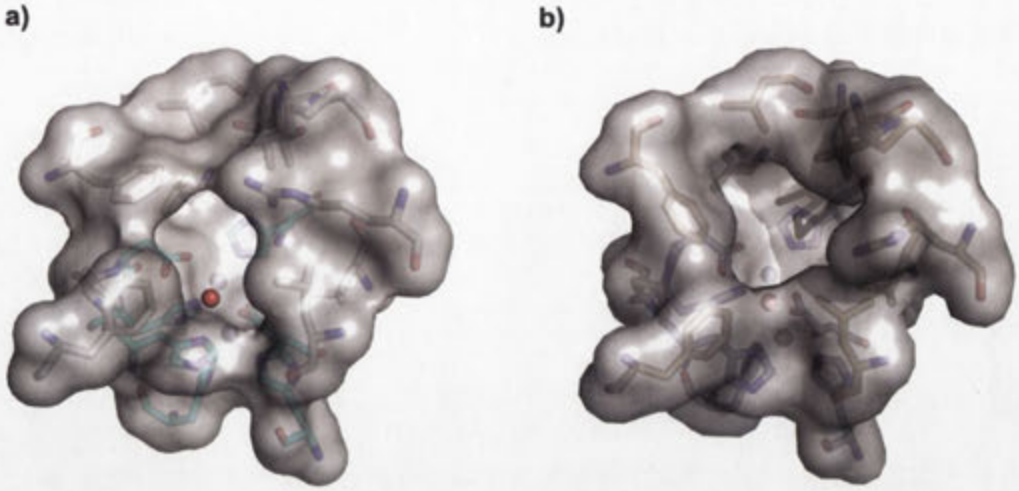


Figure 5.33: The different active site conformations of the two MPH monomeric subunits.
(a) shows the active site lining residues for subunit A and (b) shows the active site lining residues for subunit B. In both figures, the grey surface represents the solvent excluded surface of the active site.

Computer simulation of substrate docking was carried out with SwissDock[22,23] and AutoDock Vina[24]. During the course of the study on both subunits of MPH structure, it was concluded that the active site conformation observed in subunit B will allow substrate binding while access to the A subunit was restricted, suggesting that the conformation observed in subunit A obstructs the substrate from entering the active site. Methyl parathion was successfully docked into the subunit B of MPH with SwissDock and Autodock Vina, and a potential substrate-binding model from each method was shortlisted **(Figure 5.34)**. The terminally bound water in the active site was removed from the structure before virtual docking experiments were performed to prevent it from obstructing the docking procedure.

The two models selected are similar to each other, with some minor differences. The two models were selected based on the current understanding of MPH catalytic mechanism, that was derived from how substrate docked into OPDA (PDB ID: 2R1N)[25] **(Figure 5.35)**, as well as observations obtained from the mutagenesis studies of the active site lining residues. In both models, the sulphur group of the substrate is located near the β-metal, L273 and L258, displacing the β-metal terminally bound water. This is consistent with the proposed catalytic mechanism for MPH and OPH.[3,26] The *p*-nitrophenolate ring of the substrate is located near F119, where π-stacking is possible with the more essential phenylalanine residue. Active site close-up of substrate docking models obtained from Swiss-Dock and AutoDock Vina, as well as OPDA with a slow

substrate, are given in **Figure 5.36** to show the interaction between the substrate and the metal centre.

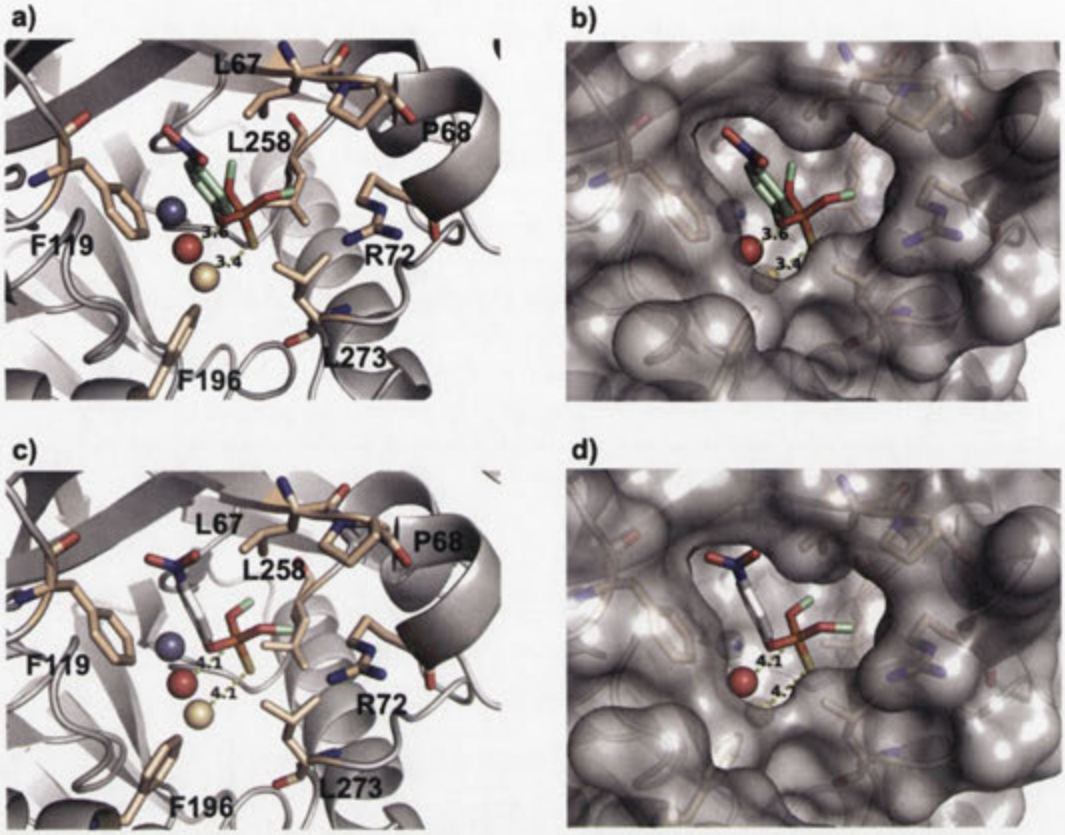


Figure 5.34: The substrate-binding models of MPS obtained from virtual docking with SwissDock and AutoDock Vina.
(a) and (b) show the MPS substrate-binding model of MPH subunit B obtained from virtual docking with SwissDock, while (c) and (d) show the model obtained from virtual docking with AutoDock Vina. (a) and (c) show the position of the active site lining residues and the substrate while (b) and (d) show the same structure with solvent excluded surface displayed to show how the substrate sits in the active site pocket. The numbers and the dashed lines indicate the distances of the phosphate centre and the sulphur group from the bridging water and β-metal respectively. The red sphere represents the bridging water; grey sphere represents the Zn metal and yellow sphere represents the Cd metal.

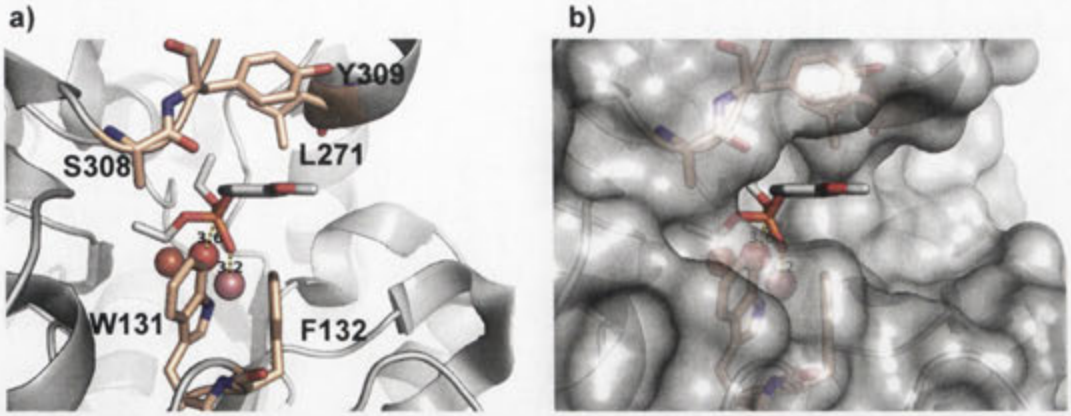


Figure 5.35: The substrate-binding of diethyl 4-methyoxyphenyl phosphate in OPDA.
(a) and (b) show how diethyl 4-methoxyphenyl phosphate sits in the active site of OPDA. (a) show the position of the active site lining residues and the substrate, while (b) shows the same structure with solvent excluded surface displayed to show how the substrate sits in the active site pocket. The number and the dashed lines indicate the distances of the phosphate centre and the phosphoryl oxygen from the bridging water and β-metal, respectively. The orange sphere represents Fe^{2+}; the pink sphere represents Mn^{2+} and the red sphere represents the bridging water.

