DIRECTED EVOLUTION, RATIONAL DESIGN AND MECHANISTIC STUDIES OF A PHOSPHOHYDROLASE FROM *ENTEROBACTER AEROGENES*

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

by

Sylvia Hsu-Chen Yip

February 2010

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. I have made an effort to acknowledge the contributions of individuals to this thesis as listed below as well as in the relevant parts of the thesis:

- Plasmid DNA of *gpd*Q in pCY76 was obtained from Dr Colin Jackson and Dr Hye-Kyung Kim.
- *E. coli* DH10β cells carrying the putative *ugp* operon cloned in the plasmid pBlueScript were prepared by Dr Sean Yu-McLoughlin.
- Primers for PCR sequencing, namely 4pCY76 and M13 were originally obtained from Dr Jian-Wei Liu who also designed the latter. 4pCY76 was designed by Dr Bradley Stevenson.
- Plasmid DNA of *opd*A in pETMCSI was obtained courtesy of Dr Jee-Loon Foo.
- Magnetic circular dichroism studies on wild-type GpdQ, mutants N80D and N80A were performed by Dr Kieran Hadler during his PhD candidature at the University of Queensland.
- Dr Paul Carr and Kelly Nguyen were involved in the X-ray diffraction studies, data collection and analysis of wild-type GpdQ crystal structure 3D03.
- Dr Jee-Loon Foo performed the soaking experiments for all the GpdQ crystals made in this study.
- Crystallisation screens of GpdQ mutant F21K were set up by Miss Tracy Murray.

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

Sylvia Hsu-Chen Yip

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my heartfelt gratitude to my supervisor Professor David Ollis, whose support, guidance and help from the initial to final stages of my candidature have helped me to develop a solid understanding and passion for protein research. I am also greatly indebted to Dr Jian-Wei Liu for his technical help in the laboratory; Dr Paul Carr for sharing his time and expertise in protein crystallography and structural analysis and for putting up with my lack of knowledge in these areas. A very huge 'thank you' goes to Drs Sean Yu-McLoughlin and Colin Jackson, my 'predecessors' in GpdQ research, whose previous work made the enzyme such an interesting subject to work on.

I would like to thank all the past and present members of the Ollis and Otting research groups – Jee-Loon (my saviour each time I messed up my experiments), Chris, Morgan, Hye-Kyung, Amalia, Tracy, Sapna, Joe, Bradley, Ruhu, Karin, Kiyoshi, Hiromasa, Xuncheng, Tiffany, Laura, Xinying and Kelly. It has been sheer joy working with you all these 3.5 years.

Special thanks to Associate Professors Gerhard Schenk and Lawrence Gahan of the University of Queensland, their students Kieran, Eleanor and Fernanda who collaborate with the Ollis group on GpdQ research. I am also grateful to the Australian National University and CSIRO for financially supporting my studies.

Last but not least, I thank my family for their unconditional support and love. This thesis is dedicated to them. Very special thanks go to my friends in Toad Hall, particularly Om (for proofreading and formatting my thesis), Sumila, Sandun, Sudewa, Hana and Akchousanh for showing me the true meaning of friendship when it mattered the most to me. This thesis happened because my family and good friends believed in me.

ABSTRACT

GpdQ is a phosphohydrolase from *Enterobacter aerogenes* whose rare ability to hydrolyse simple diesters of phosphoric acid like dimethyl phosphate and diethyl phosphate first caught the attention and interest of researchers in the 1970s. When GpdQ was characterised further in more recent times, it was found to have an exceptionally broad substrate repertoire. The enzyme has primarily phosphodiesterase activity – capable of hydrolysing a wide range of phosphodiesters including glycerophosphodiesters that are believed to be its natural substrates. Additionally, GpdQ has low phosphomonoesterase, phosphotriesterase and phosphonate monoesterase activities.

In this study, the promiscuity of GpdQ was exploited and we set out to improve the enzyme's activity on one of the non-physiological phosphodiesters, bis (*p*-nitrophenyl) phosphate. Directed evolution was the main strategy used to achieve this goal. After the first four rounds of evolution, the hexameric quaternary structure of the catalytically improved GpdQ mutants was found to have broken down and high activity was associated with the dimer. Rational design was then employed to identify mutations that would promote dissociation and form a stable GpdQ dimer. These rationally-designed mutants were included for gene shuffling from the fifth round of directed evolution onwards. At the end of this study, eight full rounds of evolution were completed. Three GpdQ dimers displaying ~500-fold k_{cat}/K_m improvement were successfully engineered.

In the course of this work, mutagenesis studies of GpdQ active site residues were also performed. In collaboration with fellow GpdQ investigators in the University of Queensland, the non-identical binding affinities of the two metal-coordinating sites was successfully characterised, providing more insight into the catalytic mechanism of the enzyme. Finally, future research

directions are also proposed in order to further characterise the substrate specificity of the mutants with improved b*p*NPP activity from this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS USED IN THIS DISSERTATION

Abbrevations and acronyms	Full names
3-25	GpdQ mutant with mutations C54G, Y221H and G259R
°C	Degree Celsius
Å	Angstrom
ABC	ATP-binding cassette
AchE	Acetylcholinesterase
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AmS	Ammonium sulfate
b	Nucleotide base
bp	Nucleotide base pair
b*p*NPP	bis (*para*-nitrophenyl) phosphate
Bis-Tris propane	1,3-bis(tris(hydroxymethyl)methylamino)propane
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
cAMP	Cyclic adenosine monophosphate
Cd^{2+}	Cadmium (II) ion
CIP	Calf intestinal alkaline phosphatase
cm	centimeter
Co^{2+}	Cobalt (II) ion
$CoCl_2$	Cobalt chloride
$CuSO_4$	Cupric sulfate
Da	Dalton
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DEP	Diethyl phosphate
DMP	Dimethyl phosphate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
E. aerogenes	*Enterobacter aerogenes*
E. coli	*Escherichia coli*
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error-prone polymerase chain reaction
EPR	Electron paramagnetic resonance
EtBr	Ethidium bromide
EtOH	Ethanol
Fe^{2+}	Ferrous or iron (II) ion
Fe^{3+}	Ferric or iron (III) ion
FPLC	Fast protein liquid chromatography

g	Gram or force during centrifugation, depending on context
G3P	*sn*-glycerol-3-phosphate
GDPD	Glycerophosphodiester phosphodiesterase
GlpQ	Periplasmic GDPD from *E. coli*
GpdQ	Phosphohydrolase from *Enterobacter aerogenes*
*gpd*Q	Gene expressing the phosphohydrolase from *E. aerogenes*
GPD	Glycerophosphodiester
GPE	*sn*-glycerolphosphoethanolamine
HCl	Hydrochloric acid
H_3PO_4	Phosphoric acid
HEPES	4-2-hydroxyethyl-1-piperazineethanesulfonic acid
k	10^3 or kilo
kDa	Kilodalton
K_2HPO_4	Dipotassium hydrogen orthophosphate
K_2SO_4	Potassium sulfate
KH_2PO_4	Potassium dihydrogen orthophosphate
k_{cat}	Catalytic constant
K_M	Michaelis constant
L	Leucine or liter, depending on context
l	Liter
LB	Luria-Bertani medium
LBA	Luria-Bertani-ampicillin medium
M	Methionine or molar, depending on context
m	10^{-3} or mili
μ	10^{-6} or micro
M. tuberculosis	*Mycobacterium tuberculosis*
MCD	Magnetic circular dichroism
MES	2-(N-morpholino)ethanesulfonic acid
Mg^{2+}	Magnesium (II) ion
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate
min	Minute
MJ0936	Phosphodiesterase from *Methanococcus jannaschii*
Mn^{2+}	Manganese (II) ion
$MnCl_2$	Manganese chloride
MOPS	3-(N-morpholino)propanesulfonic acid
MPO	Methyl paraoxon or paraoxon-methyl
mQH_2O	Reagent grade water purified and deionised using the Milli-Q system
n	10^{-9} or nano
nm	nanometer
NaCl	Sodium chloride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
Na_2HPO_4	Sodium phosphate dibasic
NH_4Cl	Ammonium chloride

Ni^{2+}	Nickel (II) ion
OD	Optical density
OP	Organophosphate
OPAA	Organophosphorus acid anhydrase (OPAA) from *Altermonas sp.*
OpdA	Phosphotriesterase from *Agrobacterium radiobacter*
*opd*A	Gene expressing the phosphotriesterase from *A. radiobacter*
OPH	Phosphotriesterase from *Pseudomonas diminuta*
ORF	Open reading frame
P. furiosus	*Pyrococcus furiosus*
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
pDNA	Plasmid DNA
P_i	Inorganic phosphate
pI	Isoelectric point
pho	Bacterial regulon that expresses genes responsible for the transport and assimilation of inorganic phosphate
*pho*A	Gene from *pho* regulon that expresses alkaline phosphatase
*pho*B	Positive regulatory gene in *pho* regulons
*p*NPP	*para*-nitrophenyl phosphate
PP-1	Protein phosphatase 1
PTE	Phosphotriesterase
RE	Restriction endonuclease or enzyme
RNA	Ribonucleic acid
rpm	Revolutions per minute
Rv0805	3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis*
s	Second
SB	Sodium hydroxide-boric acid buffer
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSM	Site-saturation mutagenesis
StEP	Staggered extension process
T. maritima	*Thermotoga maritima*
TB	Terrific broth medium
TBA	Terrific broth-ampicillin medium
TBE	Tris-boric acid-EDTA buffer
TEMED	N,N,N',N',-tetramethylethylenediamine
TIM	Common $(\alpha/\beta)_8$ tertiary fold in triosephosphateisomerases and other proteins
T_m	Melting temperature
Tricine	N-tris(hydroxymethyl)methylglycine
Tris	Tris(hydroxymethyl)methylamine
U	Enzyme unit

ugp	Uptake of glycerol-3-phosphate, name of operon in *E. coli*
UgpQ	Cytoplasmic GDPD from *E. coli*, part of *ugp* operon
UV	Ultraviolet
V	Valine or volt, depending on context
v/v	Volume per volume ratio
vs	Versus
w/v	Weight per volume ratio
YENB	Yeast extract and nutrient broth medium
Zn^{2+}	Zinc (II) ion
$ZnSO_4$	Zinc sulfate

ABBREVIATIONS FOR AMINO ACIDS

Throughout this dissertation, amino acids are often referred to using single-letter codes:

A	Alanine
C	Cysteine
D	Aspartic acid/aspartate
E	Glutamic acid/glutamate
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

LIST OF PUBLICATIONS

1. Jackson, C. J., Hadler, K. S., Carr, P. D., Oakley, A. J., Yip, S., Schenk, G., and Ollis, D. L. (2008) Malonate-bound structure of the glycerophosphodiesterase from Enterobacter aerogenes (GpdQ) and characterization of the native Fe2+ metal-ion preference. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications 64*, 681-685.
2. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from Enterobacter aerogenes. *Journal of the American Chemical Society 130*, 14129-14138.
3. Hadler, K. S., Mitic, N., Yip. S. H. C., Gahan, L. R., Ollis, D. L. Schenk, G. and Larrabee, J. A. (2009) Electronic structure analysis of the binuclear metal center in the bioremediator glycerophosphodiesterase (GpdQ) from Enterobacter aerogenes. *Manuscript under review.*

1

INTRODUCTION AND LITERATURE REVIEW

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

This dissertation is about directed evolution, rational design and mechanistic studies of a phosphohydrolase enzyme from *Enterobacter aerogenes*. The enzyme in question is GpdQ, which has very high levels of activity with phosphodiesters and low levels of activity with other types of phosphate esters.

This chapter presents introductions and overviews of:

1.1 Properties and limitations of enzymes: the long-term aims of this work.

1.2 Organophosphate compounds that are substrates of GpdQ.

1.3 An introduction to GpdQ – its origin, isolation and cloning, catalytic properties.

1.4 The structure of GpdQ.

1.5 Rv0805, a 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* that has the closest sequence and structural similarity to GpdQ.

1.6 Protein engineering strategies used in this study, namely directed evolution and rational design. This section also explains how rational design complements directed evolution in engineering enzymes.

1.7 The overview of this thesis.

1.1 ENZYMES

Enzymes are biological catalysts – they are usually proteins. There are numerous reviews and biochemical texts that give detailed description of the properties of these molecules and a detailed coverage will not be provided here. For our purposes it is sufficient to say that enzymes are molecules that catalyse or increase the rate of chemical reactions so that life is possible. The ability of enzymes to accelerate reactions has not escaped the attention

of chemists and applications for enzymes are increasingly being found in industrial and academic settings. However, their properties impose practical limits upon their use. There are a few obvious examples. For instance, enzymes are not stable over wide temperature ranges. Most enzymes from enteric bacteria work at the optimum temperature of 37°C and suffer from denaturation at high temperatures. More importantly, not all reactions of industrial importance are found in nature and an appropriate enzyme to catalyse a reaction may not exist naturally. If we could design or evolve enzymes for particular applications then these molecules would provide an efficient and clean way to carry out chemical conversions. This thesis is concerned with the process by which the substrate specificity of an enzyme can be altered. The field of directed evolution is still rapidly developing and much work needs to be done in understanding how to evolve enzymes and understanding the molecular consequences of this evolution. For reasons outlined later, I chose to work with an enzyme that had a very broad specificity and set out to increase its activity towards a relatively poor substrate. Specifically I was interested in enhancing the ability of GpdQ to degrade non-physiological substrates – simple phosphate esters. This study was intended to be the fist step in evolving GpdQ to degrade a variety of phosphate esters. It was also intended to provide some understanding how the enzymes could possibly be redesigned to catalyse new hydrolysis reactions.

In summary, the aim of my work was to first evolve an enzyme and secondly to try and understand how improvements in activity came about. To achieve this second aim we need to have some knowledge of the enzyme structure and in particular the enzyme active site. Would attempts to change the substrate specificity of an enzyme merely change residues in the active site or would the changes be more widespread? If the changes were widespread *ie.* not restricted to the active site then we would like to know what effect these changes had on the active site – how did the changes

enhance activity? These are the types of questions that we would like to address with the results from the directed evolution studies.

The structure of a protein is traditionally described on four levels – primary (the amino acid sequence of a protein), the secondary structure (α-helices and β-sheets), tertiary structure (subunit or chain structure – the arrangement of secondary structure elements) and quaternary structure (the arrangement of a number of subunits to give a functional enzyme molecule complex) [1]. Much has been written about protein structure and little needs to be added here. However, it is worth mentioning that protein subunits (tertiary structure) are often divided into domains – discrete portions of proteins made up of compact globular folds that are thought to be important in protein folding, function and evolution. Protein chains larger than ~200 residues usually fold into such globular clusters [1, 2]. Similar domains are often found in a variety of proteins. For example, the α/β hydrolase fold is an ideal fold for forming hydrolytic enzymes [3]. Similarly, the structure of the catalytic domain of GpdQ has been found in a variety of phosphohydrolases and it is worth looking at this structure looking at it in some detail before analysing the results of directed evolution. We do this in section 1.4 of this introductory chapter.

1.1.1 Enzyme active sites

The active site of an enzyme is where catalysis takes place [4] and it is usually a cleft that is surrounded by residues whose protruding side chains form the surface of the site and interact directly with ligand molecules (substrate, product or co-factor). The shape, structure and chemical properties of the active site allow the recognition of the substrate molecule, thus giving enzymes the specificity or selectivity (stereo-specificity and/or geometric specificity) [1]. A protein engineer would like to understand how the specificity can be altered – what residues can be altered and which residues need to be conserved to maintain catalysis. In the case of GpdQ, it

was known that it had a low level of activity to a wide range of substrates – would evolution alter residues on the periphery of the active site so or would resides in the centre of the active site be altered? These are the type of questions that we would like to address in this thesis.

The active sites of most enzymes are usually contained within a single domain, but in many enzymes the active site at the interface between two domains so that changes in enzyme activity may also involve changes in the oligomeric structure of the enzyme [5, 6]. The catalytic properties of an enzyme may depend upon the structure of its active site as well as the manner in which the subunits are arranged.

There are two models that describe enzyme specificity: the lock-and-key model and the induced fit model. In the lock-and-key model, Emil Fischer likened the active site of an enzyme to a lock and the substrate molecule to a key: their shapes are rigid, geometrically complementary and fit exactly into each other [7]. This model is an adequate description of many enzymes. In particular hydrolytic enzymes usually have active sites that conform to the lock and key model and usually obey Michaelis-Menten kinetics. GpdQ was expected to conform to this model.

Highly regulated enzymes or enzymes that need to exclude water from the active site usually do not conform to this model. Instead, the behaviour of these enzymes can be explained by the induced fit model proposed by Koshland. In this model, the active site of an enzyme is flexible and undergoes modifications upon substrate binding and is finally moulded to the shape of the substrate molecule to catalyse a reaction [8]. Enzymes that undergo induced fit are usually highly regulated and do not exhibit Michaelis-Menten kinetics – they are often described as cooperative enzymes. Figure 1.1 shows the Michaelis-Menten plots for three forms of cooperativity in enzyme kinetics – positive, negative and no cooperativity. Non-cooperative enzymes give a hyperbolic Michaelis-Menten plot.

Michaelis-Menten plots of positively cooperative enzymes are sigmoidal in shape while negatively cooperative ones are 'biphasic', where the velocity continues to increase at high substrate concentrations.

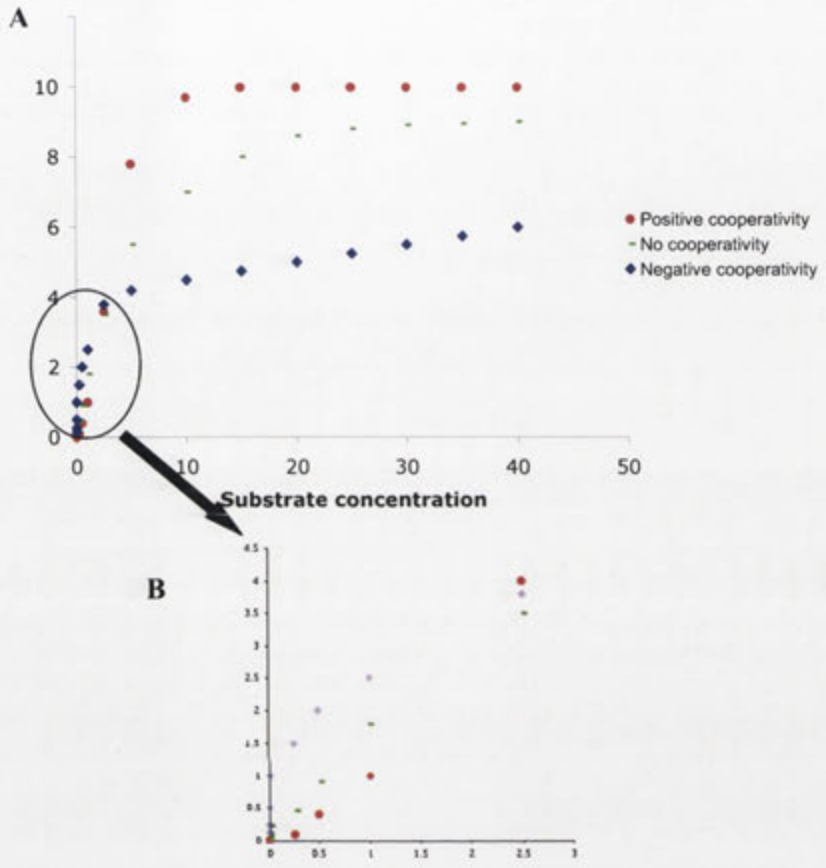


Figure 1.1: Michaelis-Menten plots of cooperative/non-cooperative enzymes. Michaelis-Menten plots of three enzymatic behaviours that exhibit positive, negative and no cooperativity respectively. The three curves at low substrate concentrations (circled in **A**) are magnified in **B**.

During the course of this study, deviations from Michaelis-Menten kinetics were observed in some GpdQ mutants and attempts were made to understand the molecular basis of this behaviour. The Hill equation, a common method to describe cooperativity was used to analyse these non-Michaelis-Menten kinetics, as will be described in Chapters 2 and 3.

1.1.2 Metalloenzymes

GpdQ contains two metal ions as co-factors and is a metalloenzyme. This class of enzymes is very common; more than one third of all known enzymes require metal ions for enzyme activity. A database dedicated to metalloproteins was set up in recent years to offer scientists information about the geometrical parameters of metal-binding sites in protein structures available from the Protein Data Bank (http://metallo.scripps.edu/) [9].

Residues that coordinate the metal ions are known as ligands. Amino acids that regularly act as metal ligands are cysteine, histidine, aspartic and glutamic acids as well as their amides, and tyrosine. The functional groups of these residues form a shell of hydrophilic atomic groups of oxygen, nitrogen and sulfur atoms, where metal ions can bind comfortably [10]. Metals that are commonly found in hydrolytic enzymes like GpdQ are transition metals like Co^{2+}, Zn^{2+}, Fe^{2+}, Fe^{3+}, Mn^{2+}; alkaline earth metal ions like Mg^{2+} that have a tendency to bind less tightly [1]. A review of the coordination chemistry relevant to metalloenzymes is available [11].

Metal ions can have at least two functions in hydrolytic reactions. They can act as Lewis acids and activate nucleophiles for attacking chemical bonds in the substrate [1]. Metal ions are also known to participate in catalysis by interacting with the substrate (binding) and ensuring that it is oriented correctly for catalysis [12].

Metalloenzymes are generally thought of as binding their metal with high affinity. Indeed this is frequently the case. In fact some proteins have demonstrated the need for metal binding to occur prior to their folding, thus playing a structural role. There are also examples of metal ions controlling conformational changes to the active site [13-15]. However, metal ions do not always bind tightly to an enzyme and they do not always bind prior to the substrate. For example, metals are essential for DNA polymerase

activity, but the metals are not bound tightly to the protein in the absence of the DNA substrate [16, 17]. Clearly, the affinity of an enzyme for metals is important for its function and how this affinity can be modulated is of considerable interest to the protein engineer. In the case of GpdQ I have tried to understand how the binding affinities of the metal ions can be modulated by mutating metal ligating residues and measuring changes in activity.

1.2 ORGANOPHOSPHATE COMPOUNDS

1.2.1 The chemistry of organophosphate compounds

Organophosphate compounds (OPs) of relevance to this thesis are esters of phosphoric acid (H_3PO_4). The term also applies loosely to sulfur analogues of phosphoric acid as well as phosphonic acids. Phosphoric acid is triprotic, meaning it contains three potential protons to donate. Therefore, phosphoric acid can be esterified from one to three stages, forming a monoester, diester or triester, respectively. Structures of the main classes of organophosphates are shown below.