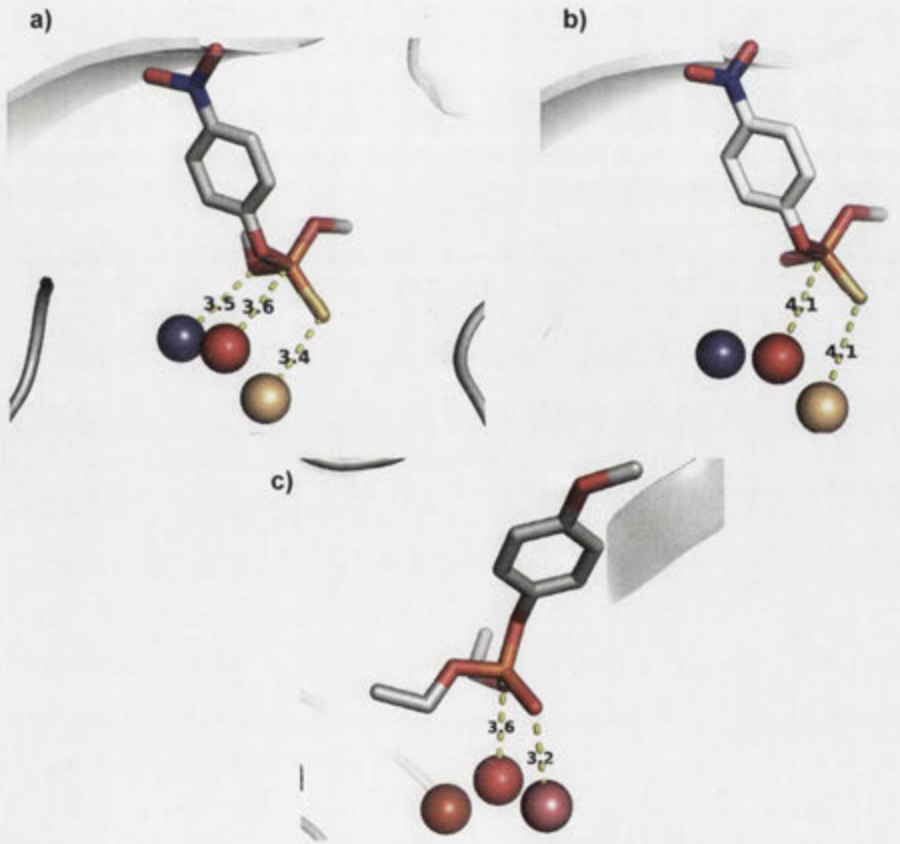


Figure 5.36: Active site close up of (a) MPS docked MPH simulated by Swiss-Dock, (b) MPS docked MPH simulated by AutoDock Vina and (c) OPDA soaked with diethyl 4-methyoxyphenyl phosphate.
The substrates in (a) and (b) are MPS while the substrate in (c) is diethyl methoxyphenyl phosphate. The red spheres represent the bridging hydroxides in all three structures; the grey spheres represent Zn^{2+} ions, the wheat spheres represent Cd^{2+} ions, the orange sphere represents Fe^{2+} ion and the pink sphere represents Mn^{2+} ion. The numbers and dashed lines indicate the distances of the substrate to the metal centre.

5.9. Summary

This chapter describes the site saturation mutagenesis on MPH active site lining residues and subsequently the directed evolution of MPH for improved EPO hydrolysis. To my best knowledge, the work described in this chapter is the most extensive substrate profile characterisation on MPH to date. Single site mutants F119Y, F196Y, F196I, F196L, L258S, L258N, L258I, L258H, L273I and L273V were the improved variants found in EPO hydrolysis. The substrate profiles of these mutants were characterised with various OP substrates in an attempt to understand the roles of these residues in substrate interaction. All of the selected single site mutants have modest improvements on EPO hydrolysis. Interestingly, unexpected but admirable improvements were seen for L258 mutants on hydrolysis of non-selected substrates (MCO and ECO). By coupling the kinetics data and the mutant activity distribution of single site libraries, insights into the possible roles of some residues were suggested. Two rounds of directed evolution aimed to further enhance EPO hydrolysing efficiency were conducted with DNA shuffling, and numerous multiple site mutants with better EPO activity were obtained. Among the multiple site mutants characterised, four interesting multiple site mutants will be singled out for special mention. R1A6 (F119Y, L258S) was found to be a TCPy leaving group substrate specialist, with 46 and 267 folds catalytic efficiency improvement on MPO and EPO. R2F3 (F119Y, F196I, L258H) was found to be a diethyl substituted oxon specialist, with 97 and 84 folds catalytic efficiency improvement on EPO and ECO. R2D4 (F119Y, F196I, L273I) is an example of moderately beneficial mutations stack, which had significant improvement on a selected substrate. This mutant had 41-fold catalytic efficiency improvement towards EPO. The final interesting mutant is R2D2 (F119Y, L258H), a generalist. Five of the six substrates characterised see improvement in overall catalytic efficiencies. The substrate specificity of R2D2 is vastly broadened compared to WT MPH, as five of the substrates show catalytic efficiencies of at least 10^5 $s^{-1}M^{-1}$. The results we observe here highlight the "plasticity" of MPH active site – a few amino acid changes in the active site pocket result in drastic changes in OP specificity. Therefore, it is no surprise that various MPH variants with high sequence identity have been isolated from different microorganisms. Based on the current understanding of MPH mechanism, as well as the data we obtained in this study, potential substrate binding models obtained from Swiss-Dock and AutoDock Vina were selected.

5.10. References

1. Neuwald, A. F., Liu, J. S., Lipman, D. J. & Lawrence, C. E. Extracting protein alignment models from the sequence database. *Nucleic Acids Research* **25,** 1665–1677 (1997).
2. Daiyasu, H., Osaka, K., Ishino, Y. & Toh, H. Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Letters* **503,** 1–6 (2001).
3. Dong, Y.-J. *et al.* Crystal structure of methyl parathion hydrolase from *Pseudomonas sp.* WBC-3. *Journal of Molecular Biology* **353,** 655–663 (2005).
4. Leiros, H.-K. S. *et al.* Crystal structure of the mobile metallo-β-lactamase AIM-1 from Pseudomonas aeruginosa : insights into antibiotic binding and the role of Gln 157. *Antimicrobial Agents and Chemotherapy* **56,** 4341–53 (2012).
5. Carfi, A., Duée, E., Galleni, M., Frère, J. M. & Dideberg, O. 1.85 Å resolution structure of the zincII β-lactamase from *Bacillus cereus*. *Acta Crystallographica Section D Biological Crystallography* **D54,** 313–323 (1998).
6. Garau, G. *et al.* A metallo-β-lactamase enzyme in action: crystal structures of the monozinc carbapenemase CphA and its complex with biapenem. *Journal of Molecular Biology* **345,** 785–795 (2005).
7. Ullah, J. H. *et al.* The crystal structure of the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* at 1.7 A resolution. *Journal of Molecular Biology* **284,** 125–136 (1998).
8. Cameron, a D., Ridderström, M., Olin, B. & Mannervik, B. Crystal structure of human glyoxalase II and its complex with a glutathione thiolester substrate analogue. *Structure* **7,** 1067–1078 (1999).
9. Campos-Bermudez, V. a *et al.* Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: new insights into metal ion selectivity. *Biochemistry* **46,** 11069–11079 (2007).
10. Kim, M. H. *et al.* The molecular structure and catalytic mechanism of a quorum-quenching *N*-acyl-L-homoserine lactone hydrolase. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 17606–17611 (2005).
11. Kostelecky, B., Pohl, E., Vogel, A., Schilling, O. & Meyer-klaucke, W. The crystal structure of the zinc phosphodiesterase from *Escherichia coli* provides insight into function and cooperativity of tRNase Z-family proteins. *Journal of Bacteriology* **188,** 1607–1614 (2006).
12. Vogel, A., Schilling, O., Niecke, M., Bettmer, J. & Meyer-Klaucke, W. *ElaC* encodes a novel binuclear zinc phosphodiesterase. *Journal of Biological Chemistry* **277,** 29078–29085 (2002).
13. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* **38,** W529–W533 (2010).
14. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *Chembiochem* **9,** 1797–1804 (2008).
15. Zhao, H. & Zha, W. *In vitro* "sexual" evolution through the PCR-based staggered extension process (StEP). *Nature Protocols* **1,** 1865–1871 (2006).
16. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).
17. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* **31,** 3381–3385 (2003).

18. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).
19. Jeong, Y.-S. *et al.* High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes. *Protein Engineering, Design & Selection* **25,** 725–731 (2012).
20. Boiteux, A. & Hess, B. Allosteric properties of yeast pyruvate decarboxylase. *FEBS letters* **9,** 293–296 (1970).
21. Stevenson, B. J., Liu, J.-W. & Ollis, D. L. Directed evolution of yeast pyruvate decarboxylase 1 for attenuated regulation and increased stability. *Biochemistry* **47,** 3013–3025 (2008).
22. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* **39,** W270–W277 (2011).
23. Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *Journal of Compuational Chemistry* **32,** 2149–2159 (2011).
24. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Compuational Chemistry* **31,** 455–461 (2010).
25. Jackson, C. J. *et al.* In crystallo capture of a Michaelis complex and product-binding modes of a bacterial phosphotriesterase. *Journal of Molecular Biology* **375,** 1189–1196 (2008).
26. Benning, M. M., Shim, H., Raushel, F. M. & Holden, H. M. High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry* **40,** 2712–2722 (2001).

FUTURE DIRECTIONS

6. Future directions

6.1. MPH proteolytic degradation

A protocol that allowed rapid purification of MPH was developed so that proteolytic degradation can be minimised. While it was a viable protocol to produce MPH for the purpose of characterisation, a more direct solution to address the problem is required, if MPH is to be used for bioremediation and medicinal applications. Site directed mutagenesis on the predicted cleavage site was performed but it was not effective in eliminating the degradation. As discussed previously in **Chapter 3**, there is a possibility that the mutation was not performed on the key residues for protease recognition. One possible approach is to perform site directed mutagenesis on multiple sites around the proteolytic cleavage region to eliminate the protease recognition site. However, since autoproteolysis was ruled out as the source of degradation, identifying a suitable protease inhibitor for use during cell lysis and protein purification might be another possible solution. Serine inhibitors, such as PMSF and AEBSF, were not effective in stopping the degradation, indicating that the protease responsible was not a serine protease.

Since the enzyme purified with immobilised-metal affinity chromatography (IMAC) using Ni^{2+} charged column gave good purity, the possibility of using this purification protocol in the future should be explored. However, efforts should be made to investigate the effects of his-tag on the dimeric assembly, as well as the metal content of MPH before employing IMAC for routine MPH purification. An ubiquitinase cleavable His-Tag construct, such as pHUE[1] vector can be considered to facilitate the fusion tag removal. By removing the His-Tag after purification, unwanted fusion tag associated effects can be eliminated.

6.2. Directed evolution of MPH for enhanced thermal stability

Ideas to improve the thermostability evolution were given in **section 4.4.5** and will not be discussed in this section. As mentioned earlier, stability is a desired trait for any enzyme that is to be used for practical applications. While the work done in evolving MPH to be more thermostable has met with little success, it gave some

indications on the effects of Zn^{2+} on the protein stability. At this point, little is known about the metal binding properties of MPH or how different metals affect the stability. Miss Nur Hafizah Azizan is investigating this front by performing site directed mutagenesis on primary and secondary shells metal-binding ligands, and investigate the associated changes in metal binding affinity. Rapid activity independent stability assay, such as Hot-CoFi[2] agar screening method, developed by Asial and co-workers should be employed to directly probe the thermostability changes during library screen. This method will also eliminate the laborious nature of the thermostability screening employed in this work.

6.3. Altering the substrate specificity of methyl parathion hydrolase

Mutants with moderate improvement in ethyl paraoxon (EPO) activity have been isolated from site saturation mutagenesis of the active site lining residues. Some of these mutants exhibited altered substrate specificity compared to the wild type variant. Shuffling of these mutations produced multiple site mutants that gave respectable improvements in EPO activity and unexpectedly, improvements in MCO and ECO activities. The evolution of MPH should continue to isolate MPH variants with quadruple mutations. The 10 single site mutants isolated consist of mutations from F119, F196, L258 and L273. The multiple site mutants isolated after two rounds of StEP shuffling consist of predominantly double-site and triple-site mutants. We expect that the quadruple mutants isolated from further shuffling will give further enhancement to what we already achieved in this work.

The anomalous reaction progression for R2D2 at high substrate concentrations should be further investigated to identify the underlying reason. R2D2 is an interesting "generalist" mutant that can hydrolyse many OPs efficiently and therefore should be further studied. Other characteristics, such as metal content and thermostability of the isolated mutants, should also be studied to ensure that the activity enhancement does not come at the cost of stability.

Crystal structures of some interesting mutants should be obtained to better understand the structural changes introduced by the mutations. The possible effects of various mutations discussed in this work were based on modelled structure produced

with SWISS-MODEL. While the trend of altered substrate specificity for some mutations are clear, e.g., L258S and L258N shifted the specificity towards chlorpyrifos based OPs, it was not immediately clear how those mutations resulted in such changes. Obtaining the crystal structures of some mutants will give some insights as to how these mutations change the active site pocket of MPH and result in altered substrate specificity. Co-crystallisation of MPH with a slow substrate or substrate mimic should also be attempted to verify the substrate-binding models obtained with Swiss-Dock and AutoDock Vina. One potential slow substrate that can be used for such study is coumaphos oxon. MPH has very low activity against coumaphos oxon. The different conformations observed in R72 and F119 between the monomeric subunits of published MPH structure should be further investigated to ensure that the difference observed was not due to crystallisation artefact. It is important to obtain such information since the difference in conformation will result in different active site shapes, and thus the substrate-binding model and mechanism.

From a practical application point of view, the substrate profiles of some mutants isolated in this work are broad and efficient enough for bioremediation purposes. The substrate profile can be further enhanced for bioremediation by formulating mixtures of MPH mutants with different substrate profiles, or mixtures of MPH and OPDA.

6.4. References

1. Catanzariti, A., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science* **13,** 1331–1339 (2004).
2. Asial, I. *et al.* Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nature Communications* **4,** 2901 (2013).

APPENDICES

Appendix A: Recipes

Bromophenol Blue Loading Dye (for DNA agarose electrophoresis)
The solution contained 30% (v/v) glycerol and 0.25% bromophenol blue in dH_2O.

LB (Lysogeny Broth)
Dissolve 10 g of tryptone (Difco), 5 g of Bacto yeast extract (Difco) and 10g of NaCl in 1000 mL of water and sterilise by autoclaving. Add 15 g/L of agar prior to autoclaving to make LB agar plates. Add antibiotics to final concentrations of 100 μg/mL for ampicilin (Amresco) and 50 μg/mL for kanamycin (AG Scientific) as require once the media had cooled to approximately 60°C.

Organophosphate Substrate Stock
18.99 mM of methyl parathion, methyl paraoxon, ethyl parathion, ethyl paraoxon, methyl chlropyrifos oxon or ethyl chlorpyrifos oxon was prepared by dissolving or diluting the organophosphates in 100% methanol.

SB Buffer (for DNA agarose electrophoresis)
20x stock solution was prepared by dissolving 8 g of NaOH in 0.9 L of dH_2O and adjusting the pH to 8.0 with boric acid and making the volume of 1 L with dH_2O.

SDS Running Buffer (for SDS-PAGE)
10x stock solution was prepared by dissolving 144 g glycine, 30 g Tris base and 10 g SDS in 1 L of dH_2O.

SDS-PAGE Stock Sample Buffer
The solution was prepared by mixing 120 μL 0.1% (w/v) bromophenol blue, 1.12 mL 80% (v/v) glycerol, 800 μL of 2 M Tris buffer at pH 6.8, 2 mL of 10% (w/v) SDS and 0.31 g of dithiothreitol. dH_2O was added to a final volume of 10 mL. 10 μL of β-mercapoethanol was added to every 190 μL of stock sample buffer immediately before use to obtain SDS-reducing sample buffer.