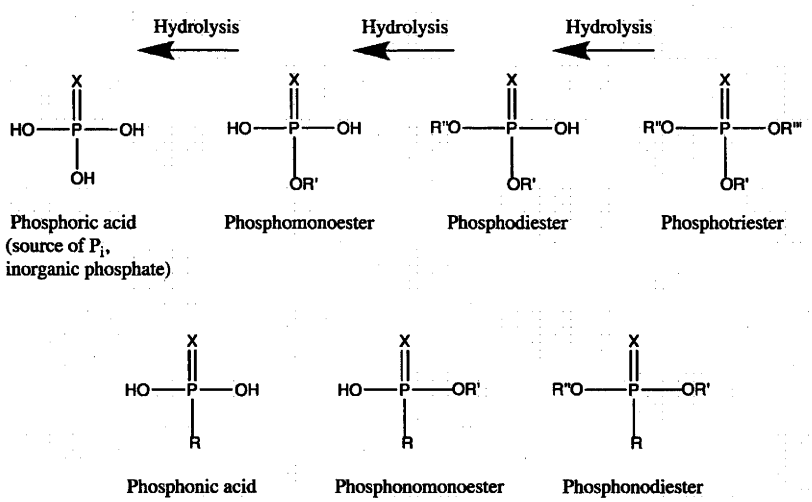


Figure 1.2: Chemical structures of the common classes of OPs.
X can be oxygen or sulfur while R, R' and R" are alkyl, alkenyl, alkynyl, aryl and even phenol groups.

1.2.2 Organophosphate compounds and their uses

Phosphotriesters

Organophosphate triesters are synthetic and do not occur naturally in the environment. They are used extensively in the production of insecticides and chemical warfare agents. In both applications, OP triesters act as potent nerve agents by inhibiting acetylcholinesterase (AChE) and ultimately causing death. AChE is a serine esterase that catalyses the hydrolysis of the neurotransmitter acetylcholine into choline and acetic acid [1, 18, 19]. Figure 1.3 shows examples of OP triesters that are commonly used in pesticides and chemical warfare agents.

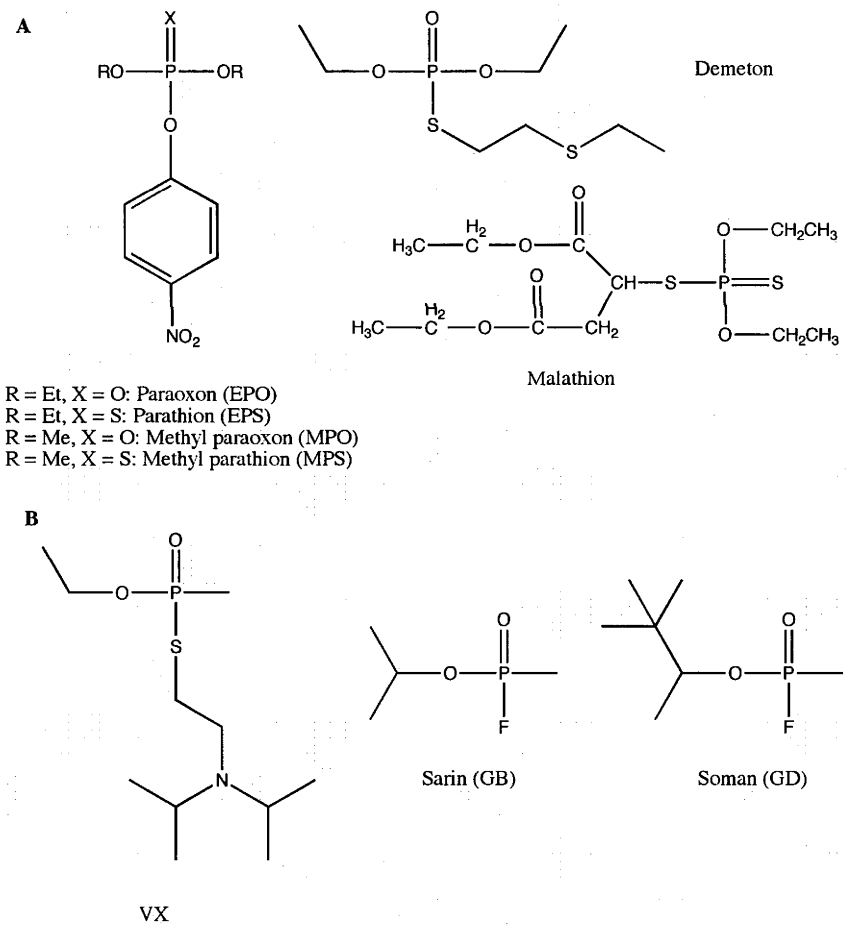


Figure 1.3: Examples of phosphotriesters.
Examples of phosphotriesters that are commonly used on pesticides (**A**) and chemical warfare agents (**B**).

Phosphodiesters and phosphomonoesters

While phosphotriesters are not naturally occurring compounds, diesters and monoesters are abundant in nature and have play important biochemical roles, including in genetic materials like DNA and RNA (Figure 1.4), enzyme co-factors like ADP, intracellular energy transfer (ATP) and as intermediates in biochemical reaction pathways [19]. Phosphodiesterases that specifically react upon the phosphodiester bonds in DNA are known as nucleases.

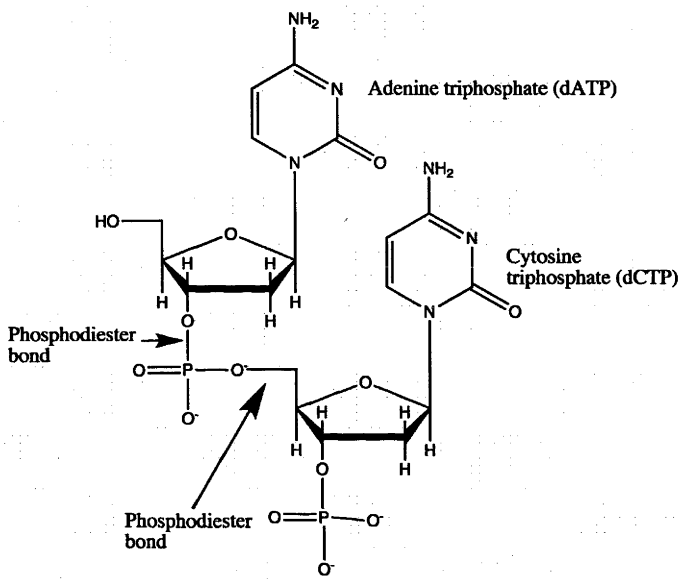


Figure 1.4: The phosphodiester bonds in DNA.

The phosphodiesters that are of interest in this study are those from the direct breakdown of triesters or have the same leaving group as the triesters, for example dimethyl phosphate (DMP), diethyl phosphate (DEP) and bis (*p*-nitrophenyl) phosphate (b*p*NPP) (Figure 1.5).

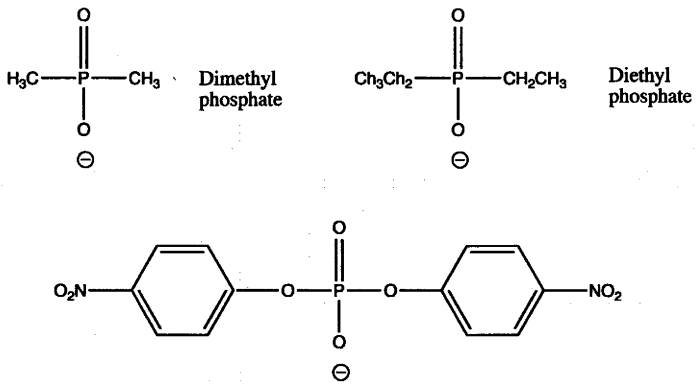


Figure 1.5: Examples of phosphodiesters.
Chemical structures of the ionised forms of dimethyl phosphate, diethyl phosphate and bis (*p*-nitrophenyl) phosphate.

1.2.3 The toxicity of organophosphate triesters

As discussed earlier, the popularity and commercial success of OP triesters are mainly due to their inherent ability to deactivate AChE. Unfortunately, the toxicity of OP triesters extends beyond the targeted organisms to pose serious threats to the environment, ecology and health [18-20]. As a result, many scientific research groups have set out to develop methods to degrade or to remove OP triesters from the environment.

Although the use of OP triesters as pesticides and nerve agents has a history of less than a century, some bacteria have evolved enzymes that are capable of hydrolysing a wide range of these compounds. These phosphotriesterases (PTEs) include the identical organophosphate hydrolases from *Flavobacterium sp.* [21] and *Pseudomonas diminuta* [22], organophosphorus acid anhydrase (OPAA) from *Altermonas sp.* [23] and OpdA from *Agrobacterium radiobacter* P230 [24]. OpdA and OPH have ~90% sequence identity but the former has additional polypeptide chain 20 amino acids in length at the C-terminus and there are three other amino acid differences located in the substrate binding pocket [25, 26]. In the Ollis

laboratory, the C-terminus extension was deleted from the *opd*A gene so that the expressed enzyme would have an increased similarity to OPH [27].

All these PTEs, due to their individual substrate specificities, catalyse the hydrolysis of different phosphotriesters with various efficiencies. For instance, OpdA activities with triesters like methyl derivatives of paraoxon and parathion are nearly diffusion limited whereas its methyl chloropyrifos activity is relatively low [28]. Various research groups, including the Ollis laboratory, have set out to engineer their respective PTEs towards higher activity with poor substrates, since PTEs provide a promising OP detoxification and bioremediation technology that is both efficient and safe [29-32].

1.3 GpdQ - A NOVEL GLYCEROPHOSPHODIESTER PHOSPHODIESTERASE FROM *ENTEROBACTER AEROGENES*

This section 1.3 is a summary of previous work on GpdQ, including the isolation, purification, cloning of the *gpd*Q gene and characterisation of the enzyme activity.

1.3.1 Growth of *Escherichia coli* co-expressing OpdA and GpdQ using paraoxon as the sole phosphorus source

In 2003, Dr Sean Yu-McLoughlin created an *Escherichia coli* system that was capable of utilising dimethyl paraoxon (MPO), a phosphotriester, as the sole phosphorus source. The *E. coli* system would go through the following proposed catabolic pathway (Figure 1.6) for the breakdown of MPO to phosphate, P_i [32].

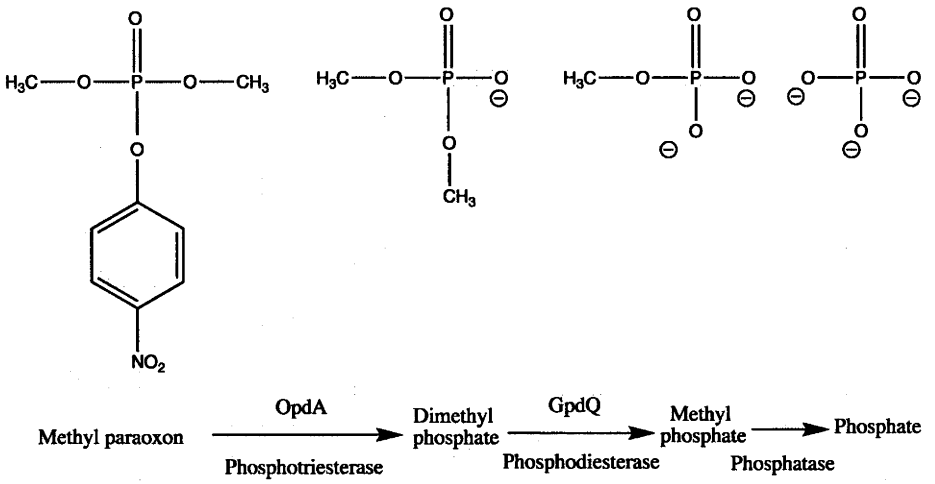


Figure 1.6: The proposed catabolic pathway for the breakdown of methyl paraoxon (MPO).
This rationally designed pathway was for the mineralisation of MPO in an *E. coli* that co-expresses recombinant OpdA and GpdQ. Phosphatases such as alkaline phosphatase are naturally expressed by *E. coli.*

E. coli transformed with the recombinant *opd*A gene would express the PTE to readily degrade MPO. However, it had been previously reported that *E. coli* was unable to use dimethyl phosphate (DMP), the hydrolysis product of MPO, as its sole phosphorus source [33]. Therefore, a phosphodiesterase from an external source that is capable of hydrolysing DMP was sought.

Enterobacter aerogenes was reported to be able to survive on DMP as its sole phosphorus source [34]. Gerlt and co-workers had isolated and characterised the phosphohydrolase that was responsible for DMP hydrolysis [35, 36]. The enzyme was found to be hexameric with a subunit molecular weight of 29 kDa. The phosphohydrolase was also found to be capable of hydrolysing a range of monoesters and diesters of phosphoric acid.

Yu-McLoughlin *et al.* isolated the gene for this enzyme from the chromosome of a strain of *E. aerogenes* obtained from a wallaby. The genomic DNA was isolated and digested into fragments sized between 2 and 10 kb with the restriction endonuclease *Sau*3AI. The DNA fragments were then cloned into plasmids. The library of *E. aerogenes* genomic DNA

fragments was used to transform DH10β *E. coli* cells that were then selected on minimal medium supplemented with 0.1 dimethyl phosphate (DMP) as the sole phosphorus source. A single colony of *E. coli* grew significantly larger than the rest. The plasmid was isolated and characterised as described in the next section. This new plasmid was later used to transform fresh DH10β *E. coli* cells that already bore the *opd*A gene. These cells were then tested positive for growth on medium with dimethyl paraoxon (MPO) as the sole phosphorus source. The ability of the cells to grow on MPO as its sole phosphorus source would signify the success of the catabolic pathway outlined in Figure 1.7, where the breakdown of methyl paraoxon (MPO) occurred in three catalysed hydrolysis steps. OpdA catalysed the hydrolysis of MPO to DMP that then served as a substrate for GpdQ to be hydrolysed to release methyl phosphate, a monoester. *E. coli* readily expressed phosphomonoesterases with the likes of alkaline phosphatase (encoded by the *pho*A gene) [37] that would hydrolyse methyl phosphate to phosphate, that could be directly utilised by the bacteria as an energy source.

1.3.2 Characteristics and properties of GpdQ

The phosphohydrolase gene was isolated as part of a putative operon

The plasmid from the single *E. coli* colony that grew using DMP as the sole phosphorus source was isolated and studied. The presence of a 5.6-kb fragment was revealed and its sequence was determined. BLAST analysis of the sequence indicated that the fragment was partially homologous to glycerol-3-phosphate (G3P) uptake (*ugp*) operon of *E. coli* (Figure 1.7) [32]. The *ugp* operon contained four genes (A, B, C and E) that code for an ABC transporter that utilises the energy of adenosine triphosphate (ATP) hydrolysis to transport molecules across the cell wall [38, 39]. The *E. aerogenes* DNA fragment could be divided into the promoter region and five putative open reading frames (ORFs) four of which were homologous to proteins in the *ugp* operon of *E. coli* as indicated in the figure below (Figure 1.7). Additionally, a nucleotide sequence upstream of the putative *E.*

aerogenes ugp operon has a 78% sequence match to the *E. coli* phosphate box, which is where the transcription activator PhoB binds when phosphate is limited [32, 40, 41].

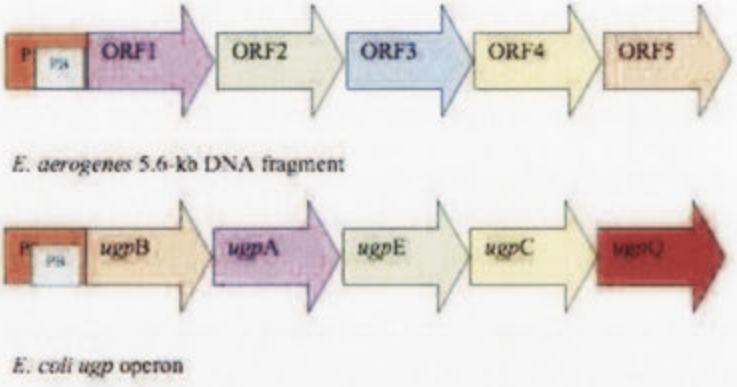


Figure 1.7: Comparison of the 5.6-kb DNA fragment isolated from *E. aerogenes* and the *E. coli ugp* operon.
ORFs in the *E. aerogenes* DNA fragment and their respective homologous genes in the *E. coli ugp* operon are shaded in the same colours, *eg.* ORF1 and *ugp*A are shaded in purple. P indicates the promoter region of each operon while PB indicates phosphate box. ORF3 is the 825-bp long *gpd*Q gene.

The remaining gene in the ugp operon codes for UgpQ, a glycerophosphodiesterase (GDPD) that had a high level of sequence similarity to GlpQ, a periplasmic GDPD whose sequence and structure is known [42-44]. GlpQ was one of more than 10 bacterial GDPD proteins whose structures were to be found in the Protein Data Bank (http://www.rcsb.org/pdb/). The GDPDs from *Thermotoga maritima* [45], *Agrobacterium tumefaciens* [46] and *Thermoanaerobacter tengcongensis* [47] and a homolog of GlpQ found in *Borrelia hermsii* [48] all adopt a central $(\alpha/\beta)_8$ barrel tertiary fold with the conserved sequence motif $HR(X)^nEXD(X)^nH(X)^nEXK$ that coordinates metals in the active site. The structure of UgpQ, the cytosolic GDPD in *E. coli* has not been solved but multiple sequence alignments with other GDPDs predict the protein to assume a similar TIM barrel structure [49]. An example of a GDPD-catalysed hydrolytic breakdown of glycerophosphodiesters that typically yields glycerol-3-phosphate (G3P) and an alcohol is shown in Figure 1.8.

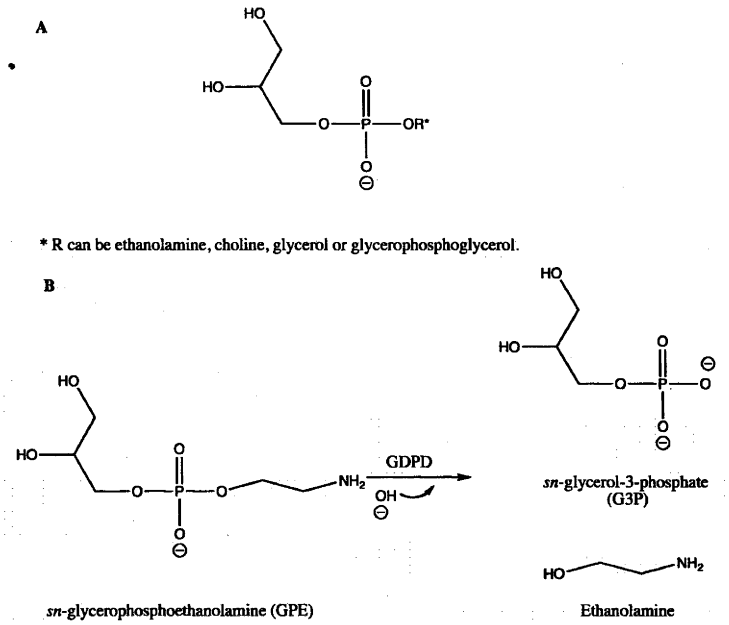


Figure 1.8: Glycerophosphodiesters and their hydrolysis.
(A) The general chemical structure of glycerophosphodiesters; **(B)** Hydrolysis of *sn*-glycerophosphoethanolamine (GPE) catalysed by GDPDs.

ORF3 of the *E. aerogenes* DNA fragment had 825 base pairs and encoded a 30.8-kDa protein that was designated GpdQ. No sequence similarity was detected between GpdQ and UgpQ. Instead, it was found to exhibit sequence similarity to a wide range of phosphatases, including 5' nucleotidase from *E. coli*, Ser/Thr protein phosphatase 1 from rabbit, kidney bean purple acid phosphatase and rat protein acid phosphatase. These phosphatases are functionally diverse yet have similar structures. They fold into what is called an α/β sandwich motif that contains a binuclear metal centre in their active site. Multiple alignment of the amino acid sequences of GpdQ and four examples of these dimetallophosphatases reveal the conserved sequence motif of $\underline{D}XH(X)^nG\underline{D}(X)^nG\underline{N}HD/E(X)^n\underline{H}(X)^n\underline{H}X\underline{H}$, where residues that coordinate the two metal ions have been underlined [32, 50]. The binuclear metal center at the active sites of these enzymes is capable of coordinating a wide range of divalent metal ions including Mn^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+}, and also Fe^{3+} at the site with higher metal binding affinity. Subsequent over-expression of the GpdQ protein showed, as expected, activities with phosphodiester and phosphomonoester substrates such as *sn*-

glycerophosphoethanolamine (GPE), DMP, bis (*p*-nitrophenyl) phosphate (b*p*NPP) and *p*-nitrophenyl phosphate (*p*NPP). Additionally, the protein exhibited very low activity with paraoxon, a phosphotriester.

GpdQ, a highly promiscuous enzyme

GpdQ is a catalytically promiscuous enzyme with an exceptionally broad substrate specificity. All other known GDPDs generally show activity with a range of GPDs [45-49]. With GpdQ, the enzyme's substrate repertoire is more extensive than the majority of GDPDs. GpdQ has primarily phosphodiesterase activity, capable of hydrolysing the phosphodiester bonds in simple alkyl phosphodiesters like dimethyl phosphate (DMP) and diethyl phosphate (DEP), other model phosphodiester substrates like bis (*p*-nitrophenyl) phosphate (b*p*NPP) and *O*-ethyl (*p*-nitrophenyl) phosphate (*O*-Et-*p*NPP). Additionally, GpdQ shows low activity towards phosphomonoesters, phosphotriesters, phosphonate monoesters and notably, an analogue of EA 2192, a highly toxic and stable hydrolysis product of the nerve agent VX [32, 51, 52].

1.4 THE STRUCTURE OF GpdQ

Three crystal structures of native GpdQ have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb/). Jackson *et al.* first solved the structure in 2006 (PDB accession codes 2DXL and 2DXN) at 2.9 Å and 3.0 Å respectively [53]. Another structure of a higher resolution (1.9 Å) was obtained from this study with the PDB code 3D03 [52]. Structural resolution analysis confirmed the α/β sandwich tertiary fold.