YENB (Yeast Extract Nutrient Broth)
Dissolve 7.5 g Bacto yeast extract (Difco) and 8 g Bacto Nutrient Broth (Difco) in 1000 mL water and sterilise with autoclaving.

Appendix B: Suppliers of Equipments, Consumables and Chemicals

Equipment and Consumable Suppliers

Equipments and consumables	Manufacturers and suppliers
5415 Microcentrifuge	Eppendorf
5804 Centrifuge	Eppendorf
AKTA FPLC System	G.E. Healthcare
Amicon Ultra-15 Centrifugal Filter Unit	Millipore
ASB270BT Autoclave	Astell Scientific
C1000 Thermal Cycler	Bio-Rad
CO8000 Cell Density Meter	WPA Biowave
French Pressure Cell Press	SLM Instrucments
GeneAmp PCR System 9700	Applied Biosystems
HisTrap FF Column	G.E. Healthcare
HiTrap DEAE FF Column	G.E. Healthcare
HiTrap Desalting Column	G.E. Healthcare
HiTrap SP FF Column	G.E. Healthcare
I-Cycler	Bio-Rad
MicroPulser Electroporator	Bio-Rad
MJ Mini Personal Thermal Cycler	Bio-Rad
NanoDrop® ND-1000 Spectrophotometer	NanoDrop
Orion ROSS Combination pH Electrode	Orion Pacific Pty Ltd
Q Sepharose Affinity Column	Amersham Biosciences
QIAprep Spin Miniprep kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
R20A2 and R9A Rotors for VX22G	Hitachi
Rotorfix 32 Centrifuge	Hettich
SE250 Mighty Small II Mini Vertical Electrophoresis Unit	Hoefer
SpectraMax® M2/M2e Microplate Reader	Molecular Devices
Superdex 200 Size Exclusion Column	Amersham Biosciences
UV transilluminator 312 nm with UVItec camera	Hanimax Statesman
Varian Cary UV/Vis Spectrophotometer	Varian
Veriti 96-Well Thermal Cycler	Applied Biosystems
VX22G High Speed Centrifuge	VWR
Wide Mini Sub Cell Electrophoresis Tank	Bio-Rad
Wizard® SV Gel and PCR Clean-Up System	Promega

Chemical Suppliers

Chemcals	Manufacturers and suppliers
1 kb DNA Marker	New England Biolabs
1,10-phenanthroline	Sigma Corporation
2-methyl-2,4-pentanediol	Sigma Corporation
2-propanol	Merck
2,6-pyridine dicarboxylate	Sigma Corporation
3-methyl-1-butanol	Sigma Corporation
3,5,6-trichloro-2-pyridinol	Sigma Corporation
4-(2-Aminoethyl) Benzenesulfonyl Fluoride	Sigma Corporation
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Amresco
8-hydroxyquionoline-5-sulfonic acid	Sigma
Acetone	Merck
Acetonitrile	Sigma Corporation
Acryl/Bis 37.5:1, 40% (w/v)	Amresco
Agar	Difco
Agarose	Sigma Corporation
Ammonium Persulfate	Sigma Corporation
Ampicilin	Amresco
*Ase*I and Buffer	New England Biolabs
BigDye® Terminator and Buffer	Biomolecular Resource Facility, JCSMR
BioTaq™ Polymerase	Bioline
bis-*p*-(nitrophenyl) Phosphate	Sigma Corporation
Bromophenol Blue	Sigma Corporation
BugBuster® Protein Extraction Reagent	Novagen
Calf Intestinal Alkaline Phosphatase and Buffer	New England Biolabs
Chelex	Bio-Rad
Dimethyl Sulfoxide	Sigma Corporation
Dithiothreitol	Sigma Corporation
dNTPs	Roche
*Dpn*I and Buffer	New England Biolabs
Dual Color Precision Plus Protein Standards	Bio-Rad
*Eco*RI and Buffer	New England Biolabs
EDTA	Ajax Chemicals
Ethanol	Merck
Ethidium Bromide	Sigma Corporation
Ethyl acetate	Merck
Ethyl Chlorpyrifos Oxon	Chemservice
Ethyl Parathion	Chemservice
Ethyl Paraxon	Sigma
Ethylene Glycol	Fluka
Gel Filtration Markers Kit 12,000 - 200,000 Da	Sigma Corporation
Glycerol	Merck
Glycine	Amresco
HCl	Ajax Chemicals
HNO_3	Merck
Imidazole	Sigma Corporation

Isopropyl β-D-1-thiogalactopyranoside	Astral
K_2HPO_4	Ajax Chemicals
Kanamycin	AG Scientific Inc
KH_2PO_4	Ajax Chemicals
Low Range SDS-PAGE Marker	Bio-Rad
Methanol	Merck
Methoxy Ethanol	Ajax Chemicals
Methyl Chlorpyrifos Oxon	Chemservice
Methyl Paraoxon	Sigma Corporation
Methyl Parathion	Chemservice
$MgCl_2$	Sigma Corporation
$MnCl_2$	Ajax Chemicals
N,N,N',N',-Tetramethylethylenediamine	Sigma Corporation
NaCl	Chem Supply
NaOH	Ajax Chemicals
Native *Pfu P*olymerase and Buffer	Thermo Scientific; Agilent Technologies
*Nde*I and Buffer	New England Biolabs
Nutrient Broth	Difco
p-nitrophenol	Sigma Corporation
p-nitrophenyl Acetate	Sigma Corporation
p-nitrophenyl Phophate	Sigma Corporation
Petroleum jelly	Shell
Phenylmethylsulfonyl fluoride	Sigma Corporation
Phusion Polymerase	Finnzyme; New England Biolabs
Piperazine	Sigma Corporation
Polyethylene Glycol 8000	Fluka
Propylene Glycol	Fluka
*Pst*I and Buffer	New England Biolabs
RedSafe™ DNA Stain	Chembio
rLysozyme™	Novagen
Sodium Acetate	ICN Biomedical
Sodium Dodecyl Sulfate	Amresco
T4 DNA Ligase	Thermo Scientific; Fermentas; New England Biolabs
Taq Polymerase	Roche
tert-butanol	Sigma Corporation
Tris Base	Amresco
Tryptone	Difco
Tween 20	Sigma Corporation
Yeast Extract	Difco
β-mercaptoethanol	Sigma Corporation

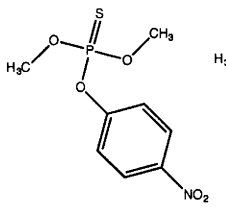
Appendix C: Oligonucleotide Primers

All oligos were supplied by GeneWorks or Integrated DNA Technologies

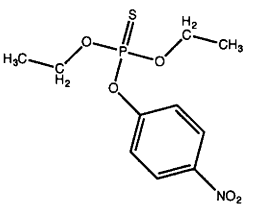
Name	Description	Sequence (5'-3')
(4)pCY76F	Forward sequencing primer for pCY76 and pJWL1030	GCC GGC GAT ATC GGA TCC G
F_SSM_Flank_M1T_Ase	Forward flanking primer for MPH without signal in pETMCSI and pCY76. Contains *Ase*I site	GCT TCT ATT AAT GGC AGC ACC GCA GGT T
F-M1T-pJWL1030flank-Ase	Forward flanking primer for MPH without signal in pJWL1030. Contains *Ase*I site	GAT ATA ATT AAT GGC AGC ACC GCA GGT T
F-MPH-NDTSSM-F119	Forward site saturation mutagenesis primer with NDT randomisation at F119	TGC AGC AGG TCT G**ND** **T**GG TCC GAC ACT
F-MPH-NDTSSM-L67P68R72	Forward site saturation mutagenesis primer with NDT randomisation at sites L67, P68 and R72 for MPH	ACC GTT GCA **NDT NDT** GTT GAT AAA **NDT** CTG AAT CAG
F-MPH-NDTSSM-P150	Forward site saturation mutagenesis primer with NDT randomisation at P150	ACC CAT ATG CAT **NDT** GAT CAT GTT GGT
F-MPH-SSM-R72	Forward site saturation mutagenesis primer with NNK randomisation at R72	CCG GTT GAT AAA **NNK** CTG AAT CAG CCT
F-MPH-SSM-W179	Forward site saturation mutagenesis primer with NNK randomisation at W179	GAA GCA GAT TTT **NNK** CTG AGC CAG ACC
F-SDM-M1T-F54A	Forward site directed mutagenesis primer at F88	CTG GCA AAA AGC **GCA** CAG AAA GCA CC
FSac2Sp2	Forward flanking primer for MPH without signal in pET47b. Contains *Sac*II site	ACC ATC ACT CCG CGG CTG CAG CAC CGC AGG TT
pCY76R(M13)	Reverse sequencing primer for pCY76 and pJWL1030	GTA AAA CGA CGG CCA GTG
pET3	Forward sequencing primer for pETMCSI	CGA CTC ACT ATA GGG AGA CCA CAA C
pET4	Reverse sequencing primer for pETMCSI	CCT TTC GGG CTT TGT TAG CAG
R_SSM_Flank_M1T_Eco	Reverse flanking primer for MPH without signal in pETMCSI and pCY76. Contains *Eco*RI site	TGC CAT GAA TTC TTA TTA TTT CGG ATT CAC
R-M1T-pJWL1030flank-Pst	Reverse flanking primer for MPH without signal in pJWL1030. Contains *Pst*I site	AAG CTT CTG CAG TTA TTA TTT CGG ATT CAC
R-MPH-NDTSSM-F196	Reverse site saturation mutagenesis primer with NDT randomisation at F196	GCA CCT TTA AA**A** **HNG** CCT TTG CTT TC
R-MPH-NDTSSM-L258	Reverse site saturation mutagenesis primer with NDT randomisation at L258	AAC GGC TGC AAC **AHN** AAT CAG ATC ACC
R-MPH-NDTSSM-L273	Reverse site saturation mutagenesis primer with NDT randomisation at L273	GCT ATC GCT ATC **AHN** CTG GGT GGT AAC
R-SDM-M1T-F54A	Reverse site directed mutagenesis primer at F88	GGT GCT TTC TGT GCG CTT TTT GCC AG
RSacSp	Reverse flanking primer for MPH without signal in pET47b. Contains *Sac*I site	ATT TAA GCG AGC TCT TAT TTC GGA TTC ACA ACG C
T7Fwd	Forward sequencing primer for pET47b	TAA TAC GAC TCA CTA TAG GG
T7Rev	Reverse sequencing primer for pET47b	GCT AGT TAT TGC TCA GCG G

*Bolded nucleotides are the sites subjected to mutagenesis.
*Restriction sites are underlined.

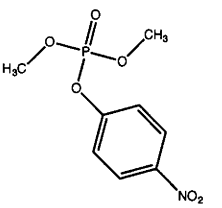
Appendix D: Organophosphates Described in This Thesis



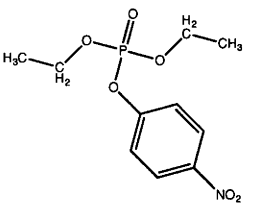
Methyl parathion (MPS)



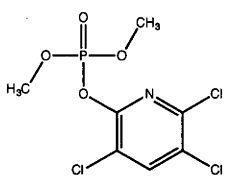
Ethyl parathion (EPS)



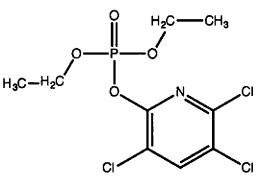
Methyl paraoxon (MPO)



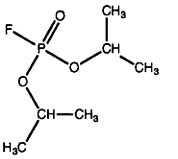
Ethyl paraoxon (EPO)



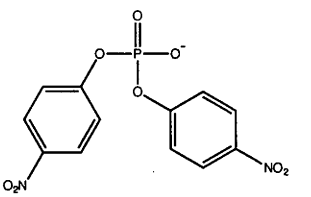
Methyl chlorpyrifos oxon (MCO)



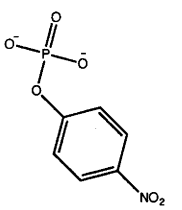
Ethyl chlorpyrifos oxon (ECO)



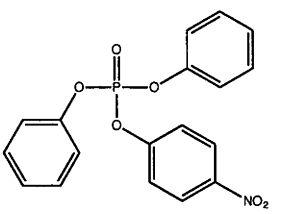
Diisopropyl fluorophosphate



bis-(*p*-nitrophenyl) phosphate (bpNPP)



p-nitrophenyl phosphate (pNPP)



p-nitrophenyl diphenylphosphate (pNPDPP)

Appendix E: MPH Homologues Sequence Alignments

Closely related MPH homologues

```
OCHMPH      MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
M6MPH       MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
WBC3MPH     MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
YC1MPH      MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
FDS1MPH     MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
BCMPH       MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
            ************************************************************

OCHMPH      SDGTVALPVDKRLNQPAPKTQSALAKYFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
M6MPH       SDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
WBC3MPH     SDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
YC1MPH      SDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
FDS1MPH     SDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
BCMPH       SDGTVALPVDKRLNQPAPKTQSALAKSFQKAPLETSVTGYLVNTGSKLVLVDTGAAGLFG 120
            ************************* ********************************

OCHMPH      PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQKEADFWL 180
M6MPH       PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQKEADFWL 180
WBC3MPH     PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQKEADFWL 180
YC1MPH      PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQKEADFWL 180
FDS1MPH     PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVGEQLAFPNAVVRADQKEADFWL 180
BCMPH       PTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMAGEQLAFPNAVVRADQKEADFWL 180
            ************************************.**********************

OCHMPH      SQTNLDKAPDD-SKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTAGHTT 239
M6MPH       SQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTT 240
WBC3MPH     SQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTT 240
YC1MPH      SQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTT 240
FDS1MPH     SQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTT 240
BCMPH       SQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPGHTT 240
            *********** ****************************************** ****

OCHMPH      YVVESQGQKLALLGDLILVAAVQFDDPSVTNQLDSDSKSAAVERKKAFADAAKGGYLIAA 299
M6MPH       YVVESQGQKLALLGDLILVAAVQFDDPSVTNQLDIDGKSAAVERKKAFADAAKGGYLIAA 300
WBC3MPH     YVVESQGQKLALLGDLILVAAVQFDDPSVTTQLDSDSKSVAVERKKAFADAAKGGYLIAA 300
YC1MPH      YVVESQGQKLALLGDLILVAAVQFDDPSVTTQLDSDSKSVAVERKKAFADAAKGGYLIAA 300
FDS1MPH     YVVESQGQKLALLGDLILVAAVQLDDPSVTIQLDSDSKSAAVERKKAFADAAKGGYLIAA 300
BCMPH       YVVESQGQKLALLGDLILVAAVQFDDPSVTIQLDSDSKSAAVERKKAFADAAKGGYLIAA 300
            ************************:****** *** *.**.********************

OCHMPH      AHLSFPGIGHIRAEGKGYRFVPVNYSVVNPK   330
M6MPH       SHLPFPGIGHIRAEGKGYRFVPVNYSVVNPK   331
WBC3MPH     SHLSFPGIGHIRAEGKGYRFVPVNYSVVNPK   331
YC1MPH      SHLSFPGIGHIRAEGKGYRFVPVNYSVVNPK   331
FDS1MPH     THLSFPGIGHIRAEGKGYRFVPVNYSVVNPK   331
BCMPH       THLSFPGIGHIRAEGKGYRFVPVNYSVVNPK   331
            :** ***************************
```

Legend:
OCHMPH – *Ochrobactrum* sp. M231
M6MPH – *Plesiomonas* sp. M6
WBC3MPH – *Pseudomonas* sp. WBC-3
YC1MPH – *Stenotrophomonas* sp. YC-1
FDS1MPH – *Bulkhoderia* sp. FDS-1
BCMPH – *Bulkhoderia cepacia*

Residues in blue are the metal-binding ligands and residues that are different from WBC3-MPH are coloured red.