1.4.1 Overall quaternary structure

Native GpdQ is homo-hexamer that is made up of six monomers of ~30.8 kDa each. The structure of GpdQ was initially determined with crystals

belonging to the $P2_13$ space group. The GpdQ molecule was found to have $D3$ (dihedral hexamer) symmetry like aspartate carbamoyl transferase.

The quaternary organisation of GpdQ can be described as a trimer of dimers:

i. Two subunits are related by a two-fold axis and are linked by strong interactions including an inter-chain disulfide bond, to form a dimer (Figure 1.9). The GpdQ dimer has a C_2 symmetry.

ii. Three dimers form a hexamer. The hexamer is held together by weak interactions.

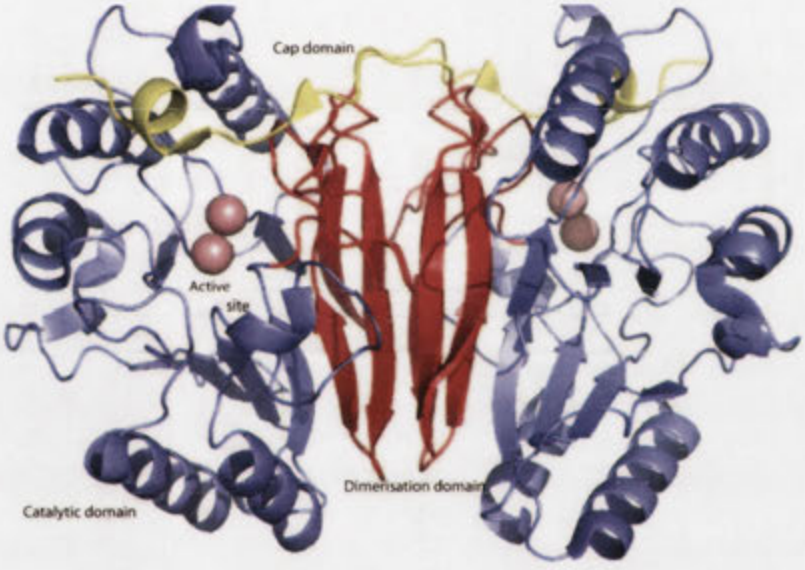


Figure 1.9: The GpdQ dimer.
The GpdQ displays a C_2 symmetry. Here, the catalytic domains are coloured in blue, the dimerisation domains in red, the cap domains in yellow, the Co^{2+} metal ions in pink.

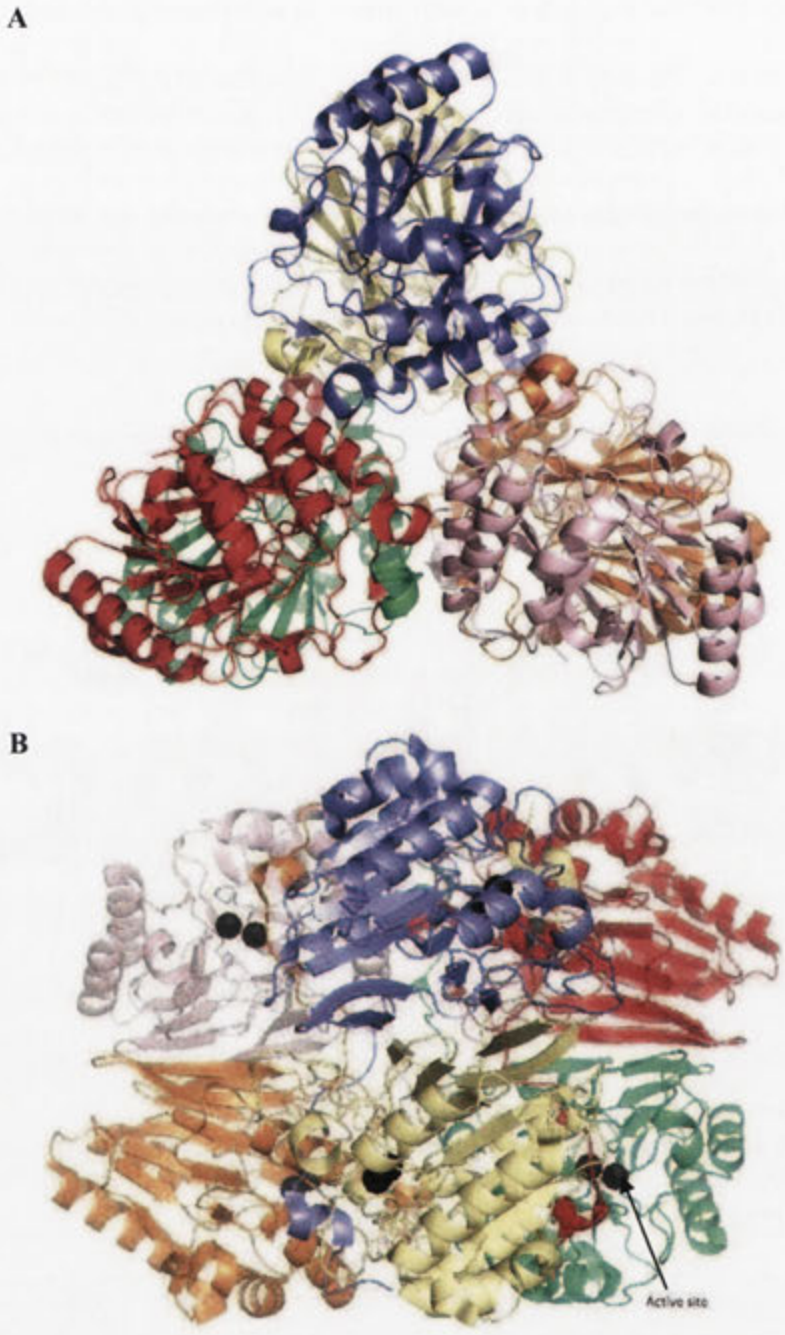


Figure 1.10: **The GpdQ hexamer.**
Ribbon diagrams showing the *D*3 symmetry of the GpdQ hexamer as viewed from the 3-fold axis (**A**) and the perpendicular 2-fold axis (**B**). Different colours are used to highlight the six chains that form the hexamer. In (**B**), metal ions (black spheres) are displayed to highlight the location of the active sites in the GpdQ hexamer.

Protein oligomerisation

It is widely believed that protein oligomerisation happens as a result of the evolutionary process. The importance of protein oligomerisation is evidenced by its prevalence in nature [54-56]. It has been proposed that protein oligomerisation plays many roles that include foremostly, the functionality of a protein. Oligomerisation enables allosteric regulation involving cooperativity between subunits, a complete active catalytic site between subunits, or binding of substrate/product/effector at an additional regulatory site that exists between subunits. Oligomerisation also contributes to conformational and thermal stability, thus making the protein less susceptible to degradation. The reduced surface area also protects the protein from degradation and is thought to enhance the diffusion of substrates to the enzyme active site. Additionally, the increased complexity of an oligomer allows a more efficient control over the accessibility and specificity of the active site [54-57].

On the molecular level, oligomerisation reduces translation errors by providing an extra proofreading step in rejecting defective subunits. Moreover, association of homo-oligomers provides a compact and efficient way for large proteins to be encoded, thus minimising the genome size. In other words, only a small genetic or DNA space is required to encode small protein subunits which will later self-assemble to form a large protein complex [54-57].

Oligomerisation results in the burial of surface area of each protein monomer. For example in GpdQ, the surface area of a monomer (12886.6 $Å^2$) is reduced by ~23% in a dimer and ~32% in a hexamer, respectively. These values were calculated using the software AREAIMOL in the CCP4 Program Suite using a 1.4-Å probe that was about the size of a water molecule. This method was based on the formula devised by Lee and Richards [58].

Dissociation of protein oligomers

Protein oligomers are also known to dissociate, along with a concomitant change in enzyme activity. The process is often reversible; therefore these enzymes are able to re-associate to form the oligomer, as influenced by regulatory effectors [57]. Dissociation could lead to both positive and negative changes towards toward the enzyme activity, depending on individual enzymes.

It has also been shown that dilution has an effect on the specific activities of purine nucleoside phosphorylase with 50 μM inosine and 50 mM phosphate, respectively [59]. It is also likely for enzyme concentration to shift the equilibrium of the associated/dissociated enzyme mixture.

Despite the numerous advantages that oligomeric proteins have over their individual subunits, monomers also have a few functional niches. Many transfer and transport proteins like Cytochrome c are small and they diffuse rapidly to their sites of action within the cell. Small proteins are also more stable at low concentrations. Furthermore, unwanted oligomerisation of proteins can result in the formation of pathogenic aggregates [54, 55, 60].

Advances in protein engineering have enabled enzymes with stable quaternary structures to dissociate [61, 62], including GpdQ in this thesis. A brief review on protein engineering is given in section 1.6 of this chapter.

1.4.2 Structural domains in a GpdQ monomer

Jackson *et al.* divided the 274-residue GpdQ monomer into the following three domains [53] (Figures 1.11 and 1.12):

i. Catalytic (residues 1-195)
ii. Dimerisation (residues 196-255)
iii. Cap (residues 256-274)

These domain definitions are consistent with SCOP classifications [63].

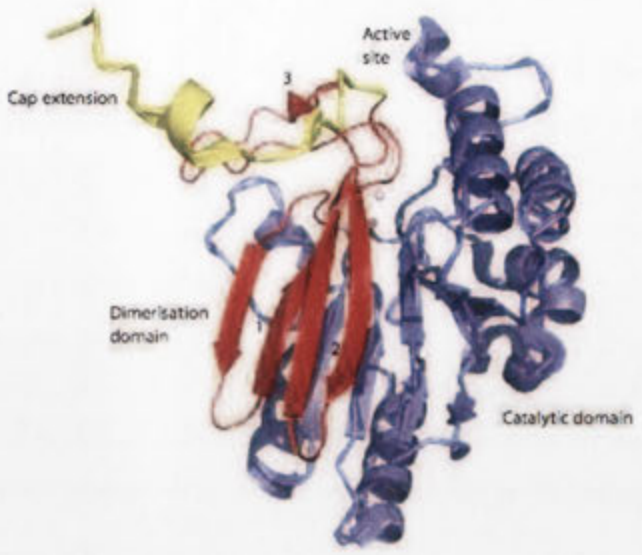


Figure 1.11: The GpdQ monomer.
Ribbon representation of a GpdQ monomer, with the three domains in different colours for clarity reasons. Numbers 1 and 2 indicate the pair of two-stranded anti-parallel β-sheets in the dimerisation domain while 3 shows the third shorter strand.

Crystal structures showed that the catalytic domain exhibited α/β or αββα sandwich fold that had been predicted earlier through the amino acid sequence analysis of GpdQ. This means GpdQ is a structurally novel GDPD [53]. Other bacterial GDPDs that have been previously characterised are either dimeric or monomeric [45, 47, 64, 65], with the exception of the GDPD from *A. tumafaciens*, which is a hexamer. However, each chain in the hexamer adopts the classical TIM barrel fold [46].

The other two additional domains, namely the dimerisation domain and the cap domain, contribute to the quaternary structure and the *D*3 symmetry of the hexameric enzyme.

1.4.3 The conserved catalytic domain, active site and binuclear metal center

The catalytic domain exhibits the conserved α/β sandwich structural motif; the active site and the binuclear metal center of the enzyme lie within this domain. Apart from the tertiary structure of the catalytic domain, the active

site residues of this superfamily of α/β sandwich phosphohydrolases are also conserved, including GpdQ and kidney bean purple acid phosphatase [66]. Additionally, two phosphodiesterases, whose structures have been solved, are known to adopt this α/β sandwich and the active residues are conserved too. The two proteins are the Rv0805 Class III 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* [67] and the and the MJ0936 phosphodiesterase from *Methanococcus jannaschii* [68]. The binuclear metal center at the active sites of these enzymes is capable of coordinating a wide range of divalent metal ions including Mn^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+} and also Fe^{3+} at the site with a higher metal binding affinity.

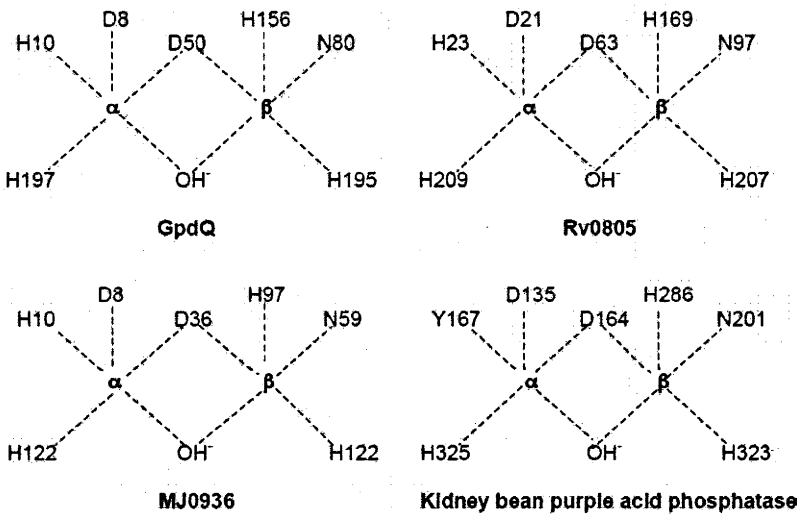


Figure 1.12: The active sites of α/β sandwich dimetallo-phosphohydrolases. Simplified diagrams showing ligand coordination geometries at the active sites of GpdQ, Rv0805, MJ0936 phosphodiesterases and kidney bean protein acid phosphatase.

The three phosphodiesterases, GpdQ, Rv0805 and MJ0936 consist of the same metal ion ligands: an aspartate and two histidines at the α-site; an asparagine and two histidines as the β-site as well as another aspartate residue acting as the bridging ligand for the two metal ions. The active site of kidney bean purple acid phosphatase differs in one residue: instead of two histidines at the α-site it has a tyrosine, which connects to the Fe^{3+} ion via a charge transfer, thus giving the enzyme the characteristic purple colour

[66]. The active sites of other related phosphatases like the Ser/Thr protein phosphatase from bacteriophage λ and 5'-nucleotidase from *E. coli* also vary slightly [69, 70].

Figure 1.13 shows the active site architecture of GpdQ.

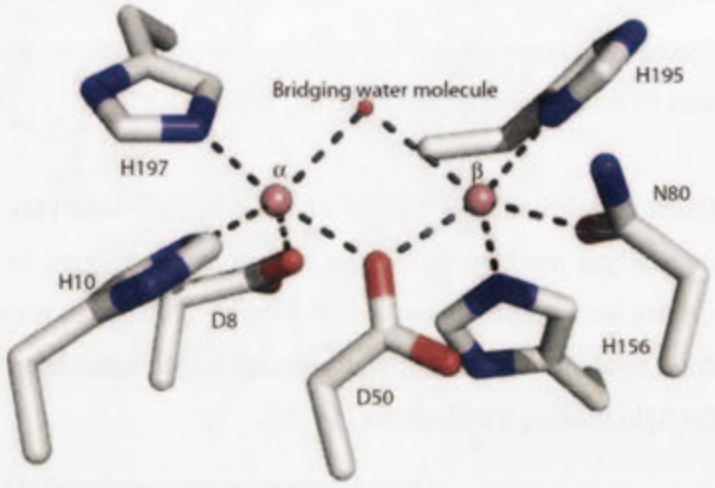


Figure 1.13: The active site of GpdQ.

It has been shown that the active site of GpdQ was able to coordinate Mn^{2+}, Co^{2+}, Zn^{2+}, Cd^{2+} and Fe^{2+} [51-53, 71-73]. The b*p*NPP activities of GpdQ reconstituted with the first four metals have been characterized. The order of activities is $Mn^{2+} > Cd^{2+} > Co^{2+} > Zn^{2+}$.

Non-identical metal binding sites

It is common for metalloenzymes that bind more than one metal ion to have binding sites that are distinguished by differences in their affinities for the metal ions [74]. In many members of the α/β sandwich binuclear hydrolases, including PAPs, and *E. coli* 5'-nucleotidase [75-79], the affinity of the α-site (thus so-called) is always greater than that of the β-site. However, for the bacteriophage λ Ser/Thr protein phosphatase, the affinities of the two sites are reversed as the α-site lacks one ligand, that is the C-terminal histidine [80, 81].

While analysing the structure of GpdQ, Jackson *et al.* observed full and partial metal occupancies at the α- and β-sites respectively [53]. This observation will be discussed in more detail in Chapter 3. The observation agreed well with the ligand coordination geometries at the GpdQ active site: although both α- and β-metal ions are coordinated by two histidine ligands each and jointly, the bridging D50 ligand, the α-metal has another strong acidic ligand in aspartate (D8) whereas the β-metal is additionally coordinated by an asparagine residue (N80) (Figure 1.13).

The different metal binding affinities of the α- and β-sites play a pivotal role in the catalytic mechanism of binuclear metallohydrolases. In PAPs for instance, it has been proposed that the loose-binding metal ion is responsible for substrate binding [77, 82, 83] while the nucleophile has been shown to bind to the tight binding trivalent one [84, 85].

1.4.4 The dimerisation domain

The dimerisation domain of GpdQ consists of a pair of two-stranded, anti-parellel β-sheets and a third shorter strand between them. The shorter strand is not directly involved in the dimeric interface between two GpdQ subunits. (Figure 1.11) [53]. This domain is made up of approximately 60 residues (residues 196-255) and a significant number of them are large aromatic residues like trytophan, tyrosine and phenylalanine, which are favoured in β-sheet secondary structures as well as other non-polar residues like methionine and valine [86]. The predominance of these non-polar residues creates a hydrophobic core within the dimeric interface (Figure 1.14).

Each GpdQ subunit most likely folds into a globular subunit and the protein-protein interface is formed by very specific interactions (*eg.* hydrogen and hydrophobic) between many residues on the surface of the domain. This is evidenced by exclusion of water at the interface (Figure 1.14) and in the figure showing the electrostatic potential of the GpdQ

monomer (Figure 1.15). Specific interacting residues are typically complementary in shape and chemical properties (Figures 1.14 and 1.16).

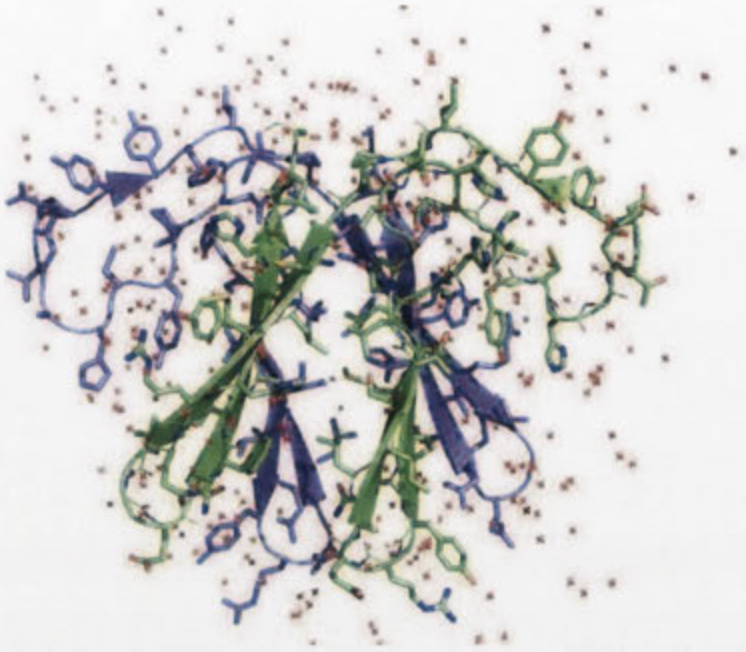


Figure 1.14: The dimerisation domain of GpdQ.
A diagram of the dimerisation domain of GpdQ, highlighting the lack of water molecules (pink dots) in the hydrophobic core of the interface (centre) and the predominance of non-polar and aromatic amino acids in the domain. The dimeric interface results in 1667 $Å^2$ of the total surface area of each chain [53].

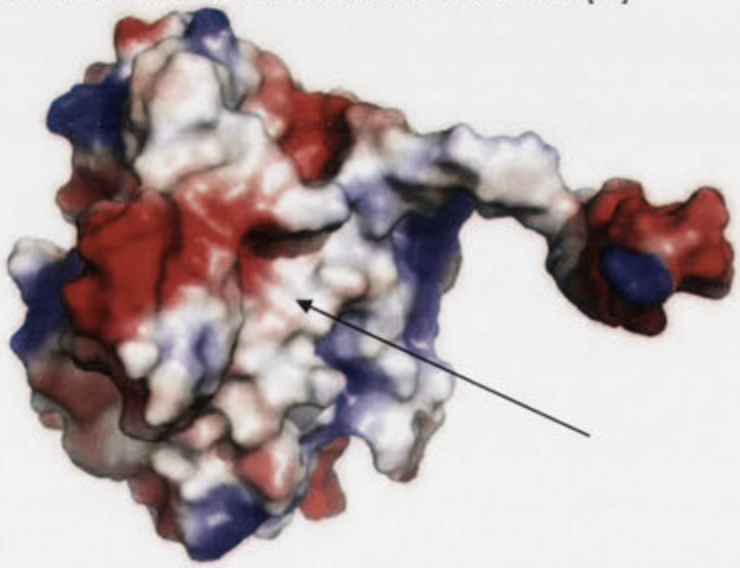


Figure 1.15: Electrostatic potential mapped to the molecular surface of the GpdQ monomer.
Red indicates negatively charged surface; blue, positively charged; white uncharged. The arrow points to the hydrophobic core of the dimer interface.

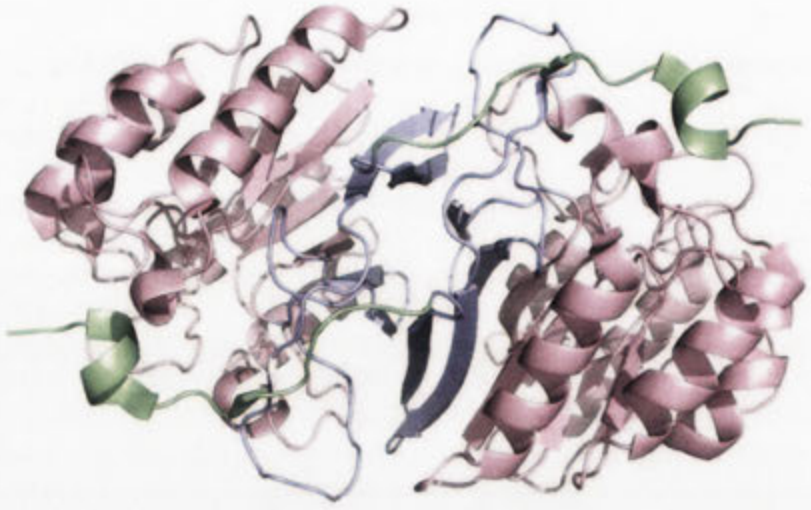


Figure 1.16: The GpdQ dimer and the intertwining dimerisation domains.
A diagram of a GpdQ dimer demonstrating how the β-strands donated from the dimerisation domain of each monomer (in blue) form a barrel at the interface. Catalytic and cap domains are in pink and green respectively.

1.4.5 The cap domain

The cap domain of GpdQ is made up of the last 19 (256-274) residues of the C-terminus. This domain is a relatively small piece of peptide with a short β-sheet followed by a short α-helix. Despite its size, the cap domain plays important structural roles in GpdQ and is also possibly involved in the catalytic mechanism of the enzyme.

Formation of the GpdQ dimer through 3D domain swapping

Domain swapping is a common structural phenomenon in protein dimers. This mechanism is defined by the replacement of one portion or domain of a protein subunit by the same domain from another identical chain [87, 88]. In general, exchanged domains can vary from a single β-strand to a globular, tertiary structural domain that may cause a sizable burial area in the 'closed' interface [87, 89].

Domain swapping in GpdQ is formed through C-terminal the cap domain. Domain swapping via C-terminal extensions has been reported in proteins

like the bovine pancreatic ribonuclease (RNAse A), human alpha-crystallins and the capsid of the HIV-1 virus [90-93].

This cap structure in GpdQ extends from the dimerisation domain and sits above the active site of the other dimerising chain. The stability of this domain is strengthened by an inter-chain disulfide bond (between C54 of one monomer and C269 of the other monomer) (Figure 1.16). Intramolecular disulfide bonds play an important role in regulating protein activity, as well as maintaining the stability and folding of the tertiary structure [94]. Intermolecular disulfide bonds, on the other hand, usually take part in the quaternary structure of a protein, by stabilising the assembly of subunits to form an oligomer [95-97].

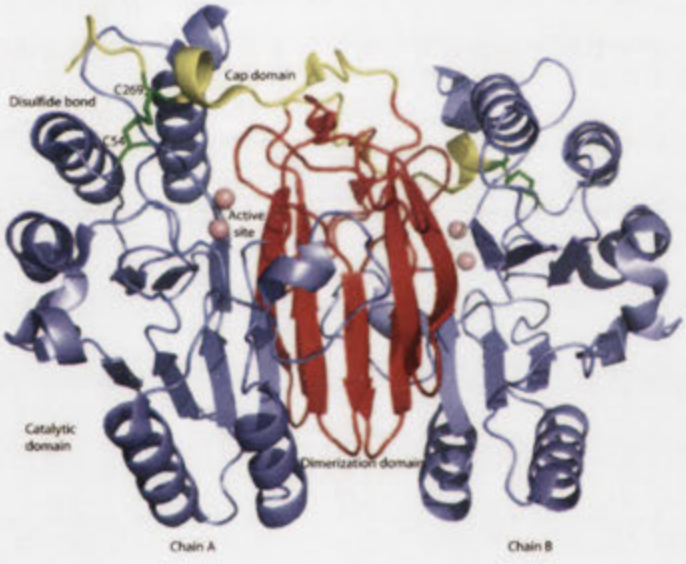


Figure 1.17: The domain swapping mechanism by the cap domain and the inter-chain disulfide bond.
The cap domains are highlighted in yellow while the disulfide bond is in green.

Apart from the inter-chain disulfide bridge, there are also symmetric interactions between N53 and R56 of one chain with S268, the backbone oxygen of P270 of another chain within the same dimeric unit (Figure 1.17).

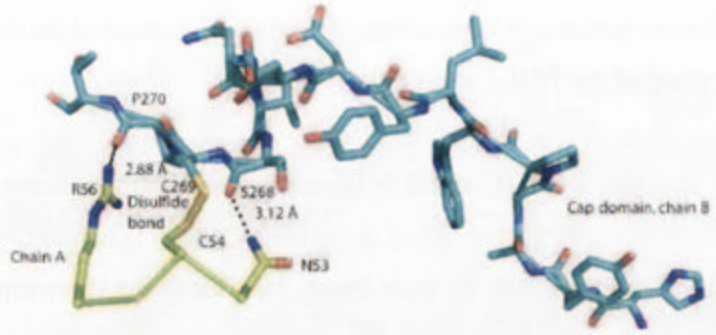


Figure 1.18: Inter-chain interactions by the cap domain of GpdQ.
Diagram showing the three inter-chain bonds contributed by the cap domain of GpdQ towards the formation of the dimer.

Formation of the GpdQ hexamer

In the GpdQ hexamer, the cap domains of all the six chains are located at the interfaces between one GpdQ dimer and another (Figure 1.18). The hexamer is held together weakly with electrostatic interactions, hydrogen bonds and an inter-chain hydrophobic patch that is explained in detail in Chapter 5.

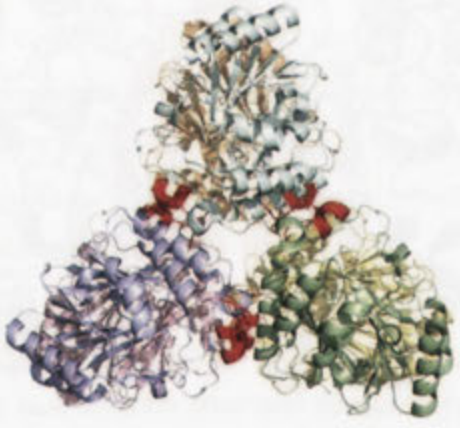


Figure 1.19: The GpdQ hexamer with the cap domains highlighted.
The cap domains of all the six subunits in red colour to highlight their locations at the interface between the three dimeric units. The interaction between the three dimeric units results in the burial of 1131 Å^2 of the total surface area of each chain [53].

Controlling access into the active site

A cap domain structure is also present in several other phosphohydrolases from both the TIM barrel GDPD and (α/β) sandwich dimetallo-phosphoesterase superfamilies of enzymes. This domain, which is additional

to the structurally conserved catalytic domain (primary domain), is often smaller and located near the active site. Examples of enzymes with this fold include the GDPD from *Thermotoga maritima* [45] (Figure 1.19), periplasmic 5'-nucleotidase from *Escherichia coli* [70] and Mre11 nuclease from *Pyrococcus furiosus* [98].

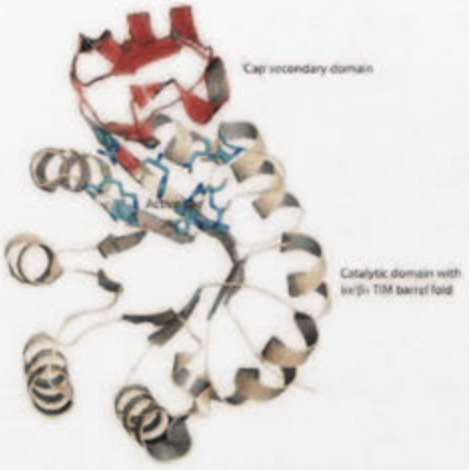


Figure 1.19: The cap domain in a TIM barrel GDPD.
A diagram of the GDPD from *T. maritima* (PDB accession code 1O1Z), with the cap domain coloured in red.

The cap domains of GpdQ (Figure 1.16) and the GDPD from *T. maritima* shown in Figure 1.19 are both located at the entrance to the active site. It is widely proposed that these domains have evolved to constrain the accessibility of the active site in order to enhance the substrate specificity of the enzymes [45, 70, 98]. In particular the complexity of the cap domain seemingly correlates with the size and other chemical properties of the physiological substrate.

The cap sub-domain in GpdQ also differs from the other enzyme examples in several other ways. Foremostly in GpdQ, the structure is a small peptide (a short β-strand followed by one turn of α-helix) made up of approximately the last 20 residues of the C-terminus. In the enzymes shown in Figure 1.19, their cap domains are much larger and they form small globular domains. The cap domain in the GDPD from *Thermotoga maritima* consists of ~50 residues.

Interestingly, a cap domain is not found in the Rv0805 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* [67] and the MJ0936 phosphodiesterase from *Methanococcus janaaschii* [68]. These two phosphodiesterases have been shown previously in section 1.4.3 to be structurally related to GpdQ. The homology exists only between the catalytic domains, where (α/β) sandwich fold and all the active residues are conserved.

1.5 Rv0805 - A 3',5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASE FROM *MYCOBACTERIUM TUBERCULOSIS*

3',5'-cyclic nucleotide phosphodiesterases (PDEs) catalyse the hydrolysis of the phosphodiester bond of second messenger cAMP to yield the biologically inert 5'-AMP, as a form of homeostatic regulation of signal transduction. Previously, this superfamily of enzymes was divided by their amino acid sequences and tertiary structure into two major classes of I and II. A third class was reported by Richter and the catalytic domain of this PDE class III of enzymes closely resembles that of dimetallophosphatases like purple acid phosphatase and calcineurin [99]. Examples of PDEs class III enzymes have been found in *E. coli* (CpdA) [100], *Vibrio fischeri* [101] and *Mycobacterium tuberculosis* (Rv0805) [66, 102].

A search on the DALI server [103] nominated the PDE from *M. tuberculosis*, Rv0805, as the enzyme with the highest sequence similarity to GpdQ. Sequence alignment of Rv0805 [104] and GpdQ using the BLOSUM comparison matrix of the ClustalW multiple alignment program (http://www.ebi.ac.uk/Tools/clustalw) shows approximately 22% similarity (Figure 1.20). All the seven residues coordinating the metal ions are conserved. Superimposition of the X-ray structures of these two enzymes confirmed that the catalytic domains are homologous (Figure 1.21). It should be noted that only the proteolytic fragment of the first 278 out of a

total of 318 residues of Rv0805 were crystallised. Therefore, it cannot be ascertained if the cap domain is present in the complete protein.

```
Rv0805          MHRLRAAEHPRPDYVLLHISDTHLIGGDRRLYGAVDADDRLGELLEQLNQSGLRPDAIVF
GpdQ            -------------MLLAHISDTHFRSRGEKLYGFIDVNAANADVVSQLNALRERPDAVVV
                           :* ******: . ..:*** :*.:   .:::.***    ****:*.

Rv0805          TGDLADKGEPAAYRKLRGLVEPFAAQLGAELVWVMGNHDDRAELRKFLLDEAPSMAPLDR
GpdQ            SGDIVNCGRPEEYQVARQILG----SLNYPLYLIPGNHDDKALFLEYLQPLCPQLGSDAN
                :**:.: *.*  *:  * ::     .*.  *  : *****:* : ::*   .*.:..  .

Rv0805          --VCMIDG--LRIIVLDTSVPGHHHGEIRASQLGWLAEELATPAPDGTILALHHPPIPS-
GpdQ            NMRCAVDDFATRLLFIDSSRAGTSKGWLTDETISWLEAQLFEGGDKPATIFMHHPPLPLG
                   * :*.   *::.:*:* .*  :* :  . :.**  :*   . . : : :****:*

Rv0805          VLDMAVTVELRDQAALGRVLRGTDVRAILAGHLHYSTNATFVGIPVSVASATCYTQDLTV
GpdQ            NAQMDPIACENGHRLLALVERFPSLTRIFCGHNHSLT-------------MTQYRQALIS
                  :*   .  ..:  *. * * ..:  *:.** *  *     1        * * * *

Rv0805          AAGGTRGRDGAQGCNLVHVYPDTVVHSVIPLGGGETVGTFVSPGQARRKIAESGIFIEPS
GpdQ            TLPGT--------------------VHQVPYCHEDTRPYYDLSPASCLMHRQVGEQWVSYQ
                :  **         2         **.*     .      :**...  :   .  ::. .

Rv0805          RRDSLFKHPPMVLTSSAPRSPVD
GpdQ            HSLAHYAGPWLYDENISCPTEER
                :  : :  * :   . :  :
```

Figure 1.21: Amino acid sequence alignment of Rv0805 and GpdQ.

The two large insertions (1, 2) in GpdQ are part of the extensive dimerisation domain.

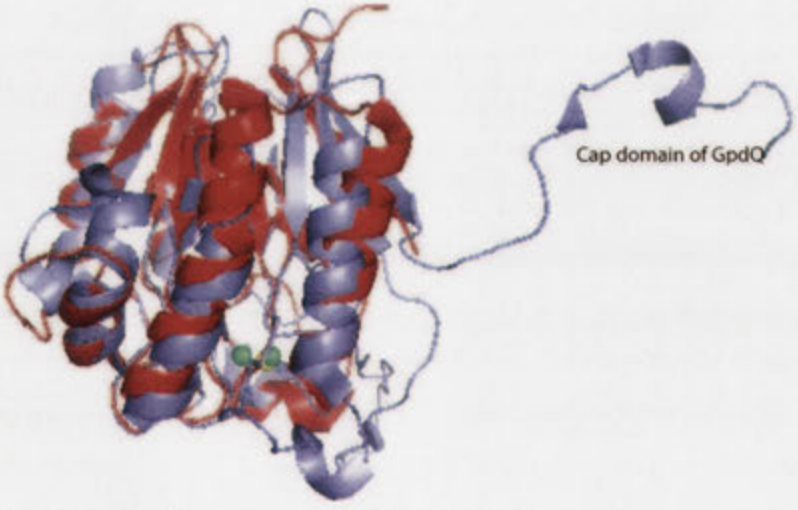


Figure 1.22: Structural alignment of Rv0805 and GpdQ.
Structural superimposition of Rv0805 (red) and GpdQ (blue) monomers (RMSD = 2.298, 179 to 179 atoms).

Throughout this thesis, kinetic and structural comparisons are often drawn between GpdQ site-specific mutants and their equivalents in Rv0805.

1.6 PROTEIN ENGINEERING STRATEGIES

This study is largely concerned with modifying the activity of an enzyme. As a concept, this is fairly easy to understand, but the tools for this work have only been available for a few years and as a consequence our understanding of how to proceed with this type of study is growing rapidly. This type of "protein engineering" offers the potential of being able to convert our knowledge of proteins into products for industry and medicine. It is worth spending some time going over the need for this type of work and the recent history that led to the current activity.

1.6.1 Protein engineering

History and background

Enzymes undergo biological evolution, a very slow process that takes place over times scales of several million years, with mutation, selection, recombination and drifts as the driving forces. It has been suggested that during the course of this enormous length of time that the resultant enzymes should be near perfect – in fact they are probably ideally suited to the requirements of the host organism. However, this does not mean that enzymes are ideally suited for industrial applications. For example, an enzyme that operates in a pathway may be regulated so that it only functions under a specific set of conditions – these conditions may not be appropriate for laboratory or industrial applications. Furthermore, nature can generate a particular level of enzyme activity by increasing (or decreasing) the amount of enzyme produced whereas in the laboratory efficient enzymes are desired. Nature can also adjust the level of enzyme activity by changing the stability of the enzyme. One could find other examples of enzymes that are ideally suited to the cellular environment, but that are not ideally suited to the needs of researchers. Even if an enzyme is available to catalyse a reaction, a great deal of work may be required to optimise it activity for practical applications.

A time line for the development of techniques useful for protein engineering is given in Figure 1.22.

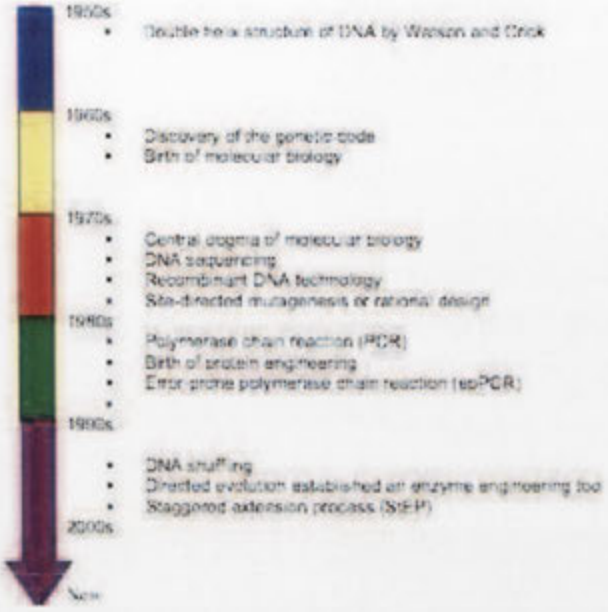


Figure 1.23: Time line for the development of techniques in DNA research and molecular biology that are useful for protein engineering.

It can be seen that by the mid-1980s, a great deal had been learnt about proteins – the structures of numerous proteins and their functions were known. However, the ability to modify proteins was limited and the idea of generating new proteins was not a practical objective. Chemical modification could be used to alter some types of residues and proteases could be used to cleave proteins, but in both cases the processes were not specific and usually resulted in a considerable loss of protein. Around this time, rapid advances in molecular biology, especially in DNA sequencing and recombinant DNA technology started to offer some hope that proteins could be modified in a specific way. Such a development was logical, considering the irreversible process of how DNA transcribes into RNA that later translates into proteins, as stated in the central dogma of molecular biology [105].

Initially the developments in molecular biology led to studies that were aimed at understanding how proteins functioned – residues were mutated site specifically and changes in activity were monitored. Eventually the idea

emerged of modifying proteins in a useful way and the term "protein engineering" was coined [106, 107]. Some of the first studies were directed at improving the stability of enzymes used in laundry detergent, but as the tools have improved the objectives of the practitioners have also become more ambitious.

The two main strategies

Two general strategies for protein engineering have emerged: *rational design* or redesign and *directed evolution* (directed molecular evolution).

1.6.2 Rational design

The rational approach to protein engineering requires a thorough knowledge of protein structure and a good understanding of how sequence, structure and function are related. One examines the structure with a view of rationally changing some property of the protein and then uses site-directed mutagenesis to make changes.

Site-directed mutagenesis is a technique that introduces a mutation at a defined site of a DNA molecule. This technique was developed in the late 1970s [108]. Site-directed mutagenesis is also commonly known as site-specific mutagenesis and oligonucleotide-directed mutagenesis. PCR-based site-directed mutagenesis protocols are the most commonly used method. The sequence of the DNA needs to be known so that oligonucleotide pairs bearing the desired mutation can be designed [109, 110]. Detailed procedures in site-directed mutagenesis are presented in Chapter 2. Stratagene developed several QuikChange™ Site-Directed Mutagenesis systems that allow not only single but multiple mutations to be created simultaneously. Some of these kits even facilitate frame-shift mutations such as insertion and deletion [111].

In this study, site-directed mutagenesis was used for two purposes. The first was to generate active site mutants where residues at the active site of GpdQ were substituted by other amino acids (Chapter 3). The effects of the mutations were then tested in order to better understand the roles of these residues and the catalytic mechanism. Secondly, site-directed mutagenesis was used to rationally design GpdQ mutants to develop a stable dimer (Chapter 5).

Although rational approaches to protein engineering have yielded some impressive results, they pale into insignificance when compared with the results achieved by natural selection over the course of evolutionary time. The tools used by nature to evolve protein, have till recent times been much more powerful than those available to protein engineers. Directed evolution seeks to use tools that mimic natural evolution for the purpose of evolving proteins for human purposes.

1.6.3 Directed evolution

Directed evolution or *in vitro* evolution is a strategy that mimics natural evolution and has the aim of developing enzymes on the laboratory timescale of months, with desirable properties that are not normally found or not optimal in nature. The main advantage directed evolution has over rational design is that the approach requires no prior understanding of the enzyme structure and mechanism. Random mutagenesis in directed evolution leads to the discovery of other valuable and beneficial mutations which otherwise will not be identified via rational design, since with rational design, the focus is often primarily on the active site [112-119].

Over the recent years, the application of directed evolution has successfully engineered enzymes with properties such as altered substrate specificities, which includes expanded substrate repertoires and improved weak promiscuous activities [28, 120-124], antibiotic resistance [125], improved

protein solubility [126], improved protein expression [127] organic solvent resistance [128], thermal stability [129] and enantioselectivity [130].

Directed evolution is usually comprised of sequential rounds of two processes: *generating diversity* (through mutagenesis and/or recombination) and *screening/selection*. Figure 1.23 illustrates the general concept of directed evolution studies. A complete cycle of generating diversity and screening/selection is usually called one round of evolution, with the best variants selected commonly referred to as first, second or third generation mutants.

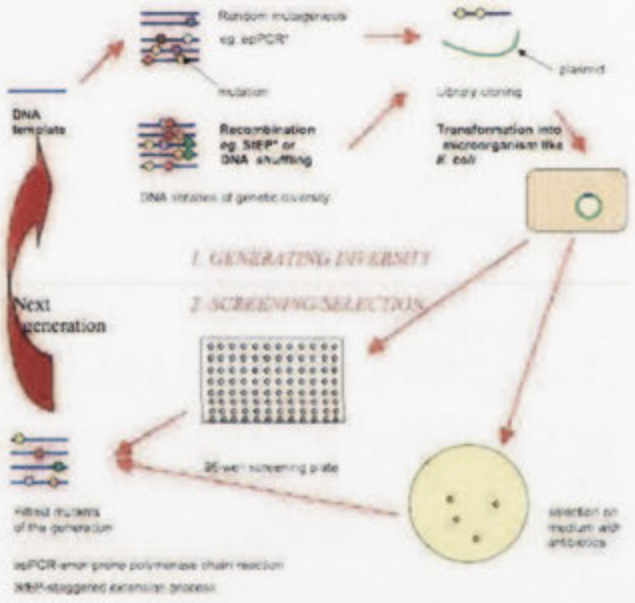


Figure 1.24: A schematic representation of directed evolution studies.
Directed evolution consists two independent areas of technology: generating diversity (creation of a library of genetic variants) and screening/selection of variants for desired properties.

Generating diversity

The invention of the polymerase chain reaction (PCR) in 1984 was followed by the development of error-prone PCR (epPCR) five years later. epPCR is a straightforward method for library construction, where the amplified full-length genes of an enzyme are subject to random mutagenesis [131-137].

Specific methods of introducing mutations to the amplified genes will be discussed in Chapter 4. The generated library of mutants is then screened or selected for the desired function. Each mutant must be assayed.