Alignment between WBC3MPH and OPHC2

```
WBC3MPH     MPLKNRLLARLSCVAAVVAATAAVAPLTLVSTAHAAAPQVRTSAPGYYRMLLGDFEITAL 60
OPHC2       ----MRLF---SLSTA---LSSAMIALVSLPLQAAAPAQQKTQVPGYYRMALGDFEVTAL 50
                 **:   *  :*    ::*:  *. :    **  * :*..****** *****:***

WBC3MPH     SDGTVALPVDKRLNQPAPKTQSALAKSFQ--KAPLETSVTGYLVNTGSKLVLVDTGAAGL 118
OPHC2       YDGYVDLPASLLKGIDDKDLQSLLARMFVASEKGVQTAVNAYLINTGDNLVLIDTGAAQC 110
             ** * **..        . ** **: *   :  ::*:*..**:***.:***:*****

WBC3MPH     FGPTLGRLAANLKAAGYQPEQVDEIYITHMHPDHVGGLMVG-EQLAFPNAVVRADQKEAD 177
OPHC2       FGPTLGVVQTNLKASGYQPEQVDTVLLTHLHPDHACGLVNADGSPAYPNATVEVPQAEAE 170
            ****** : :****:******** : :**:****. **: .  . *:***.*.. * **:

WBC3MPH     FWLSQTNLDKAPDDESKGFFKGAMASLNPYVKAGKFKPFSGNTDLVPGIKALASHGHTPG 237
OPHC2       FWLDEATMAKAPEG-MQGMFKMAQQAVAPYAKMNKLKPYKTEGELLPGVSLVASPGHTPG 229
            ***.::.: ***:   :*:** *  :: **.*  *:**:. : :*:**:. :** *****

WBC3MPH     HTTYVVESQGQKLALLGDLILVAAVQFDDPSVTTQLDSDSKSVAVERKKAFADAAKGGYL 297
OPHC2       HTSYLFKSGGQSLLVWGDILLNHAVQFAKPEVVFEFDVDSDQARQSRQRILAEAATDKLW 289
            **:*:.:* **.* : **::*  **** .*.*. ::* **...  .*:: :*:**.

WBC3MPH     IAASHLSFPGIGHIRAEGKGYRFVPVNYSVVNPK-     331
OPHC2       VAGAHLPFPGLGHVRKEAQGYAWVPVEFSPIRSDR     324
            :*.:** ***:**:* *.:** :***::* :. .
```

Legend:
WBC3MPH – *Pseudomonas* sp. WBC-3
OPHC2 – *Pseudomonas pseudoalcaligenes*

Residues in blue are the metal-binding ligands

Appendix F: Full List of Mutants from Chapter 4

The full list of all mutations present in mutant *mpd* genes generated during the process of directed evolution described in Chapter 4 are presented across the following pages. The asterisk (*) on Round 1 mutants indicates the mutants contain only silent mutations.

	34	50	90	107	109	113	116	128	162	164	191	202	209	211	212	226	237	256
WT	**S**	**M**	**K**	**K**	**V**	**T**	**A**	**L**	**L**	**F**	**D**	**A**	**K**	**G**	**K**	**I**	**G**	**L**
Round 1																		
1D4																		Q
1G6					A													
2B3																		
3A2*																		
3B5																		
3B10															K			
3C2																		
3D2																		
3 E2						A												
3G2*																		
3H4*																		
4A6																		
4C6				N														
5A7		V										S						
5B8											G							
5F11																		
Round 2																		
1A3																		
1A7																V		
1B3																		Q
1D2																		Q
1 E8			E															
1F8			E															Q
1G5																		Q
1H11						A												
2A7																		Q
2B8																		Q
2C4																		
2D4																		
2D8																		Q
2D9																		
2 E11																		Q
2G9																		
2H5																		
2H8																		
3A5																		
3A8																		Q
3B4						A												
3B8													R					
3C6																		Q
3C10																		
3 E9											G							
3F2																		Q
3G5												S						
3H4																		Q
4A7																		
4A8																		
4C9																		Q
4D8								P										Q
4 E3				N														
4 E8						A												
4F3																		Q
4F11																		Q
4G5																		Q
4G10																		
4H5																		
4H9																		Q
5A8																		
5B6											G							
5B8																		
5C2																		Q
5 E2																		Q
5F6																		
5F10																		Q
5G6			E									S						
6C3																		Q
6D6												S						
6 E4													R					Q
6 E7													R					Q
6F10			E									S						
6G5																		Q
6H5																		Q
7B5																		Q
7B9				N														
7C8																		
7D11																		

	264	265	266	268	275	278	281	282	285	293	300	307	313	315	316	318	320
WT	**F**	**D**	**D**	**S**	**S**	**K**	**A**	**V**	**K**	**K**	**A**	**A**	**A**	**G**	**K**	**Y**	**F**
Round 1																	
1D4																	
1G6			E					A									
2B3																	
3A2*																	
3B5													D				
3B10																	
3C2													D				
3D2		E															
3 E2																	
3G2*																	
3H4*																	
4A6																H	
4C6																	
5A7														D			
5B8															E		
5F11																	Y
Round 2																	
1A3													D				
1A7																	
1B3																	
1D2															E		
1 E8																	
1F8																	
1G5																	
1H11															E		
2A7																	
2B8																	
2C4															E		
2D4													D				
2D8											G						
2D9				N													
2 E11																	
2G9															E		
2H5															E		
2H8															E		
3A5													D				
3A8																	
3B4															E		
3B8													D				
3C6																	
3C10													D				
3 E9															E		
3F2																	
3G5																	
3H4	L																
4A7													D				
4A8															E		
4C9															E		
4D8															E		
4 E3															E		
4 E8														D			
4F3																	
4F11																	
4G5															E		
4G10															E		
4H5															E		
4H9																	
5A8													D				
5B6									E								
5B8															E		
5C2																	
5 E2																	
5F6														D			
5F10														D			
5G6																	
6C3																	
6D6														D			
6 E4															E		
6 E7															E		
6F10														D			
6G5																	
6H5																	
7B5																	
7B9						N									E		
7C8															E		
7D11															E		

	34	50	90	107	109	113	116	128	162	164	191	202	209	211	212	226	237	256
WT	S	M	K	K	V	T	A	L	L	F	D	A	K	G	K	I	G	L
Round 3																		
1A2									P									Q
1A4				N							G							Q
1A10																		Q
1B2												S						Q
1B8																		Q
1D10																		Q
1 E2																		Q
1 E3																		Q
1F8																		Q
1G4																		Q
1G6																		Q
1G7													R					Q
1H5												S						
2A6																		Q
2A7																		Q
2B4										Y								Q
2B11																		Q
2C11						A												
2D6																		Q
2D10																		Q
2 E6																		
2F6																		Q
2F7												S						Q
2H3						A												Q
3A4																		Q
3C6																		Q
3D5			E															Q
3 E8	T																	Q
3G10							G											Q
3H8			E															Q
3H9																		
4C2																		Q
4C7																		Q
4D9																		
4 E3			E										R					Q
4 E5																		Q
4G8			E															Q
4H11																		Q
5A9			E									S						
5B4												S						Q
Round 4																		
1A4			E															Q
1D5																		Q
1 E6						A												Q
5H6																		Q
6A10																		Q
6B3																		Q
6C11																		Q
8A6						A								D			D	Q
8C8																		Q
8H6																		Q
9B8												S						Q
11A3										Y								Q
10H2																		Q
10G8																		Q
13D4																		Q
13D11			E															Q
13G6												S						Q
13G9																		Q
14A7																		Q
14H9																		Q
15 E2																		Q
15F3																		Q
15H5																		Q
17F11																		Q
16A4																		Q
20A6			E															Q
16H7																		Q
17 E10																		Q
20H9																		Q

	264	265	266	268	275	278	281	282	285	293	300	307	313	315	316	318	320
WT	**F**	**D**	**D**	**S**	**S**	**K**	**A**	**V**	**K**	**K**	**A**	**A**	**A**	**G**	**K**	**Y**	**F**
Round 3																	
1A2															E		
1A4							V								E		
1A10																	
1B2																	
1B8															E		
1D10															E		
1 E2																	
1 E3											G				E		
1F8								A									
1G4															E		
1G6											G	D					
1G7															E		
1H5															E		
2A6																	
2A7																	
2B4																	
2B11																	
2C11															E		
2D6																	
2D10											G						
2 E6						N									E		
2F6															E		
2F7							T								E		
2H3									E						E		
3A4	L														E		
3C6																	
3D5																	
3 E8																	
3G10														D			
3H8															E		
3H9															E		
4C2															E		
4C7															E		
4D9																	
4 E3															E		
4 E5																	
4G8										N							
4H11																	
5A9				D											E		
5B4															E		
Round 4																	
1A4															E		
1D5											G				E		
1 E6							V								E		
5H6															E		
6A10											G				E		
6B3															E		
6C11															E		
8A6															E		
8C8															E		
8H6															E		
9B8					I										E		
11A3															E		
10H2															E		
10G8															E		
13D4															E		
13D11															E		
13G6							T								E		
13G9						N									E		
14A7															E		
14H9											G				E		
15 E2															E		
15F3							T								E		
15H5															E		
17F11							V								E		
16A4															E		
20A6															E		
16H7						N									E		
17 E10											G				E		
20H9															E		

Appendix G: Kinetics Overview for MPH Variants from Chapter 5

Kinetics parameters for the hydrolysis of various OP various MPH variants.
The relative k_{cat}, K_m and k_{cat}/K_m values are colour-coded for easy reference. Green: Increase in value; Pink: Decrease in value; ND: Not determined.