For subsequent rounds of evolution, the best variants selected or screened from the previous generations are subject to *in vitro* sexual recombination. The two main techniques in recombining genes are Stemmer's DNA shuffling [138-140] and Zhao's staggered extension process (StEP) [137, 141-143]. In this study, StEP was used and the concept of this technique is discussed in Chapter 2.

Screening/selection

Finding a suitable method to screen or select a library of mutants is often a challenging process. In selection, the target enzyme property is essential for survival and the mutant carrying the desired function can be easily distinguished from the background of reactions of all other mutants, *eg.* antibiotic resistance. Selection methods are generally less laborious and allow larger libraries to be evaluated. However, the target feature sometimes cannot be easily linked to cell survival. A screen is more versatile than selection, involving a direct measurement of enzymatic activity. The throughput is usually lower for a screen than it is for a selection [144].

1.6.4 A semi-rational approach to enzyme engineering – combining rational design and directed evolution

As pointed out earlier, rational design relies on structural knowledge of a protein and molecular modelling to predict desired amino acid changes. These changes are introduced into the gene of the protein by methods such as site-directed mutagenesis. Another disadvantage of rational design is that it can practically test only a small number of modifications [134].

Random mutagenesis, on the other hand, is the most vital step in generating diversity in directed evolution. However, random mutagenesis is biased and has limitations due to factors like transition bias and codon degeneracy.

Directed evolution is the only possible enzyme engineering method when the structure and the mechanism of the enzyme are not available. When the crystal structure of the enzyme has been solved, like GpdQ, rational design will present itself as a quick and intelligent approach to improve the enzyme. Rational design and directed evolution are not mutually exclusive; they can potentially complement each other. Whilst rational design can offer speed and precision, directed evolution may enable novel non-intuitive solutions to be found. Researchers will often apply both strategies to achieve optimal results [145]. Both approaches have been used in this study to modify GpdQ's activity towards a non-physiological phosphodiester substrate, bis (*p*-nitrophenyl) phosphate.

1.7 OVERVIEW OF THESIS

Based on the findings obtained from previous studies on GpdQ, especially the exceptionally broad substrate specificity of the enzyme, its structure, active site and the non-identical binding affinities at the binuclear metal center, this research project set out to achieve the following objectives:

i. To characterise the active site residues in order to achieve a better understanding of their roles and the catalytic mechanism of GpdQ. As part of this aim, the binding affinities of the two metal sites were also studied.

ii. To improve GpdQ's weak and promiscuous activities with poor, non-physiological substrates using directed evolution and rational design. Initially, priority was given to phosphodiester substrates like dimethyl phosphate and diethyl phosphate as well as phosphotriesters like methyl paraoxon and ethyl paraoxon. Enhanced catalytic activities with these substrates would have more

relevant bioremediation and industrial applications. We aimed to further develop Dr Sean Yu-McLoughlin's agar plate based growth assay (section 1.2.1, Chapter 1) as a method to preliminarily screen GpdQ mutant libraries with these four phosphate ester compounds. On the other hand, we were able to easily screen GpdQ's activity with bis (*p*-nitrophenyl) phosphate (b*p*NPP) due to reasons that it is a good and chromogenic substrate. GpdQ was then eventually evolved towards enhanced activity with this phosphodiester substrate. Several mutants were also designed to complement directed evolution in developing catalytically efficient GpdQ enzymes. At the end of this project, eight full rounds of directed evolution of GpdQ had been completed.

CHAPTER ONE REFERENCES

1. Voet, D. and Voet, J. G. (1990) Biochemistry, 1st edition. *John Wiley & Sons, Inc.*
2. Taylor, W. R. (1999) Protein structural domain identification. *Protein Engineering 12*, 203-216.
3. Carr, P. D., and Ollis, D. L. (2009) Alpha/beta hydrolase fold: An Update. *Protein and Peptide Letters 16*, 1137-1148.
4. Taylor, H. S. (1925) A theory of the catalytic surface. *Proceedings of the Royal Society of London A 108,* 105-111.
5. Jung, W. S., Hong, C. K., Lee, S., Kin, C. S., Kim, S. J., Kim, S. I., and Rhee, S. (2007) Structural and functional insights into intramolecular fructosyl transfer by inulin fructotransferase, *Journal of Biological Chemistry 282*, 8414-8423.
6. Muller, J. J., Barbirz, S., Heinle, K., Freiberg, A., Seckler, R., and Heinemann, U. (2008) An intersubunit active site between supercoiled parallel beta helices in the trimeric tailspike endorhamnosidase of Shigella flexneri phage Sf6. *Structure 16*, 766-775.
7. Lichtenthaler, F. W. (2002) Emil Fischer, his personality, his achievements, and his scientific progeny. *European Journal of Organic Chemistry*, 4095-4122.
8. Koshland, D. E. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America 44*, 98-104.
9. Castagnetto, J. M., Hennessy, S. W., Roberts, V. A., Getzoff, E. D., Tainer, J. A., and Pique, M. E. (2002) MDB: The metalloprotein database and browser at the Scripps Research Institute. *Nucleic Acids Research 30*, 379-382.
10. Yamashita, M. M., Wesson, L., Eisenman, G., and Eisenberg, D. (1990) Where metal ions bind in proteins. *Proceedings of the National Academy of Sciences of the United States of America 87*, 5648-5652.
11. Bertino, I., Drago, R. S., Luchinat, C. (1983) The coordination chemistry of metalloenzymes: The role of metals in reactions involving water, dihydrogen, and related species. *Proceedings of the NATO Advanced Study Institute.*
12. Lehninger, A. L. (1950) Role of metal ions in enzyme systems. *Physiological Reviews 30*, 393-429.
13. Wilson, C. J., Apiyo, D., and Wittung-Stafshede, P. (2004) Role of cofactors in metalloprotein folding. *Quarterly Reviews of Biophysics 37*, 285-314.
14. Bren, K. L., Pecoraro, V. L., and Gray, H. B. (2004) Metalloprotein folding. *Inorganic Chemistry 43*, 7894-7896.
15. Barondeau, D. P., and Getzoff, E. D. (2004) Structural insights into protein-metal ion partnerships. *Current Opinion in Structural Biology 14*, 765-774.

16. Frey, M. W., Frey, S. T., Horrocks, W. D., Kaboord, B. F., and Benkovic, S. J. (1996) Elucidation of the metal-binding properties of the Klenow fragment of Escherichia coli polymerase I and bacteriophage T4 DNA polymerase by lanthanide (III) luminescence spectroscopy. *Chemistry & Biology 3*, 393-403.
17. Steitz, T. A. (1999) DNA polymerases: Structural diversity and common mechanisms. *Journal of Biological Chemistry 274*, 17395-17398.
18. Costa, L. G. (2006) Current issues in organophosphate toxicology. *Clinica Chimica Acta 366*, 1-13.
19. Cleland, W. W., and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chemical Reviews 106*, 3252-3278.
20. Steenland, K., Jenkins, B., Ames, R. G., Omalley, M., Chrislip, D., and Russo, J. (1994) Chronic neurological sequelae to organophosphate pesticide poisoning. *American Journal of Public Health 84*, 731-736.
21. Sethunat.N, and Yoshida, T. (1973) A Flavobacterium sp that degrades diazinon and parathion. *Canadian Journal of Microbiology 19*, 873-875.
22. Tuovinen, K., Kalistekorhonen, E., Raushel, F. M., and Hanninen, O. (1994) Phosphotriesterase - a promising candidate for use in detoxification of organophosphates. *Fundamental and Applied Toxicology 23*, 578-584.
23. Defrank, J. J., and Cheng, T. C. (1991) Purification and properties of an organophosphorus acid anhydrase from a halophilic bacterial isolate. *Journal of Bacteriology 173*, 1938-1943.
24. Horne, I., Sutherland, T. D., Harcourt, R. L., Russell, R. J., and Oakeshott, J. G. (2002) Identification of an *opd* (organophosphate degradation) gene in an Agrobacterium isolate. *Applied and Environmental Microbiology 68*, 3371-3376.
25. Foo, J. L. (2009) Mechanistic study of enzymes involved in phosphate ester hydrolysis. *The Australian National University.*
26. Ely, F., Foo, J.-L., Jackson, C. J., Gahan, L. R., and Ollis, D. (2007) Enzymatic bioremediation: organophosphate degradation by binuclear metallo-hydrolases. *Current Topics in Biochemical Research 9,* 63-78.
27. Yang, H., Carr, P. D., McLoughlin, S. Y., Liu, J. W., Horne, I., Qiu, X., Jeffries, C. M. J., Russell, R. J., Oakeshott, J. G., and Ollis, D. L. (2003) Evolution of an organophosphate-degrading enzyme: A comparison of natural and directed evolution. *Protein Engineering 16*, 241-241.
28. Cho, C. M. H., Mulchandani, A., and Chen, W. (2004) Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Applied and Environmental Microbiology 70*, 4681-4685.

29. Watkins, L. M., Mahoney, H. J., McCulloch, J. K., and Raushel, F. M. (1997) Augmented hydrolysis of diisopropyl fluorophosphate in engineered mutants of phosphotriesterase. *Journal of Biological Chemistry 272*, 25596-25601.
30. Griffiths, A. D., and Tawfik, D. S. (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO Journal 22*, 24-35.
31. Raushel, F. M., and Holden, H. M. (2000) Phosphotriesterase: An enzyme in search of its natural substrate. *Advances in Enzymology 74*, 51-93.
32. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of *Escherichia coli* coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Applied and Environmental Microbiology 70*, 404-412.
33. Wolfenden, R., and Spence, G. (1967) Derepression of phosphomonoesterase and phosphodiesterase activities in *Aerobacter aerogenes*. *Biochimica et Biophysica Acta 146*, 296-8.
34. Cook, A. M., Daughton, C. G., and Alexander, M. (1978) Phosphorus-containing pesticide breakdown products - quantitative utilization as phosphorus sources by bacteria. *Applied and Environmental Microbiology 36*, 668-672.
35. Gerlt, J. A., and Westheim, F. H. (1973) Phosphodiesterase from *Enterobacter aerogenes*. *Journal of the American Chemical Society 95*, 8166-8168.
36. Gerlt, J. A., and Whitman, G. J. R. (1975) Purification and properties of a phosphohydrolase from *Enterobacter aerogenes*. *Journal of Biological Chemistry 250*, 5053-5058.
37. Inouye, H., Michaelis, S., Wright, A., and Beckwith, J. (1981) Cloning and restriction mapping of the alkaline phosphatase structural gene (*pho*A) of *Escherichia coli* and generation of deletion mutants *in vitro*. *Journal of Bacteriology 146*, 668-675.
38. Brzoska, P., Rimmele, M., Brzostek, K., and Boos, W. (1994) The pho regulon-dependent *ugp* uptake system for glycerol-3-phosphate in *Escherichia coli* is trans inhibited by Pi. *Journal of Bacteriology 176*, 15-20.
39. Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews 72*, 317-364.
40. Makino, K., Shinagawa, H., Amemura, M., Kimura, S., Nakata, A., and Ishihama, A. (1988) Regulation of the phosphate regulon of *Escherichia coli* - activation of *pst*S transcription by PhoB protein in vitro. *Journal of Molecular Biology 203*, 85-95.
41. Makino, K., Shinagawa, H., Amemura, M., and Nakata, A. (1986) Nucleotide sequence of the PhoB gene, the positive regulatory gene for the phosphate regulon of *Escherichia coli* K-12. *Journal of Molecular Biology 190*, 37-44.

42. Larson, T. J., Ehrmann, M., and Boos, W. (1983) Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the glp regulon. *Journal of Biological Chemistry 258*, 5428-5432.
43. Larson, T. J., and Vanloobhattacharya, A. T. (1988) Purification and characterization of *glp*Q-encoded glycerophosphodiester phosphodiesterase from *Escherichia coli* K-12. *Archives of Biochemistry and Biophysics 260*, 577-584.
44. Tommassen, J., Eiglmeier, K., Cole, S. T., Overduin, P., Larson, T. J., and Boos, W. (1991) Characterization of two genes, *glp*Q and *ugp*Q, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. *Molecular & General Genetics 226*, 321-327.
45. Santelli, E., Schwarzenbacher, R., McMullan, D., Biorac, T., Brinen, L. S., Canaves, J. M., Campbell, J., Dai, X. P., Deacon, A. M., Elsliger, M. A., Eshagi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S. K., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kovarik, J. S., Kreusch, A., Kuhn, P., Lesley, S. A., McPhillips, T. M., Miller, M. D., Morse, A., Moy, K., Ouyang, J., Page, R., Quijano, K., Rezezadeh, F., Robb, A., Sims, E., Spraggon, G., Stevens, R. C., van den Bedem, H., Velasquez, J., Vincent, J., von Delft, F., Wang, X. H., West, B., Wolf, G., Xu, Q. P., Hodgson, K. O., Wooley, J., and Wilson, I. A. (2004) Crystal structure of a glycerophosphodiester phosphodiesterase (GDPD) from *Thermotoga maritima* (TM1621) at 1.60 angstrom resolution. *Proteins-Structure Function and Bioinformatics 56*, 167-170.
46. Rao, K. N., Bonanno, J. B., Burley, S. K., and Subramanyam Swaminathan (2006). Crystal structure of glycerophosphodiester phosphodiesterase from *Agrobacterium tumefaciens* by SAD with a large asymmetric unit. *Proteins: Structure, Function, and Bioinfomatics 65*, 514-518.
47. Shi, L., Liu, J. F., An, X. M., and Liang, D. C. (2008) Crystal structure of glycerophosphodiester phosphodiesterase (GDPD) from *Thermoanaerobacter tengcongensis*, a metal ion-dependent enzyme: Insight into the catalytic mechanism. *Proteins-Structure Function and Bioinformatics 72*, 280-288.
48. Shang, E. S., Skare, J. T., ErdjumentBromage, H., Blanco, D. R., Tempst, P., Miller, J. N., and Lovett, M. A. (1997) Sequence analysis and characterization of a 40-kilodalton *Borrelia hermsii* glycerophosphodiester phosphodiesterase homolog. *Journal of Bacteriology 179*, 2238-2246.
49. Zheng, B., Berrie, C. P., Corda, D., and Farquhar, M. G. (2003) GDE1/MIR16 is a glycerophosphoinositol phosphodiesterase regulated by stimulation of G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America 100*, 1745-1750.
50. Koonin, E. V. (1994) Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Science 3*, 356-358.

51. Ghanem, E., Li, Y. C., Xu, C. F., and Raushel, F. M. (2007) Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry 46*, 9032-9040.
52. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of the American Chemical Society 130*, 14129-14138.
53. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.
54. Ali, M. H., and Imperiali, B. (2005) Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry 13*, 5013-5020.
55. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure 29*, 105-153.
56. Traut, T. W. (1994) Dissociation of enzyme oligomers - a mechanism for allosteric regulation. *Critical Reviews in Biochemistry and Molecular Biology 29*, 125-163.
57. Najera, H., Costas, M., and Fernandez-Velasco, D. A. (2003) Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme. *Biochemical Journal 370*, 785-792.
58. Lee, B. and Richards, F. M. (1971) The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology 55*, 379-400.
59. Ropp, P. A., and Traut, T. W. (1991) Purine nucleoside phosphorylase - allosteric regulation of a dissociating enzyme. *Journal of Biological Chemistry 266*, 7682-7687.
60. Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004) The power of two: protein dimerization in biology. *Trends Biochem Sci 29*, 618-25.
61. Vamvaca, K., Butz, M., Walter, K. U., Taylor, S. V., and Hilvert, D. (2005) Simultaneous optimization of enzyme activity and quaternary structure by directed evolution. *Protein Science 14*, 2103-2114.
62. MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution. *Science 279*, 1958-1961.
63. Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2004) SCOP database in 2004: refinements intergrate structure and sequence family data. *Nucleic Acids Research 32*, 226-229.

64. Zhang, R., Kim, Y., Dementieva, I., Duck, N., Stols, L., Donnelly, M., and Joachimiak, A. (2004). Structural genomics, crystal structure of glycerophosphoryl diester phosphodiesterase from *E. coli. RCSB Protein Data Bank 1TQ8.*
65. Malashkevic, V. M., Fedorov, E., and Almo, S. C. (2005). Crystal structure of periplasmic glycerophosphodiesterase from *Escherichia coli. RCSB Protein Data Bank 1YDY.*
66. Strater, N., Klabunde, T., Tucker, P., Witzel, H., and Krebs, B. (1995) Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site. *Science 268*, 1489-1492.
67. Shenoy, A. R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S. S., and Podobnik, M. (2007) Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis. Journal of Molecular Biology 365*, 211-225.
68. Chen, S. F., Yakunin, A. F., Kuznetsova, E., Busso, D., Pufan, R., Proudfoot, M., Kim, R., and Kim, S. H. (2004) Structural and functional characterization of a novel phosphodiesterase from *Methanococcus jannaschii. Journal of Biological Chemistry 279*, 31854-31862.
69. Voegtli, W. C., White, D. J., Reiter, N. J., Rusnak, F., and Rosenzweig, A. C. (2000) Structure of the bacteriophage lambda Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes. *Biochemistry 39*, 15365-15374.
70. Knofel, T., and Strater, N. (1999) X-ray structure of the *Escherichia coli* periplasmic 5 '-nucleotidase containing a dimetal catalytic site. *Nature Structural Biology 6*, 448-453.
71. Gerlt, J. A., and Wan, W. H. Y. (1979) Stereochemistry of the hydrolysis of the endo isomer of uridine 2',3'-cyclic phosphorothioate catalyzed by the nonspecific phosphohydrolase from *Enterobacter aerogenes. Biochemistry 18*, 4630-4638.
72. Jackson, C. J., Hadler, K. S., Carr, P. D., Oakley, A. J., Yip, S., Schenk, G., and Ollis, D. L. (2008) Malonate-bound structure of the glycerophosphodiesterase from *Enterobacter aerogenes* (GpdQ) and characterization of the native Fe2+ metal-ion preference, *Acta Crystallographica Section F-Structural Biology and Crystallization Communications 64*, 681-685.
73. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural Flexibility Enhances the Reactivity of the Bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.
74. Prescott, J. M., Wagner, F. W., Holmquist, B., and Vallee, B. L. (1985) Spectral and kinetic studies of metal substituted *Aeromonas aminopeptidase* - nonidentical, interacting metal-binding sites. *Biochemistry 24*, 5350-5356.

75. Schenk, G., Ge, Y. B., Carrington, L. E., Wynne, C. J., Searle, I. R., Carroll, B. J., Hamilton, S., and de Jersey, J. (1999) Binuclear metal centers in plant purple acid phosphatases: Fe-Mn in sweet potato and Fe-Zn in soybean, *Archives of Biochemistry and Biophysics 370*, 183-189.
76. Kimura, E. (2000) Dimetallic hydrolases and their models. *Current Opinion in Chemical Biology 4*, 207-213.
77. McMillen, L., Beacham, I. R., and Burns, D. M. (2003) Cobalt activation of *Escherichia coli* 5 '-nucleotidase is due to zinc ion displacement at only one of two metal-ion-binding sites. *Biochemical Journal 372*, 625-630.
78. Mitic, N., Smith, S. J., Neves, A., Guddat, L. W., Gahan, L. R., and Schenk, G. (2006) The catalytic mechanisms of binuclear metallohydrolases. *Chemical Reviews 106*, 3338-3363.
79. Miller, D. J., Shuvalova, L., Evdokimova, E., Savchenko, A., Yakunin, A. F., and Anderson, W. F. (2007) Structural and biochemical characterization of a novel Mn^{2+}-dependent phosphodiesterase encoded by the yfcE gene. *Protein Science 16*, 1338-1348.
80. Rusnak, F., Yu, L. A., Todorovic, S., and Mertz, P. (1999) Interaction of bacteriophage lambda protein phosphatase with Mn(II): Evidence for the formation of a [Mn(II)](2) cluster. *Biochemistry 38*, 6943-6952.
81. White, D. J., Reiter, N. J., Sikkink, R. A., Yu, L., and Rusnak, F. (2001) Identification of the high affinity Mn^{2+} binding site of bacteriophage lambda phosphoprotein phosphatase: Effects of metal ligand mutations on electron paramagnetic resonance spectra and phosphatase activities. *Biochemistry 40*, 8918-8929.
82. Aquino, M. A. S., Lim, J. S., and Sykes, A. G. (1994) Mechanism of the reaction of different phosphates with the Iron(II)Iron(III) form of purple acid phosphatase from porcine uteri (uteroferrin). *Journal of the Chemical Society-Dalton Transactions*, 429-436.
83. Twitchett, M. B., Schenk, G., Aquino, M. A. S., Yiu, D. T. Y., Lau, T. C., and Sykes, A. G. (2002) Reactivity of M-II metal-substituted derivatives of pig purple acid phosphatase (Uteroferrin) with phosphate. *Inorganic Chemistry 41*, 5787-5794.
84. Merkx, M., and Averill, B. A. (1998) Ga3+ as a functional substitute for Fe3+: Preparation and characterization of the Ga^{3+}-Fe^{2+} and Ga^{3+}-Zn^{2+} forms of bovine spleen purple acid phosphatase. *Biochemistry 37*, 8490-8497.
85. Merkx, M., and Averill, B. A. (1999) Probing the role of the trivalent metal in phosphate ester hydrolysis: Preparation and characterization of purple acid phosphatases containing (AlZnII)-Zn-III and (InZnII)-Zn-III active sites, including the first example of an active aluminum enzyme. *Journal of the American Chemical Society 121*, 6683-6689.
86. Wouters, M. A. and Curmi, P. M. G. (1995) An analysis of side-chain interactions and pair correlations within antiparellel beta-

sheets – the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins – Structure, Function and Genetics 22*, 119-131.

87. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) 3D Domain swapping - a mechanism for oligomer assembly. *Protein Science 4*, 2455-2468.
88. Xu, D., Tsai, C. J., and Nussinov, R. (1998) Mechanism and evolution of protein dimerization. *Protein Science 7*, 533-544.
89. Kundu, S., and Jernigan, R. L. (2004) Molecular mechanism of domain swapping in proteins: An analysis of slower motions. *Biophysical Journal 86*, 3846-3854.
90. Liu, Y. S., Gotte, G., Libonati, M., and Eisenberg, D. (2001) A domain-swapped RNase A dimer with implications for amyloid formation. *Nature Structural Biology 8*, 211-214.
91. Pasta, S. Y., Raman, B., Ramakrishna, T., and Rao, C. M. (2002) Role of the C-terminal extensions of alpha-crystallins - Swapping the C-terminal extension of alpha A-crystallin to alpha B-crystallin results in enhanced chaperone activity. *Journal of Biological Chemistry 277*, 45821-45828.
92. Rousseau, F., Schymkowitz, J. W. H., and Itzhaki, L. S. (2003) The unfolding story of three-dimensional domain swapping. *Structure 11*, 243-251.
93. Ivanov, D., Tsodikov, O. V., Kasanov, J., Ellenberger, T., Wagner, G., and Collins, T. (2007) Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America 104*, 4353-4358.
94. Ishikawa, H., Kim, S., Kwak, K., Wakasugi, K., and Fayer, M. D. (2007) Disulfide bond influence on protein structural dynamics probed with 2D-IR vibrational echo spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America 104*, 19309-19314.
95. van der Wijk, T., Overvoorde, J., and den Hertog, J. (2004) H_2O_2-induced intermolecular disulfide bond formation between receptor protein-tyrosine phosphatases. *Journal of Biological Chemistry 279*, 44355-44361.
96. Herscovitch, M., Comb, W., Ennis, T., Coleman, K., Yong, S., Armstead, B., Kalaitzidis, D., Chandani, S., and Gilmore, T. D. (2008) Intermolecular disulfide bond formation in the NEMO dimer requires Cys54 and Cys347. *Biochemical and Biophysical Research Communications 367*, 103-108.
97. Li, T. S., Yamane, H., Arakawa, T., Narhi, L. O., and Philo, J. (2002) Effect of the intermolecular disulfide bond on the conformation and stability of glial cell line-derived neurotrophic factor. *Protein Engineering 15*, 59-64.
98. Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P., and Tainer, J. A. (2001) Structural biochemistry and interaction

architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell 105*, 473-485.

99. Richter, W. (2002) 3 ',5 '-cyclic nucleotide phosphodiesterases class III: Members, structure, and catalytic mechanism. *Proteins-Structure Function and Genetics 46*, 278-286.
100. Imamura, R., Yamanaka, K., Ogura, T., Hiraga, S., Fujita, N., Ishihama, A., and Niki, H. (1996) Identification of the cpdA gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in *Escherichia coli*. *Journal of Biological Chemistry 271*, 25423-25429.
101. Dunlap, P. V., and Callahan, S. M. (1993) Characterization of a periplasmic 3',5'-cyclic nucleotide phosphodiesterase gene, *cpd*P, from the marine symbiotic bacterium *Vibrio fischeri*. *Journal of Bacteriology 175*, 7124-7124.
102. Shenoy, A. R., Sreenath, N., Podobnik, M., Kovacevic, M., and Visweswariah, S. S. (2005) The Rv0805 gene from *Mycobacterium tuberculosis* encodes a 3 ',5 '-cyclic nucleotide phosphodiesterase: Biochemical and mutational analysis. *Biochemistry 44*, 15695-15704.
103. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics 24*, 2780-2781.
104. Tyagi, R., Shenoy, A. R., and Visweswariah, S. S. (2009) Characterization of an Evolutionarily Conserved Metallophosphoesterase That Is Expressed in the Fetal Brain and Associated with the WAGR Syndrome. *Journal of Biological Chemistry 284*, 5217-5228.
105. Crick, F. (1970) Central dogma of molecular biology. *Nature 227*, 561-563.
106. Wetzel, R. (1986) What is protein engineering. *Protein Engineering 1*, 3-5.
107. Leatherbarrow, R. J., and Fersht, A. R. (1986) Protein engineering. *Protein Engineering 1*, 7-16.
108. Hutchinson, C. A., Phillips, S., Marshall, H. E., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry 253,* 6551-6560.
109. Kunkel, T. A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America 82*, 488-492.
110. Ling, M. M., and Robinson, B. H. (1997) Approaches to DNA mutagenesis: An overview. *Analytical Biochemistry 254*, 157-178.
111. Wang, W. Y., and Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange (TM) site-directed mutagenesis. *Biotechniques 26*, 680-682.

112. Arnold, F. H. (1996) Directed evolution: Creating biocatalysts for the future. *Chemical Engineering Science 51*, 5091-5102.
113. Arnold, F. H. (1998) Design by directed evolution. *Accounts of Chemical Research 31*, 125-131.
114. Voigt, C. A., Kauffman, S., and Wang, Z. G. (2001) Rational evolutionary design: The theory of in vitro protein evolution. *Advances in Protein Chemistry, Vol 55 55*, 79-160.
115. Dalby, P. A. (2003) Optimising enzyme function by directed evolution. *Current Opinion in Structural Biology 13*, 500-505.
116. Turner, N. J. (2003) Directed evolution of enzymes for applied biocatalysis. *Trends in Biotechnology 21*, 474-478.
117. Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) Evolving strategies for enzyme engineering. *Current Opinion in Structural Biology 15*, 447-452.
118. Kaur, J., and Sharma, R. (2006) Directed evolution: An approach to engineer enzymes. *Critical Reviews in Biotechnology 26*, 165-199.
119. Sen, S., Dasu, V. V., and Mandal, B. (2007) Developments in directed evolution for improving enzyme functions. *Applied Biochemistry and Biotechnology 143*, 212-223.
120. Yano, T., Oue, S., and Kagamiyama, H. (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proceedings of the National Academy of Sciences of the United States of America 95*, 5511-5515.
121. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology 6*, R91-R105.
122. Fa, M., Radeghieri, A., Henry, A. A., and Romesberg, F. E. (2004) Expanding the substrate repertoire of a DNA polymerase by directed evolution. *Journal of the American Chemical Society 126*, 1748-1754.
123. Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions. *Nature Genetics 37*, 73-76.
124. Lin, H. N., Tao, H. Y., and Cornish, V. W. (2004) Directed evolution of a glycosynthase via chemical complementation. *Journal of the American Chemical Society 126*, 15051-15059.
125. Watson, M., Liu, J. W., and Ollis, D. (2007) Directed evolution of trimethoprim resistance in *Escherichia coli*. *Febs Journal 274*, 2661-2671.
126. Liu, J. W., Boucher, Y., Stokes, H. W., and Ollis, D. L. (2006) Improving protein solubility: The use of the *Escherichia coli* dihydrofolate reductase gene as a fusion reporter. *Protein Expression and Purification 47*, 258-263.
127. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. (2005) Increased expression of a bacterial phosphotriesterase in

Escherichia coli through directed evolution. *Protein Expression and Purification 41*, 433-440.

128. You, L., and Arnold, F. H. (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide (vol 9, pg 78, 1996). *Protein Engineering 9*, 719-719.

129. Flores, H., and Ellington, A. D. (2002) Increasing the thermal stability of an oligomeric protein, beta-glucuronidase. *Journal of Molecular Biology 315*, 325-337.

130. Park, S., Morley, K. L., Horsman, G. P., Holmquist, M., Hult, K., and Kazlauskas, R. J. (2005) Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. *Chemistry & Biology 12*, 45-54.

131. Leung, D. W., Chen, E. and Goeddel, D. V. (1989) A method for random mutagenesis of a defined DNA fragment using a modified polymerase chain reaction. *Technique 1,* 11-15.

132. Cadwell, R. C., and Joyce, G. F. (1994) Mutagenic PCR. *PCR-Methods and Applications 3*, S136-S140.

133. Wilson, D. S., and Keefe, A. D. (2001) Random mutagenesis by PCR. *Current Protocol in Molecular Biology Chapter 8*, Unit 8.3.

134. Ness, J. E., Del Cardayre, S. B., Minshull, J., and Stemmer, W. P. C. (2001) Molecular breeding: The natural approach to protein design. *Advances in Protein Chemistry, Vol 55*, 261-292.

135. Cirino, P. C., Mayer, K. M., and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods in Molecular Biology 231,* 3-9.

136. Pritchard, L., Corne, D., Kell, D., Rowland, J., and Winson, M. (2005) A general model of error-prone PCR. *Journal of Theoretical Biology 234*, 497-509.

137. An, Y. F., Ji, J. F., Wu, W. F., Huang, R. B., Wei, Y. T., and Xiu, Z. L. (2008) Random mutagenesis and recombination of *sam*1 gene by integrating error-prone PCR with staggered extension process. *Biotechnology Letters 30*, 1227-1232.

138. Stemmer, W. P. C. (1994) DNA Shuffling by random fragmentation and reassembly - in-vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America 91*, 10747-10751.

139. Patten, P. A., Howard, R. J., and Stemmer, W. P. C. (1997) Applications of DNA shuffling to pharmaceuticals and vaccines. *Current Opinion on Biotechnology 8,* 724-733.

140. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature 391*, 288-291.

141. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology 16*, 258-261.

142. Zhao, H. M. (2004) Staggered extension process in vitro DNA recombination. *Protein Engineering 388*, 42-49.
143. Zhao, H. M., and Zha, W. J. (2006) In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature Protocols 1*, 1865-1871.
144. Arnold, F. H., and Georgiou, G. (2003) Directed enzyme evolution: Screening and selection methods. *Methods in Molecular Biology 230.*
145. Chica, R. A., Doucet, N., and Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology 16*, 378-384.

2

EXPERIMENTAL PROCEDURES

CHAPTER TWO: EXPERIMENTAL PROCEDURES

2.1 INTRODUCTION

This chapter discusses all the techniques relevant to all the experimental work performed in preparation of this dissertation, including molecular biology, protein expression and purification, enzyme kinetic analysis, protein crystallisation, equations and computer software programs.

Equipment, chemicals and reagents

Whenever required, especially for molecular biology work, all glassware instruments, microtubes, micropipette tips, growth media, buffers and solutions were sterilised by autoclaving. Distilled water used in this study (mQH_2O) was purified with a Milli-Q Reagent System. A complete list of suppliers and manufacturers of equipment, kits, chemicals and reagents is attached in Appendix A. Recipes of culture media, buffers and other reagents used in this study are given in Appendix B.

2.2 STRAINS AND VECTORS

2.2.1 *Escherichia coli* strains DH5α and DH10β

Library efficiency electro-competent *Escherichia coli* DH5α cells were used throughout this project for general cloning and protein expression procedures. Directed evolution studies in this project involved the generation of genetic libraries. For this reason DH5α cells were the *E. coli* strain of choice, as these cells are known for their high transformation efficiency, relatively simple preparation, usage and storage. Competent cells were prepared from cell stock that was originally purchased from Invitrogen, handled and stored as described [1] with minor modifications.

All glassware instruments, microtubes, centrifuge tubes, pipette tips, mQH_2O, 10% (v/v) glycerol and YENB medium were sterilised prior to the preparation of any electro-competent cells to ensure their purity. The culture flasks and measuring cylinders used in making electro-competent cells were also set aside from other glassware in the laboratory and washed separately to reduce chances of contamination with salt as high ionic conductivity greatly reduces electroporation efficiency [1, 2].

A tube of DH5α cell stock was taken out of -80°C and thawed on ice for about 10 minutes to minimise 'heat shock' towards the cells [3]. Using a flame sterilised inoculation loop, an inoculum was retrieved from the thawed cell culture and streaked onto an LB plate. The plate was then incubated at 37°C overnight (approximately 16 hours).

The following day, 5 mL of YENB medium was inoculated with a single colony of DH5α cells from the LB plate and incubated in a 37°C shaker (~200 rpm).

On the third day, the 5 mL overnight culture of DH5α cells was transferred into a 2 L baffled flask containing 1 L of YENB medium. The culture was grown in a 37°C shaker (~200 rpm) and the optical density of the culture at 595 nm (OD_{595}) was checked regularly with a cell density meter until it reached 0.6-0.8. It typically took about 4 hours to reach this density. The culture was placed on ice immediately to stop cell growth. All subsequent steps were carried out at 4°C.

After it had been cooled down to 4°C, the cells were harvested by centrifugation at 4,000 x g for 10 minutes with an SLA-3000 fixed angled rotor in two 500 mL centrifuge bottles. The cell pellet was resuspended with 200 mL mQH_2O by gently shaking or pipetting (no vortexing) to wash the remaining medium off the cells. Centrifugation at 4,000 x g for 10 minutes was performed again as before. This time, the cell pellet was resuspended

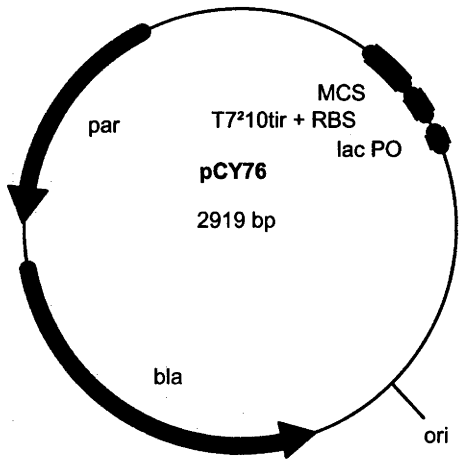
with 50 mL 10% glycerol and subjected to a third round of centrifugation. The cells were then suspended with 10% glycerol to a final volume of 1.5-2 mL and dispensed into 50 μL aliquots which were then immersed in liquid nitrogen for rapid freezing. The cell aliquots were stored at -80°C immediately for future laboratory use.

The DH10β cells used in setting up control experiments during *in vivo* agar plate-based growth assay screenings had been transformed by Dr Sean Yu-McLoughlin with a 5.6-kb DNA fragment containing the *Enterobacter aerogenes ugp* operon. Dr Yu-McLoughlin opted for DH10β cells [4] as these cells were designed for the uptake of large DNA inserts [5].

2.2.2 Vector pCY76

The *gpd*Q full-length gene (825 bp) was cloned into the *Nde*I (5' terminus) and *Eco*RI (3' terminus) sites of vector pCY76 by Dr Jian-Wei Liu. The last three nucleotides of the *Nde*I restriction site, CATATG, encode methionine, the start codon of the GpdQ protein.

The vector pCY76 (*par*$^{+}$, *bla*$^{+}$, T7ϕ10^{tir+}, *lac* PO^{+}) was constructed by Dr Cy Jeffries [6]. The physical map of pCY76 is shown in Figure 2.1. A full nucleotide sequence of the plasmid is supplied in Appendix C.



ori: Origin of replication
lac PO: Promoter and operator of the *lac* operon
T7ϕ10^{tir}: ϕ10 translational initiation region from T7 phage
RBS: Ribosome binding site
MCS: Multiple cloning site
par : Partition element that maintains plasmid segregational stability (especially high-copy-number plasmids) through many generations without antibiotic selection
bla: Expresses β-lactamase that is responsible for the resistance of the plasmid towards β-lactam antibiotics like ampicillin

Figure 2.1: The physical map of pCY76 vector.

pCY76 was engineered as a simple, high-copy-number cloning vector that also offers a constitutive protein expression system under the control of the *lac* promoter. The segregational stability of the vector was enhanced by the insertion of the partition function (*par*). The *lac* promoter was somewhat 'leaky', allowing protein expression without IPTG induction due to its ability to interact with both glucose and lactose. However, constitutive expression has certain advantages over other expression systems which use heat or IPTG induction as it avoids any cellular shock and stress. To reduce the deleterious effects of the *lac*-controlled, constitutive expression system of pCY76, a strong ribosome binding site (RBS) and a ϕ10 translational initiation region from T7 phage (as used for Studier's T7 pET vectors) were included upstream of the multiple cloning site during design of the plasmid [6-9].

GpdQ expressed satisfactorily with pCY76 (Figure 2.16).

2.3 MUTATIONAL ANALYSIS AND RATIONAL DESIGN OF GpdQ USING SITE-DIRECTED MUTAGENESIS

The site-directed mutagenesis (SDM) technique of introducing amino acid changes to GpdQ was used in this study for two purposes:

i. To explore the role of active site residues (Chapter 3).

ii. To disrupt the oligomeric structure of GpdQ (Chapter 5).

Figure 2.2 presents a summary of the methodologies used in this study in producing site-directed mutants.

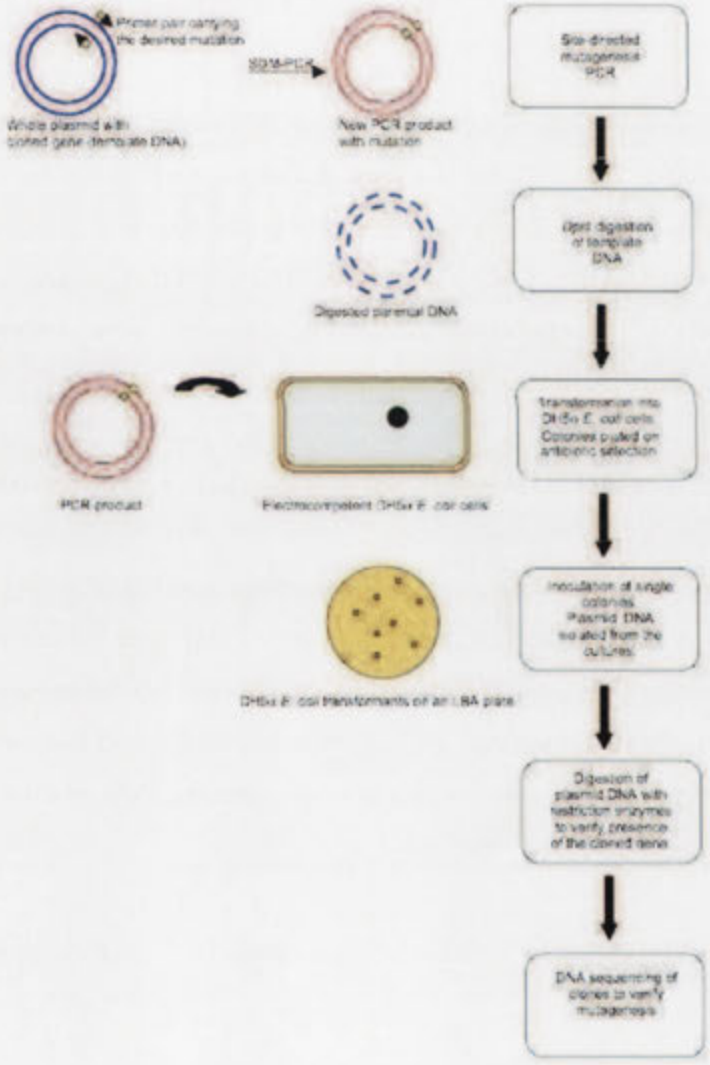


Figure 2.2: The process of making a site-directed mutant.
A flow-chart illustrating how a site-directed mutant was made in this study.

2.3.1 Design of primers

An online software program provided by Stratagene (http://www.stratagene.com/sdmdesigner/default.aspx) was used to design primers for site-directed mutagenesis. Primer pairs were typically 100% complementary to each other; carried the mutation of interest, and were of 25-45 bases in length with a melting temperature (T_m) ≥ 78°C. The calculation of T_m is based on a formula given by Stratagene:

$$T_m = 81.5 + (\%GC) - (675/N) - \%mismatch \qquad \text{(Equation 2.1)}$$

Where,

- N is the length of the oligonucleotide primer in bases
- Values for %GC and %mismatch should be whole numbers

A glossary of all the mutagenic primer pairs used in this research is given in Appendix D. Sequencing/PCR grade primers were ordered from Geneworks.

2.3.2 Site-directed mutagenesis polymerase chain reaction (SDM-PCR)

Site-directed mutagenesis polymerase chain reaction (SDM-PCR) was used to introduce point mutations in the wild-type *gpd*Q gene (or other mutant). Two models of thermocyclers were available in the laboratory: an iCycler® (Bio-Rad) and a GeneAmp® PCR System 9700 (Applied Biosystems). As both models were equipped with a hot-top assembly, PCR reactions did not have to be overlaid with mineral oil.

SDM-PCR reactions were set up in 200 µL clear PCR microtubes with:

Template pCY76-*gpd*Q	0.5 µl (20-30 ng)
(*Purified with Qiagen Miniprep kit, about 50 ng/µl*)	
Forward primer (25 µM)	1.0 µl
Reverse primer (25 µM)	1.0 µl
dNTP (40 mM)	1.0 µl
10X Native Plus buffer	5.0 µl
Pfu DNA polymerase	1.0 µl
mQH_2O	40.5 µl
Total volume	50.0 µl

*PCR reaction mixtures were given a 3-second quick spin in the microcentrifuge before they were put into the thermal cycler.

The SDM-PCR reaction mixtures were initially subject to the standard protocol recommended by Stratagene: Pre-denaturation at 95°C for 30 seconds, followed by 18 cycles of 30-second denaturation at 95°C, 1-minute annealing at 55°C and 10 minutes of extension at 68°C.

Optimisation of SDM-PCR

Towards the later part of this project, site-directed mutagenesis experiments were challenged by the increasingly prevalent phenomenon of primer dimerisation. Other users of QuikChange™ Site-Directed Mutagenesis kits have previously reported similar problems [10, 11], which highlighted one major shortcoming of this protocol that requires 100% complementary primer pairs to be 25-45 bases in length to reach a melting temperature (T_m) ≥ 78°C, so that primer self-pairing can be minimised.

The primer dimerisation phenomenon encountered in this study would typically create amplification products that have additional ~25 base pairs, which was a repetition of the full-length primer sequence, (Figures 2.3).

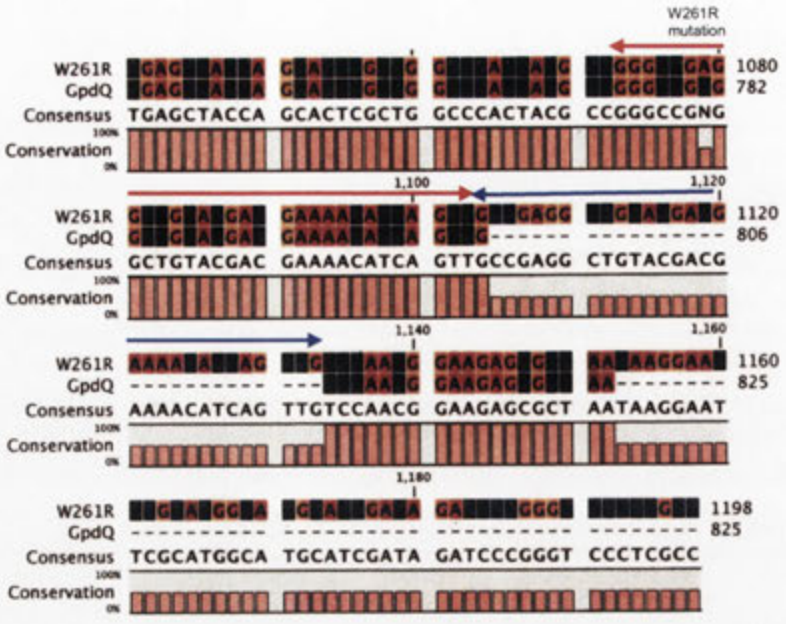


Figure 2.3: DNA sequence alignment of the W261R mutant and wild-type GpdQ.
The red arrow indicates the length of the primer, which contains the desired W261R mutation. The blue arrow shows the duplicated primer, which causes the insertion of 29 extra base pairs into the amplified PCR product.