Substrate	Variant	k_{cat} (s^{-1})			Relative k_{cat}	Km (μM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative K_{cat}/K_m
MPS	WT	27.5	±	1.800	1.00	27.3	±	6.0	1.00	1007.3	1.00
	F119Y	28.8	±	0.300	1.05	29.8	±	1.0	1.09	966.4	0.96
	F196L	1.5	±	0.060	0.05	45.6	±	4.5	1.67	32.9	0.03
	F196I	1.7	±	0.050	0.06	12.9	±	1.7	0.47	131.8	0.13
	F196Y	34.5	±	0.700	1.25	37.8	±	2.2	1.38	912.7	0.91
	L258S	1.2	±	0.050	0.04	25.8	±	3.7	0.95	46.5	0.05
	L258I	14.1	±	0.500	0.51	5.2	±	0.9	0.19	2711.5	2.69
	L258H	1.7	±	0.070	0.06	16.9	±	2.7	0.62	100.6	0.10
	L258N	3.5	±	0.300	0.13	48.2	±	10.0	1.77	72.6	0.07
	L273I	3.0	±	0.100	0.11	12.8	±	1.8	0.47	234.4	0.23
	L273V	7.7	±	0.050	0.28	8.9	±	0.3	0.33	865.2	0.86
MPO	WT	4.5	±	0.300	1.00	72.2	±	10.2	1.00	62.3	1.00
	F119Y	8.6	±	1.500	1.91	241.9	±	65.3	3.35	35.6	0.57
	F196L	1.1	±	0.100	0.24	297.9	±	53.1	4.13	3.7	0.06
	F196I	1.3	±	0.100	0.29	286.8	±	45.6	3.97	4.5	0.07
	F196Y	6.6	±	0.600	1.47	208.8	±	30.9	2.89	31.6	0.51
	L258S	6.4	±	1.000	1.42	187.7	±	47.3	2.60	34.1	0.55
	L258I	8.7	±	0.200	1.93	57.9	±	2.8	0.80	150.3	2.41
	L258H	3.0	±	0.200	0.67	27.4	±	5.9	0.38	109.5	1.76
	L258N	15.2	±	0.900	3.38	195.6	±	20.0	2.71	77.7	1.25
	L273I	16.4	±	0.700	3.64	116.5	±	10.5	1.61	140.8	2.26
	L273V	18.7	±	3.000	4.16	108.1	±	35.3	1.50	173.0	2.78
EPS	WT	1.1	±	0.004	1.00	10.8	±	1.4	1.00	101.9	1.00
	F119Y	3.6	±	0.200	3.27	14.8	±	2.2	1.37	243.2	2.39
	F196L	0.6	±	0.020	0.55	21.2	±	2.7	1.96	28.3	0.28
	F196I	1.9	±	0.100	1.73	12.9	±	1.8	1.19	147.3	1.45
	F196Y	1.8	±	0.040	1.64	16.1	±	0.8	1.49	111.8	1.10
	L258S	ND			ND	ND			ND	ND	ND
	L258I	0.7	±	0.010	0.64	4.0	±	0.5	0.37	175.0	1.72
	L258H	ND			ND	ND			ND	ND	ND
	L258N	0.1	±	0.010	0.09	18.8	±	3.7	1.74	5.3	0.05
	L273I	0.2	±	0.003	0.18	5.5	±	0.4	0.51	36.4	0.36
	L273V	0.3	±	0.030	0.27	10.3	±	4.0	0.95	29.1	0.29
EPO	WT	0.3	±	0.010	1.00	49.0	±	5.5	1.00	6.1	1.00
	F119Y	0.8	±	0.100	2.67	84.3	±	21.1	1.72	9.5	1.56
	F196L	0.7	±	0.030	2.33	96.3	±	9.5	1.97	7.3	1.19
	F196I	1.2	±	0.100	4.00	109.1	±	9.8	2.23	11.0	1.80
	F196Y	0.7	±	0.070	2.33	86.7	±	18.8	1.77	8.1	1.32
	L258S	0.5	±	0.003	1.67	62.4	±	9.0	1.27	8.0	1.31
	L258I	0.5	±	0.010	1.67	18.6	±	1.9	0.38	26.9	4.41
	L258H	0.8	±	0.040	2.67	18.6	±	3.9	0.38	43.0	7.05
	L258N	1.7	±	0.040	5.67	102.0	±	5.6	2.08	16.7	2.73
	L273I	0.7	±	0.030	2.33	46.7	±	6.1	0.95	15.0	2.46
	L273V	0.7	±	0.010	2.33	27.3	±	2.5	0.56	25.6	4.20
MCO	WT	6.9	±	0.100	1.00	21.4	±	1.6	1.00	322.4	1.00
	F119Y	3.3	±	0.040	0.48	51.5	±	1.7	2.41	64.1	0.20
	F196L	0.5	±	0.010	0.07	48.3	±	4.0	2.26	10.4	0.03
	F196I	1.5	±	0.040	0.22	57.9	±	4.1	2.71	25.9	0.08
	F196Y	9.5	±	0.400	1.38	72.5	±	7.8	3.39	131.0	0.41
	L258S	76.1	±	2.700	11.03	31.5	±	4.0	1.47	2415.9	7.49
	L258I	7.0	±	0.200	1.01	42.9	±	4.2	2.00	163.2	0.51
	L258H	23.1	±	0.200	3.35	64.5	±	1.6	3.01	358.1	1.11
	L258N	133.1	±	2.400	19.29	35.6	±	2.2	1.66	3738.8	11.60
	L273I	13.6	±	0.200	1.97	31.7	±	2.0	1.48	429.0	1.33
	L273V	17.4	±	0.500	2.52	23.4	±	2.9	1.09	743.6	2.31
ECO	WT	0.3	±	0.010	1.00	15.6	±	2.5	1.00	19.2	1.00
	F119Y	0.3	±	0.002	1.00	23.1	±	0.9	1.48	13.0	0.68
	F196L	0.2	±	0.010	0.67	37.8	±	4.4	2.42	5.3	0.28
	F196I	0.6	±	0.004	2.00	33.7	±	0.9	2.16	17.8	0.93
	F196Y	0.5	±	0.010	1.67	31.7	±	2.6	2.03	15.8	0.82
	L258S	9.7	±	0.200	32.33	10.1	±	1.5	0.65	960.4	50.02
	L258I	0.4	±	0.004	1.33	12.7	±	0.7	0.81	31.5	1.64
	L258H	1.6	±	0.010	5.33	11.8	±	0.5	0.76	135.6	7.06
	L258N	27.4	±	0.600	91.33	24.3	±	2.2	1.56	1127.6	58.73
	L273I	0.5	±	0.020	1.67	18.5	±	3.3	1.19	27.0	1.41
	L273V	0.3	±	0.006	1.00	8.5	±	1.3	0.54	35.3	1.84

Kinetic parameters for the hydrolysis of various OP by multi-site MPH variants.
The relative k_{cat}, K_m and k_{cat}/K_m values are color-coded for easy reference. Green: Increase in value; Pink: Decrease in value; ND: Not determined.