From the DNA sequence alignment of W261R mutant and wild-type GpdQ in Figure 2.3, it appeared that the primer molecules self-paired among themselves before the dimers annealed to the template DNA to start amplification, as depicted in Figure 2.4. This was surprising, considering the fact that *Pfu* DNA polymerase possesses the 3' to 5' exonuclease proofreading activity, which works to correct nucleotide misincorporation.

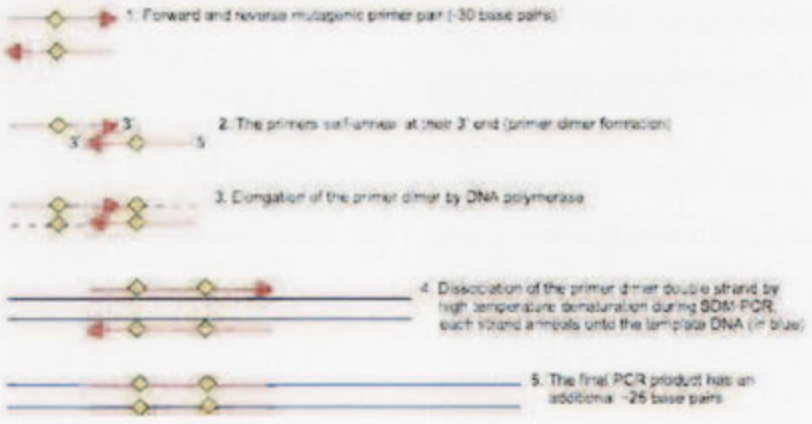


Figure 2.4: Primer dimerisation.
A diagram illustrating primer dimerisation during SDM-PCR in this study.

Alteration to cycling parameters

To overcome the primer dimerisation phenomenon, 1.5 µL of 100% DMSO (Sigma) was added into the PCR reactions and the final volume of the reactions was adjusted accordingly. DMSO is known to enhance PCR amplification by increasing the yield and specificity especially in GC-rich templates, as well as to inhibit the formation of secondary structures in the template DNA and primers [12]. Kureishi *et al.* recommended pre-boiling the primers especially those with a high GC content [13] but in this study, pre-boiled primers failed to yield any PCR product.

Cycling parameters were also altered to a program that yielded more amplification products, reduced primer self-pairing and was more time-efficient:

Pre-denaturation: 95°C	30 s		
Denaturation: 95°C	30 s	↕	25 cycles
Annealing: 55°C	30 s		
Extension: 68°C	8 min		
4°C	∞		

Partially overlapping primer pairs

Partially overlapping primer pairs eliminated the primer dimerisation problem for mutant W261R. These primer pairs complemented each other at the 5'-terminus, with the targeted mutation lying among the overlapping region. The partial overlapping also widened the melting temperature (T_m) gap between primer-to-template annealing and primer-to-primer self-annealing, so the $T_m \geq 78°C$ as proposed by Stratagene was no longer a strict requirement. This gave increased flexibility during primer design [10].

The W261R reverse primer was re-designed to give partial instead of full complementariness of the new primer pair (Figure 2.5). This strategy successfully abolished the primer dimerisation problem altogether.

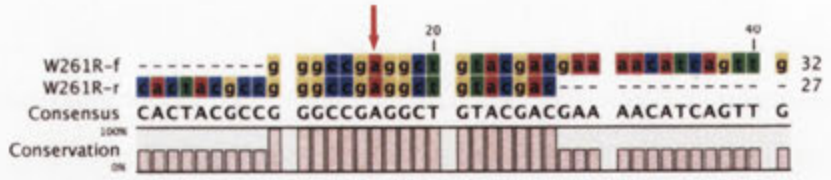


Figure 2.5: Partially complementary W261R forward (W261R-f) and reverse (W261R-r) primers.
The red arrow indicates the location of the point mutation.

2.3.3 Agarose gel electrophoresis and gel purification

The success of SDM-PCR was checked by running 2 µl of the reaction mixture on an 1.0% (w/v) agarose gel electrophoresis prepared in the SB buffer system. Ethidium bromide was added to the heat-dissolved gel solution prior to pouring to a final concentration of 0.5 µg/mL. The electrophoresis was always run for 30 minutes at 150 V, 300 mA. 6X loading dye was added to every DNA sample before it was injected into the wells of the agarose gel. The DNA was then visualised in the ultra-violet light transilluminator at a wavelength of 312 nm and photographed with the camera attached to the UVIpro transilluminator system. A photograph of a typical gel is shown in Figure 2.6.

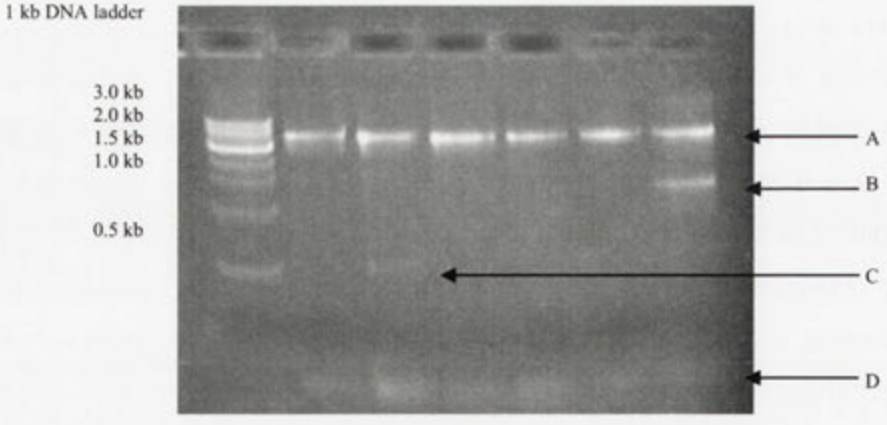


Figure 2.6: SDM-PCR product visualisation with agarose gel electrophoresis.
1 kb DNA ladder was purchased from New England Biolabs (NEB). 4 µL of a 25ng/µL dilution of the marker was used. PCR amplication products of the correct size (~3.7 kb) are indicated by arrow A; non-specific amplication products are indicated by arrows B and C while arrow D shows primer dimers.

After the success of PCR amplification was confirmed, the remaining 48 µL of each SDM-PCR reaction was subject to agarose gel electrophoresis. This time, the gel was illuminated with an UV lamp at 365 nm and the desired DNA bands (~ 3.7 kb in size) were excised. The exposure time of the gel to the weak UV light was kept as short as possible to minimise damage to the DNA. The DNA was then purified with the Qiaquick® Gel Extraction Kit according to the protocol supplied by the manufacturer. The QG buffer from the kit contains guanidine thiocyanate, a chaotropic agent that helped to solubilise agarose at 50°C and to create a suitable pH (≤ pH 7.5) for the adsorption of DNA to the QIAquick silica membrane. The final DNA elution volume was 30 µL.

The additional step of gel purifying PCR products was to ensure that only amplification products of the correct size were selected and to prepare them for a more efficient subsequent *Dpn*I digestion.

2.3.4 *Dpn*I digestion of template DNA

At the end of SDM-PCRs, the reactions would contain a mixture of amplification products and a small amount of template DNA. The template DNA was methylated as it was acquired via *E. coli* cloning. To eliminate the template DNA, *Dpn*I was added to the PCR reactions as the endonuclease specifically recognises and degrades methylated and hemi-methylated DNA cloned from *E. coli*. Although Stratagene recommends the addition of 1 µl of *Dpn*I directly into unpurified PCR products without the aid of any buffer, incomplete digestion that resulted in a high background has been reported in the literature [10]. Any undigested template would have competed with the mutant plasmids to transform competent cells. The *Dpn*I digestion mixture was set up as follows:

DNA from gel purification	30.0 µl
10X NEB buffer 4	3.5 µl
*Dpn*I	1.5 µl
Total volume	35.0 µl

The 1.5 µl *Dpn*I was not added into the reaction all at once but 0.5 µl of the restriction enzyme was added hourly for three hours. Therefore, the PCR products were treated with *Dpn*I at 37°C for three hours, after which the enzyme was inactivated at 80°C for 20 minutes.

2.3.5 Transformation

DH5α cells that were prepared as described in 2.2.1 were thawed on ice for 10 minutes. Electroporation cuvettes with a gap width 0.2 cm were also chilled on ice. 1 µl of the *Dpn*I treated PCR products was added into thawed 50 µl DH5α cell aliquots and mixed thoroughly by gently flicking the Eppendorf tube and pipetting several times. The cells were then transferred directly to the bottom of the chilled cuvettes between the two aluminium electrodes. The cuvettes were incubated on ice again for at least 1 minute. Shortly before electroporation, the cuvettes were tapped against the bench a few times to make sure that the cells were all at the bottom and the moisture on the walls was removed by wiping. The cells were then transformed with a Bio-Rad MicroPulser™ at 2.49 kV for ≥ 4.0 milli-seconds.

Transformed DH5α cells were immediately removed from the electrodes by resuspensing them with 1 mL YENB medium and transferred into a 10 mL sterile pushed cap tube. The recovery culture was then incubated for at least three hours at 37°C to allow ample time for ampicillin resistance to be developed. After that, 100 µl of the culture was spread onto LB plates supplemented with 50 µg/mL ampicillin and incubated overnight at 37°C.

Electroporation cuvettes were rinsed with mQH_2O several times, filled with 70% ethanol and incubated at room temperature for an hour, washed again with mQH_2O thoroughly then dried in a 70°C oven for reuse.

2.3.6 Single-colony mini-cultures and plasmid isolation

Four single colonies from each LBA plate were picked to each inoculate 5 mL of LBA medium and grown in a 37°C shaker overnight. The next day, the plasmid DNA in the LBA cultures was isolated and purified with the Qiaprep® Miniprep Kit by following the instructions in the handbook provided by the kit. The procedure consisted of three main steps. Firstly, bacterial cells were lysed briefly with a sodium hydroxide-containing buffer then the lysate was subsequently neutralised and cleared via centrifugation and pipette aspiration. The plasmid DNA (pDNA) sample that was retained on the QIAprep silical membrane was then purified. In the final step, pDNA was eluted with 30 µl of elution buffer (EB).

2.3.7 DNA verification

Double digestion with *Nde*I and *EcoR*I was used to verify the presence of the gene obtained from each single colony grown and purified with Qiaprep® Miniprep Kit. The digestion reaction was set up as below:

Template pDNA	1.0 µl
10X NEB buffer 4	1.0 µl
*Nde*I	0.2 µl
*EcoR*I	0.2 µl
mQH_2O	7.6 µl
Total volume	10.0 µl

Digestion reactions were incubated at 37°C for 2 hours and then analysed via agarose gel electrophoresis (Figure 2.7).

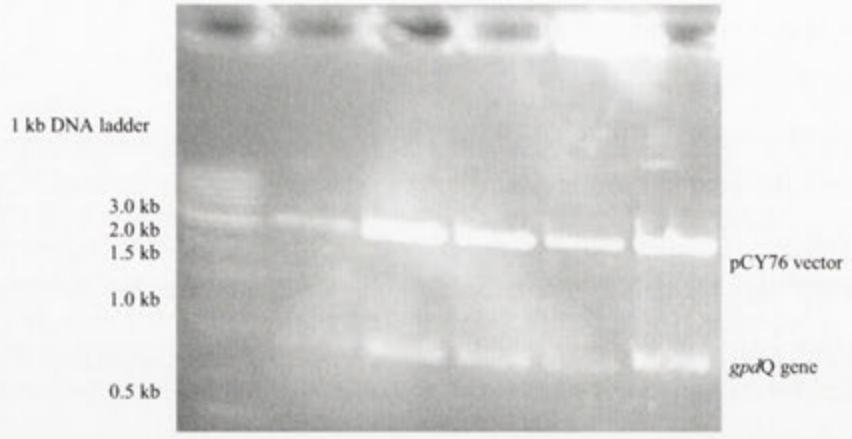


Figure 2.7: Agarose gel presentation of 5 *gpd*Q-pCY76 samples digested with *Nde*I and *Eco*RI.

2.3.8 DNA quantitation, PCR sequencing and ethanol precipitation of DNA

DNA quantitation

The concentration of plasmid DNA of each sample was measured using a NanoDrop® ND-1000 spectrophotometer. Alternatively, the DNA concentration was roughly estimated by comparing the intensity of DNA bands of the samples to the 3.0 kb band of the marker (25 ng/µl).

PCR sequencing

pDNA samples that had been verified to have the gpdQ gene incorporated by restriction enzyme analysis were sequenced. Sequencing reactions consisted of the following:

Template pDNA	~300 ng
Forward primer 4pCY76 (3.2 µM)	1.0 µl
or	
Reverse primer M13 (3.2 µM)	
5X sequencing buffer	4.0 µl
BigDye terminator v3.1	1.0 µl
mQH_2O	To top up to 20.0 µl
Total volume	20.0 µl

Oligonucleotide sequences of primers 4pCY76 and M13 are given in Appendix C. BigDye terminator v3.1 was purchased from Biomolecular Resource Facility (BRF) of John Curtin School of Medical Research (JSCMR) in the Australian National University (ANU) who also supplied the 5X sequencing buffer.

Cycling conditions were: (i) 94°C for 5 min, (ii) 96°C for 10 s, (iii) 50°C for 5 s, (iv) 60°C for 4 min, where steps (ii) to (iv) were repeated 30 times, then held at 4°C until ethanol precipitation.

Ethanol precipitation

Sequencing products were transferred to 1.5 mL Eppendorf microtubes that were clearly labelled on the lids. An ethanol/sodium acetate solution was prepared by combining the following: 3.0 µL of 3M sodium acetate (pH 4.6), 62.5 µl of 95% ethanol and 14.5 µL of mQH_2O for each sequencing sample. 80 µl of the EtOH/NaOAc solution was added to the 20 µl of every reaction mix. The tubes were closed, vortexed briefly and left to incubate at room temperature for 15 minutes so that extension products could be precipitated. The tubes were then spun at maximum speed (16,000 x g) for 20 minutes with a benchtop microcentrifuge. All microtubes were aligned in the same orientation in the centrifuge to ensure the invisible pellet was deposited in the same place within each tube. After that, the supernatant was aspirated very carefully by pipetting away from and without disturbing the pellet (precipitated extension products) which most of the time was invisible to the naked eye. 250 µl of 70% EtOH was added to the pellet and mixed briefly. The microtubes were spun again at maximum speed for 5 minutes in the same orientation as before. All supernatant was then removed carefully and samples were dried in a chemical fume hood for 1 hour or overnight in a drawer before they were sent to BRF (Biomolecular Resource Facility) for direct sequencing. It was important to minimise the exposure of the samples to light, as BigDye terminator is sensitive to light.

pDNA samples whose mutation(s) had been confirmed by sequencing results were stored at -20°C for future use. DNA sequences were analysed using the CLC Sequence Viewer 6 software.

2.3.9 Site-saturation mutagenesis

GpdQ residues C54, Y221 and G259 were subject to randomisation through site-saturation mutagenesis (SSM) [10, 14]. Degenerate primers containing the complete combination of 64 codons at the mutation points are designed to generate the SSM libraries. The oligonucleotide sequences are given in Appendix D.

SSM-PCR reactions were prepared and run as described in sections 2.3.2 to 2.3.5. Ten colonies were sequenced to examine the randomness of the point mutations.

300 single colonies were picked manually from the LBA plates and grown individually in the wells of 96-well round bottom culture plates for kinetic analysis. Screening procedures are detailed in 2.4.2.2.

Difficulties were experienced during site-saturation mutagenesis studies including primer dimerisation and lack of PCR product yield despite numerous attempts to optimise PCR conditions and to redesign primers. Primer design was challenging especially for C54 and G259 as both residues are located in regions that were extremely high in GC content. Site saturation mutagenesis experiments were not pursued any further eventually.

2.4 DIRECTED EVOLUTION

Discussion of experimental procedures used in directed evolution will be sub-divided into two independent steps: (i) genetic library creation, and (ii) developing methods to screen libraries for improved variants. Figure 2.8

summarises in the form of a flow chart, the process of making a GpdQ genetic library in this study.

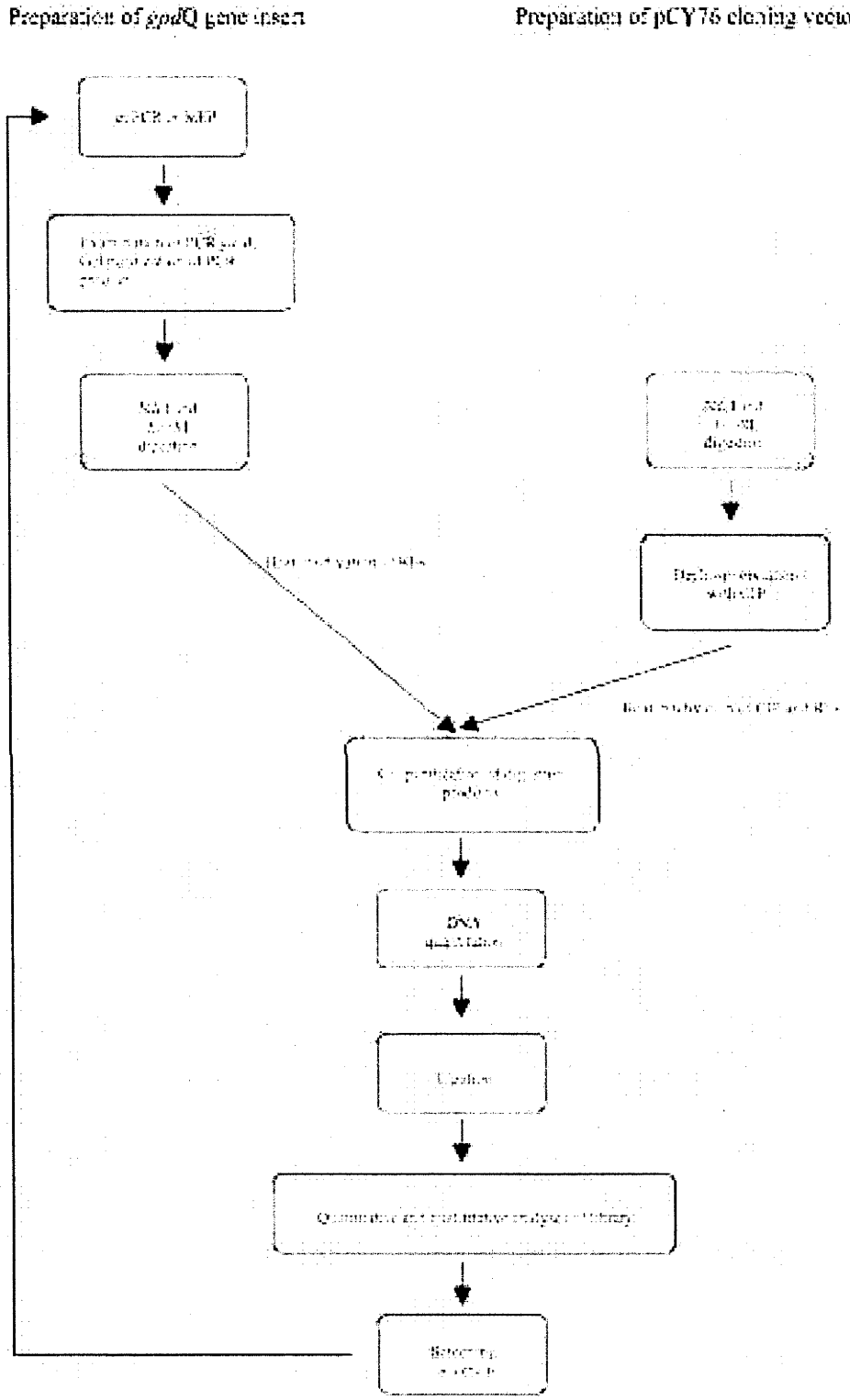


Figure 2.8: The process of creating a library of *gpd*Q mutants.

2.4.1 Genetic library creation

2.4.1.1A Preparation of gene insert

A.1 Error-prone polymerase chain reaction (epPCR)

The overall philosophy of the epPCR technique is to introduce random mutations in a gene of interest. To achieve this, epPCR practitioners use the following strategies to deliberately compromise the fidelity of the amplification process:

i. *An increased $MgCl_2$ concentration.* Mg^{2+} is an essential cofactor for *Taq* DNA polymerase to function properly. 1.5 mM of the metal is usually enough for a standard PCR. By increasing the concentration of Mg^{2+} up to 7 mM in epPCR, the formation of non-complementary pairs is stabilised. The pH is also increased, which possibly has an effect on the polymerase.

ii. *An increased and unbalanced concentration of dNTPs.* Biased dNTP concentrations promote misincorporation of base pairs. dNTPs of about 200 µM each are recommended for a normal PCR. In epPCR, the concentrations of dCTP and dTTP are increased to 1 mM each.

iii. *Addition of $MnCl_2$* to diminish the template specificity of the polymerase.

iv. *The use of Taq polymerase and the increased amount of the enzyme.* *Taq* polymerase is known for its relatively low replication fidelity as it lacks 3'-5' exonuclease proofreading activity. Therefore, *Taq* polymerase is ideal for mutagenic PCRs and its amount is increased to 5 U to promote chain extensions beyond points of mismatch.

v. *Prolonged PCR extension time in each reaction cycle* encourages chain extension past mismatched positions.

vi. *An increased number of reaction cycles* to enhance the overall error rate of the PCR [15-19].

In a pilot study, several libraries were generated with different concentrations of Mn^{2+} ranging from 0.05-0.5 mM. It was decided that the library created with 0.2 mM Mn^{2+} during the epPCR step was best suited for this project as the library was of a good size and of a desirable level of mutagenesis or diversity (Table 2.1), while retaining the protein function. A higher amount of Mn^{2+} used in epPCR greatly reduced the library size. Based on this study, two libraries generated with two different concentrations of 0.2 mM and 0.4 mM Mn^{2+},were selected for more detailed analysis. The high mutation rate also resulted in variants carrying multiple mutations and many of them were therefore inactive (Table 2.2).