Substrate	Variant	k_{cat} (s^{-1})			Relative k_{cat}	K_m (mM)			Relative K_m	k_{cat}/K_m (s^{-1} mM^{-1})	Relative k_{cat}/K_m
MPS	WT	27.5	±	1.80	1.00	27.3	±	6.0	1.00	1007.3	1.00
	R1A6	6.3	±	0.20	0.23	22.5	±	2.2	0.82	280.0	0.28
	R1B6	0.3	±	0.01	0.01	19.1	±	1.5	0.70	15.7	0.02
	R2B2	0.7	±	0.01	0.03	16.1	±	0.8	0.59	43.5	0.04
	R2C2	1.0	±	0.03	0.04	7.9	±	1.2	0.29	126.6	0.13
	R2D2	13.9	±	0.40	0.51	10.1	±	1.2	0.37	1376.2	1.37
	R2A2	2.3	±	0.30	0.08	27.8	±	7.4	1.02	82.7	0.08
	R2F3	1.5	±	0.10	0.05	50.4	±	3.4	1.85	29.8	0.03
	R2D4	1.8	±	0.10	0.07	12.0	±	2.1	0.44	150.0	0.15
MPO	WT	4.5	±	0.30	1.00	72.2	±	10.2	1.00	62.3	1.00
	R1A6	30.7	±	5.50	6.82	231.4	±	65.2	3.20	132.7	2.13
	R1B6	2.6	±	0.40	0.58	213.3	±	51.1	2.95	12.2	0.20
	R2B2	19.4	±	1.60	4.31	321.0	±	39.1	4.45	60.4	0.97
	R2C2	3.2	±	0.40	0.71	139.2	±	28.6	1.93	23.0	0.37
	R2D2	24.6	±	1.30	5.47	44.5	±	7.5	0.62	552.8	8.87
	R2A2	7.0	±	0.60	1.56	187.5	±	28.2	2.60	37.3	0.60
	R2F3	6.3	±	0.40	1.40	57.6	±	8.9	0.80	109.4	1.76
	R2D4	4.4	±	0.50	0.98	66.0	±	15.2	0.91	66.7	1.07
EPS	WT	1.1	±	0.00	1.00	10.8	±	1.4	1.00	101.9	1.00
	R1A6	1.6	±	0.10	1.45	34.5	±	4.5	3.19	46.4	0.46
	R1B6	0.3	±	0.01	0.27	11.5	±	0.7	1.06	26.1	0.26
	R2B2	ND			ND	ND			ND	ND	ND
	R2C2	ND			ND	ND			ND	ND	ND
	R2D2	2.3	±	0.20	2.09	25.4	±	6.9	2.35	90.6	0.89
	R2A2	2.0	±	0.02	1.82	2.1	±	0.3	0.19	952.4	9.35
	R2F3	ND			ND	ND			ND	ND	ND
	R2D4	ND			ND	ND			ND	ND	ND
EPO	WT	0.3	±	0.01	1.00	49.0	±	5.5	1.00	6.1	1.00
	R1A6	4.4	±	0.10	14.67	46.6	±	2.6	0.95	94.4	15.48
	R1B6	7.4	±	0.20	24.67	202.2	±	10.5	4.13	36.6	6.00
	R2B2	14.5	±	0.80	48.33	68.4	±	8.9	1.40	212.0	34.75
	R2C2	3.3	±	0.10	11.00	31.6	±	4.2	0.64	104.4	17.12
	R2D2	11.2	±	0.30	37.33	36.7	±	3.0	0.75	305.2	50.03
	R2A2	9.1	±	0.30	30.33	43.2	±	4.1	0.88	210.6	34.53
	R2F3	22.0	±	2.00	73.33	37.0	±	10.5	0.76	594.6	97.47
	R2D4	13.6	±	0.90	45.33	54.9	±	9.3	1.12	247.7	40.61
MCO	WT	6.9	±	0.10	1.00	21.4	±	1.6	1.00	322.4	1.00
	R1A6	249.1	±	10.30	36.10	16.8	±	4.1	0.79	14827.4	45.99
	R1B6	144.2	±	8.47	20.90	135.3	±	18.5	6.32	1065.8	3.31
	R2B2	60.2	±	1.70	8.72	70.4	±	5.9	3.29	855.1	2.65
	R2C2	9.8	±	0.40	1.42	55.9	±	7.6	2.61	175.3	0.54
	R2D2	117.8	±	6.60	17.07	67.0	±	11.4	3.13	1758.2	5.45
	R2A2	3.9	±	0.10	0.57	70.6	±	3.3	3.30	55.2	0.17
	R2F3	42.4	±	2.90	6.14	103.9	±	18.1	4.86	408.1	1.27
	R2D4	10.4	±	0.10	1.51	57.1	±	1.7	2.67	182.1	0.56
ECO	WT	0.3	±	0.01	1.00	15.6	±	2.5	1.00	19.2	1.00
	R1A6	91.3	±	5.20	304.33	17.8	±	5.1	1.14	5129.2	267.15
	R1B6	109.7	±	6.10	365.67	49.9	±	8.0	3.20	2198.4	114.50
	R2B2	30.3	±	0.90	101.00	12.6	±	2.8	0.81	2404.8	125.25
	R2C2	4.2	±	0.10	14.00	14.5	±	1.6	0.93	289.7	15.09
	R2D2	31.6	±	1.30	105.33	37.1	±	5.9	2.38	851.8	44.36
	R2A2	2.8	±	0.10	9.33	21.1	±	4.3	1.35	132.7	6.91
	R2F3	35.8	±	0.50	119.33	22.2	±	1.6	1.42	1612.6	83.99
	R2D4	10.2	±	0.10	34.00	19.8	±	1.5	1.27	515.2	26.83

Variants	Mutations
R1A6	F119Y, L258S
R1B6	F196I, L258N
R2B2	F196I, L258H
R2C2	F196I, L273I
R2D2	F119Y, L258H
R2A2	F119Y, F196I, L273V
R2F3	F119Y, F196Y, L258H
R2D4	F119Y, F196I, L273I

Appendix H: Oversampling Formula used in CASTer

Oversampling required was calculated using CASTer 2.0, a handy worksheet written by Reetz. (http://www.kofo.mpg.de/en/research/organic-synthesis) .The algorithm used in CASTer is approximated by Poisson distribution, described by **Equation E.1** and was discussed by Patrick *et al.*[1]:

$$P(r) = \frac{e^{-\mu}\mu^{r}}{r!} \tag{E.1}$$

Where $P(r)$ is the probability any one variant, v_i occurs exactly r times in the library, and μ is the mean number of occurrence v_i in the library, assuming all possible variants are equally probable. The mean, μ, can also be rewritten as $\mu = L/V$ where L is the library size and V is the total number of possible sequence variants. The probability of all variants occurring at least once would be 1- $P(0)$. 1-$P(0)$ can be defined as F, the fraction of completeness of the library. **Equation E.1** can be transformed into

$$\begin{aligned} 1 - P(0) &= 1 - e^{-\frac{L}{V}} \\ F &= 1 - e^{-\frac{L}{V}} \\ \frac{L}{V} &= -\ln(1-F) \\ L &= -V\ln(1-F) \end{aligned} \tag{E.2}$$

The relationship between the sampling factor, L/V, and library coverage can be estimated from **Equation E.2** and is shown in **Figure E.1**. Due to the exponential nature of the relationship, library coverage beyond 95% requires exhaustive screening efforts. To achieve 95% library coverage, the sampling factor is about three, which means three fold excess of transformants needs to be screened.

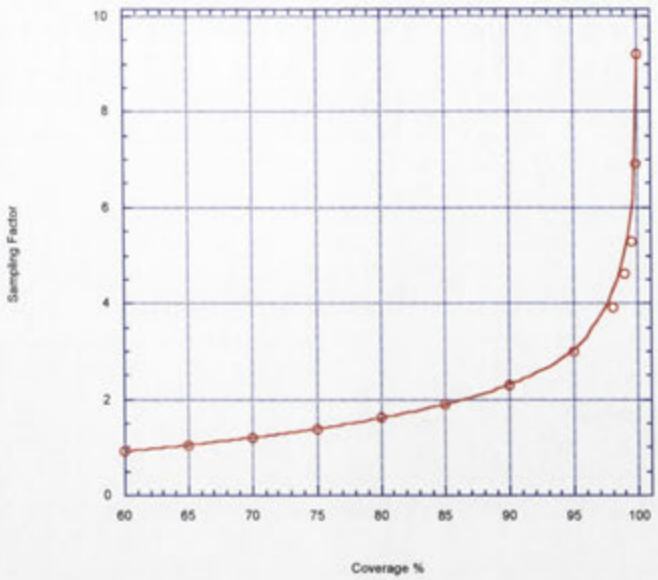


Figure E.1: Correlation between library coverage and oversampling of enzyme variants for a given codon randomization calculated based on Equation E.1.
Sampling factor, *L/V*, represents the number of folds of excess to screen in order to achieve a certain percentage of coverage.

References:

1. Patrick, W. M., Firth, A. E. & Blackburn, J. M. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Engineering* **16,** 451–457 (2003).