$[Mn^{2+}]$ (mM)	0.05	0.1	0.2	0.5
Library size	1×10^5	4×10^4	2×10^4	6×10^3
Error rate (amino acid changes per gene)	0-1	1-2	3-5	5-9

Table 2.1: A comparison of size and error rate among four different libraries generated in this study with different concentrations of Mn^{2+}.
10 clones from each library were sequenced. As the concentration of Mn^{2+} increases, a trade-off between the optimal library size and the error rate can be seen.

Library A (0.2 mM Mn^{2+})					Library B (0.4 mM Mn^{2+})				
3.19	0.85	1.54	0.72	Mean of average rate values = 1.56; Standard deviation = 1.66	0.78	0.76	0.53	0.85	Mean of average rate values = 0.92; Standard deviation = 0.80
1.58	4.49	0.77	1.21		0.66	0.78	0.85	0.58	
0.46	0.96	1.11	0.76		0.77	0.65	0.62	0.62	
0.51	1.33	0.97	1.02		0.69	0.52	4.00	0.81	
0.52	1.01	4.48	0.78		0.80	0.90	0.79	0.47	
0.67	0.68	0.46	7.28		2.85	0.48	0.50	0.83	

Table 2.2: Comparing the mean average specific activities of 24 mutants from each of Libraries A and B.

Libraries A and B from Table 4.2 were created with different concentrations of Mn^{2+}: 0.2 mM and 0.4 mM respectively. 24 variants were randomly

picked as a stochastic representation of each library to be assayed for 1.0 mM b*p*NPP hydrolysis activity with a 96-well plate spectrophotometer. The average reaction rates were measured. The values given in the table are average rates relative to that of the wild-type GpdQ, *ie.* A value of 4.00 means the average rate of the particular mutant is 4 times higher than that of the native enzyme. The mean average activity of a mutant from Library A was ~1.7 times higher than that of Library B, thus supporting the fact that a high error rate employed during epPCR correlates to less active mutants.

The epPCR recipe and program were developed based on published procedures [16-18]. The following materials were combined in a 200 µL PCR reaction tube:

Template DNA (gene within vector)	0.5 µl
(Purified with Qiagen Miniprep kit, about 50 ng/µL)	
Forward primer SY-F (50 µM)	0.5 µl
Reverse primer SY-R (50 µM)	0.5 µl
dNTP (40 mM,10 mM each dNTP)	1.0 µl
10X Mg^{2+} free PCR buffer	5.0 µl
$MgCl_2$ solution (50 mM)	5.0 µl
mQH_2O	Top up to 50 µL final volume
$MnCl_2$ solution (5 mM)	0.5 to 5.0 µl
Taq DNA polymerase	1.0 µl
Total volume	50.0 µl

Notes:

- $MnCl_2$ was added into the reaction mixture just before *Taq* polymerase as $MnCl_2$ may cause the PCR buffer to precipitate.
- Final concentration of primers given above was 0.5 µM. The concentration could be doubled if PCR yield was low.
- Primers contained *Nde*I (SY-F) and *EcoR*I (SY-R) cloning sites as well as sufficient nucleotide bases flanking them for specificity and recognition by the restriction enzymes.
- Final concentrations of $MnCl_2$ in PCR reactions were between 0.05 and 0.5 mM.

- The amplified product PCR product was the size of the gene insert and was therefore smaller than the template pDNA.

PCR parameters were as follows:

94 °C	2 min	
45 °C	30 sec	30 to 40 cycles
72 °C	2 min	
94 °C	30 sec	
45 °C	2 min	
72 °C	5 min	
4 °C	∞	

A.2 Purification of PCR products

The PCR yield was examined with a 1.0% agarose gel electrophoresis with 2 µL of the reaction mix (Figure 2.9).

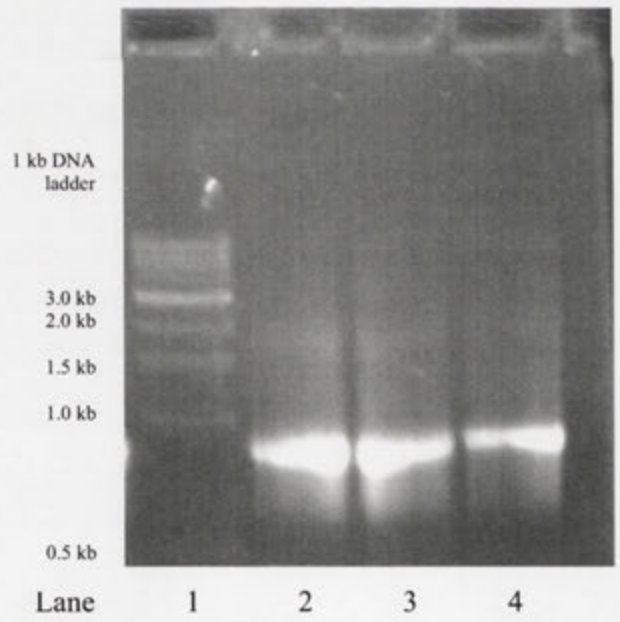


Figure 2.9: An agarose gel showing epPCR amplification products.
Lane 1 is 1 kb DNA ladder; lanes 2-4 show the ~0.825 kb amplification products from three different epPCR reaction samples.

After the PCR yield was checked, the remaining 48 µl of the reaction mix was loaded into a new agarose gel for a second electrophoresis run. The gel was viewed under the weak UV lamp and the amplification products (~ 0.825 kb in size) were excised. The DNA was then purified with the Qiaquick® Gel Extraction Kit as described previously. The final elution volume was 50 µl.

Purifying PCR products with a gel extraction kit took longer than it would with a Qiaquick® PCR Purification Kit. However, the former had an advantage in that PCR products of the correct size could be separated physically (by excising the band from the agarose gel) from other non-specific products that might be found and the wild-type parental DNA itself. Therefore *Dpn*I digestion of DNA template was not required.

Notes:

- Up to 4 PCR reactions were combined into one gel purification sample when the PCR yield was very low.
- TBE buffer system for agarose gel electrophoresis was preferred and used instead of the SB system in directed evolution to offer better a resolution of DNA.

A.3 Digestion with restriction enzymes

Double digestion reactions with *Nde*I and *Eco*RI were comprised of the following:

Purified PCR product	50.0 µl
10X NEB Buffer 4	10.0 µl
mQH_2O	36.0 µl
*Nde*I	2.0 µl
*Eco*RI	2.0 µl
Total	100.0 µl

Samples were incubated in a 37°C water bath for 4 hours after a 3-second quick spin in the microcentrifuge.

The treatment of epPCR amplification products produced DNA fragments that could be ligated into the expression plasmid pCY76. DNA fragmentation of restriction endonucleases (REs) or digestion was the most vital step in ensuring the creation of a genetic library of high stringency. During the course of this research there was difficulty encountered in the Ollis laboratory in producing large mutant libraries. Several measures that were being undertaken to make the digestion work as efficiently as possible. For example, in view of the short half-life of *Nde*I, fresh REs were added to the reaction every hour for 4 hours in 0.5 μL portions so that there was a continual fresh supply of enzymes to cleave the DNA. The following strategies were used to avoid star activity.

i. Use of excessive REs was avoided to circumvent the adverse consequences of star activity (relaxed or altered specificity in REs). The amount of REs in each reaction as given above (40 U for each RE) was already in excess.
ii. Digestion samples were spun regularly for 3 seconds throughout the entire reaction process to clear the vapor formed on the lids which would increase the glycerol concentration of the reaction mixtures, leading to severe star activity.
iii. A greater reaction volume (100 μL instead of 50 μL) was used in order to reduce water vapor during incubation [20].
iv. Digestions were run for 4 hours that was long enough to ensure a complete digestion, yet short enough to avoid star activity.

Digestion was terminated by heat inactivation of the REs at 65°C for 10 minutes.

Subsequent steps in library generation are discussed in sections 2.4.2 to 2.4.5.

2.4.1.1B Preparation of cloning vector

B.1 Preparation of *Nde*I and *EcoR*I cloning sites

Intact or uncut pCY76 plasmids were isolated from *E. coli* DH5α single colonies with a Qiagen Miniprep Kit where the DNA was eluted with 50 µl EB. To set up double digestion with *Nde*I and *EcoR*I, each DNA miniprep was divided into two aliquots. The aim was to ensure a more complete digestion by the REs by lowering the amount of DNA in each reaction, as pCY76 is a high-copy-number plasmid. pCY76 vector and *gpd*Q gene insert digestion samples (as in 2.4.1.A.1) were prepared simultaneously.

pCY76 vector	25.0 µl
10X Buffer 4	10.0 µl
mQH_2O	59.0 µl
*Nde*I	3.0 µl
*EcoR*I	3.0 µl
Total	100.0 µl

Digestion samples were incubated in a 37 °C water bath for 4 hours. 0.75 µl of *Nde*I and 0.75 µl of *EcoR*I were added to the reactions every hour for 4 hours. The procedure was repeated until the total final volumes of *Nde*I and *EcoR*I reached 3.0 µL each. The reaction tubes were quick spun every time after new enzymes were added.

B.2 Dephosphorylation of vector DNA

The pCY76 vector DNA was dephosphorylated with calf intestinal alkaline phosphatase (CIP) to prevent self-circularisation of the linear DNA. This step was performed immediately after the 4-hour digestion. Vector DNA samples were incubated at 37 °C for another 1 hour.

Vector digestion mixture	~100.0 µL
CIP	1.0 – 1.5 µL
Total	~100.0 µL

All enzymes in the reaction mixes were inactivated at 65°C for 10 minutes.

2.4.1.2 Gel purification of vector and insert DNA

Vector and insert digestion reactions were run on a 1% agarose gel with 1.0 µg/µl ethidium bromide in the TBE buffer system. For other purposes, the amount of ethidium bromide added in agarose solutions was 0.5 µg/µl. The concentration of the stain was increased for easy viewing under the mild UV lamp. EtBr is not detrimental to DNA and may even protect it from UV damage [21].

Care was being taken during gel excision to maintain the fidelity of the mutant library. After electrophoresis was completed, the gel was not placed directly under the mild UV lamp but just close enough to see the DNA marker clearly. 10 µl of the marker was used instead of 4 µl for other routine electrophoresis runs, to ensure clear visibility of all the bands in the ladder. A razor was used to mark on the gel where DNA vector and insert bands should be, respectively (by referring to the marker). To confirm that the gel was cut at the right spot, it could be exposed UV light for a short time [22-24].

The gene insert fragments were purified with the Qiagen Gel Extraction Kit as described in 2.3.3. The columns were incubated for 5 minutes at room temperature.

After adding PE buffer before they were centrifuged. In the final step, DNA was eluted with 30 µL EB.

2.4.1.3 DNA quantitation

The concentration of the insert DNA and the vector DNA (prepared separately) was measured with the NanoDrop® spectrophotometer (Figure

2.10). If all previous steps were performed carefully, there should be about 90% recovery of the DNA produced by epPCR. Agarose gel electrophoresis was run to verify the values obtained from the spectrophotometer (Figure 2.11).

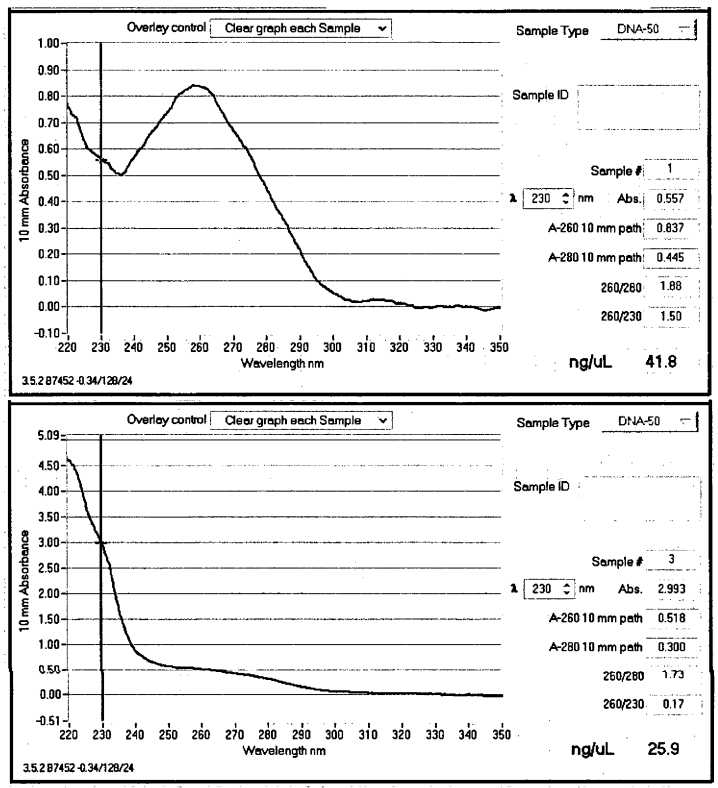
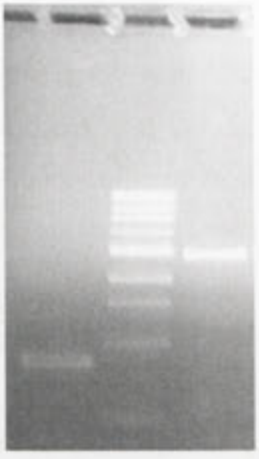


Figure 2.10: DNA UV spectrum measurement graphs.
DNA UV spectrum measurement graphs of pCY76 cloning vector (A) and *gpd*Q gene insert (B).



Lane

1: *gpd*Q gene insert
2: 1 kb DNA ladder
3: Cut and linearised pCY76 vector

Lane 1 2 3

Figure 2.11: A 1.0% agarose gel showing *gpd*Q insert and cut pCY76 plasmid just before ligation.

2.4.1.4 Ligation

Purified *gpd*Q inserts were cloned in the digested pCY76expression vector. Ligation reactions were prepared according to a molar ratio of 3:1 (Insert: vector.

DNA	Size	Molar ratio	DNA amount in ligation
pCY76	~2.9 kb	1	100 ng
*gpd*Q	0.825 kb	3	$\frac{100}{(2.9/0.825)} \times 3 = \sim 85$ ng

pCY76 vector	Adjust accordingly
*gpd*Q gene insert	Adjust accordingly
10X T4 ligase buffer	5.0 µl
mQH_2O	Top up to 50 µl
T4 DNA ligase	1.0 µl
Total	50.0 µl

Ligation mixtures were quick spun before they were incubated overnight at 16°C.

2.4.1.5 Quantitative and qualitative analyses of libraries

T4 ligase was not subject to heat inactivation. 2.5 µl of the ligation mix was used to transform DH5α cells directly without any prior purification step. The salt content in the ligation reaction reduces transformation efficiency and therefore was diluted with 5 µL mQH_2O. The DNA-bacterial cell mixture was blended well by gentle flicking and pipetting to avoid 'arcing' during electroporation. The transformed cells were immediately resuspended with 1 mL YENB medium and put into a 37 °C air incubator for 3 hours. 100 µl of the recovery culture was spread onto two LBA plates (50 µl each plate) and incubated overnight at 37°C. 100 µl of 100% glycerol was added to the remaining 900 µl of the culture. The culture was vortexed, frozen in liquid nitrogen and stored at -80°C.

2.4.1.5A Qualitative analysis

A qualitative analysis of the generated library had two underlying objectives:

i. To determine the intergrity of the library, *ie.* Quantitating the number of transformants that carry an empty plasmid and the number of transformants that do not carry any plasmid.
ii. To ascertain the rate of mutation.

Colony PCR and double digestion were the two methods used to determine the integrity of the library. In both these statistical approaches, a sizable amount of single colonies was randomly picked from the LBA plates.

Colony PCR

Colony PCR reactions consisted of the following:

Single DH5α colony	
SY-F forward primer (25 μM)	1.0 μl
SY-R reverse primer (25 μM)	1.0 μl
dNTP (40 mM)	1.0 μl
10X PCR buffer with Mg^{2+}	5.0 μl
mQH_2O	41.0 μl
Taq DNA polymerase	1.0 μl
Total volume	50.0 μl

Colony PCR parameters were:

94 °C	2 min	
45 °C	30 sec	30 cycles
72 °C	2 min	
94 °C	30 sec	
45 °C	2 min	
72 °C	5 min	
4 °C	∞	

Colony PCR reactions were then analysed on a 1.0% agarose-EtBr gel (Figure 2.12).

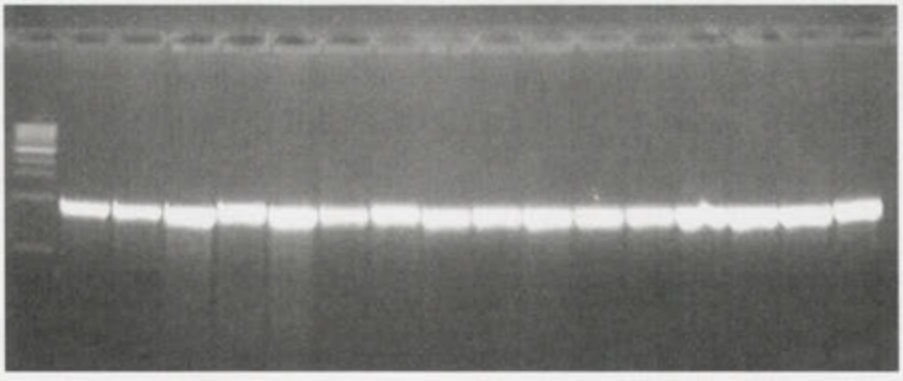


Figure 2.12: 16 colony PCR reactions in an agarose gel presentation.

DNA verification and sequencing

Incorporation of the plasmid was checked by restriction enzyme digestion and sequencing. About ten transformants were randomly selected from the LBA plates (from 2.4.5) to inoculate 5 mL LBA medium. Cultures were grown overnight for plasmid isolation the following day. The purified pDNA samples were digested with *Nde*I and *EcoR*I to check for the presence of gene inserts. Digestion reactions were prepared, run and analysed as described in 2.3.6 and 2.3.7.

The ten variants were also sequenced to determine the mutation rate.

2.4.1.5B Quantitative analysis of library size

The size of the library was estimated by counting the number of colonies on the two LBA plates, deriving the mean average and subtracting the background. The libraries generated during this study using epPCR and StEP were always between 10^4 and 10^5 colony-forming units in size.

2.4.1.6 Creating future libraries (second generation libraries and onward) using Staggered Extension Process (StEP)

StEP is an *in vitro* DNA recombination method first introduced in the late 1990s. It is technically and conceptually simpler than the other older, conventional approach – DNA shuffling. DNA shuffling requires substantial quantities of parental DNA to be digested with DNaseI to create a pool of small fragments that are then reassembled into full-length genes by two different thermocycling reactions (PCR without primers followed by PCR with primers) [25-27].

StEP uses extremely short extension times to induce template switching (Figure 2.13).

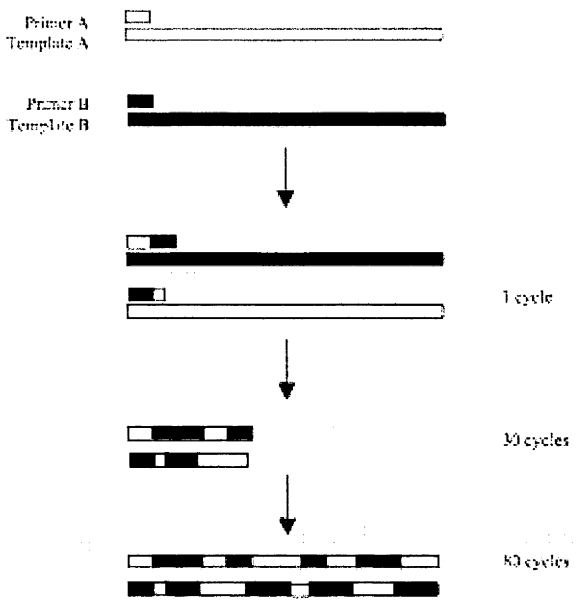


Figure 2.13: A diagram to illustrate the simplified concept of StEP.
In each cycle, the growing fragments or incomplete extension products undergo 'template switching' by annealing to different templates and further extend to become full-length, recombined genes.

In this study, the best mutants from each round of screening were shuffled using the staggered extension process PCR (StEP) method. The StEP reaction contained 0.5 ng template DNA (a equimolar mixture of selected variants), 1 μM of each of the forward (SY-F) and reverse (SY-R) primers, 1.25 mM of dNTP, 1.5 mM $MgCl_2$, 3% DMSO and 0.1 U *Taq* DNA polymerase. DMSO was added to improve the product yield and to reduce the smear in the background, as seen when StEP reactions were analysed on agarose gels.

Four different protocols of StEP were used throughout this study (StEP-A, B, C and D). Initially (second to fourth generation libraries), StEP-A reactions were run at 94°C for 2 minutes (pre-denaturation), 80 cycles of 45°C (10 seconds), 72°C (10 seconds) and 94°C (10 seconds), followed by 45°C for 2 minutes and a final extension at 72°C for 5 minutes. This program was used to create libraries up to the fourth generation. Libraries

generated with this process were small in size and had low recombination efficiency.

Cycling conditions were then optimised, based on Zhao's journals [28-30] and empirical trials. A new protocol StEP-B was introduced to produce libraries from the fifth generation to the seventh. The StEP-B protocol was developed as a two-step cycle (lacking a distinct 72°C elongation step) to minimise extension time to ensure shuffling efficiency:

94°C	2 min	
94°C 55°C	30 sec 5 sec	80 cycles
72°C 4°C	5 min ∞	

Polymerisation and recombination would take place during the extremely short ramp time (55°C to 94°C) of the thermocycler. *Taq* polymerase is one of the quickest DNA polymerases, leading to shorter annealing/extension times that are better for recombination efficiency. Therefore, whenever StEP-PCR reactions were run, the BioRad iCycler® was used as it was the faster thermocycler of the two models available in the laboratory.

DNA sequencing results of the best mutants from the seventh round of directed evolution of GpdQ indicated that, even with a 5-second annealing step at 55°C, StEP-B could not recombine mutations that were as close as 30 nucleotides (10 residues in the translated protein), *eg.* G259R and C269A. Hence, to generate libraries for the eighth round of evolution, we experimented with 3-second (StEP-C) and 1-second (StEP-D) annealing times. Table 2.3 shows the difference in activities between the two libraries when 24 variants from each were assayed for 1.0 mM b*p*NPP hydrolysis.

Library StEP-C (3-second annealing)					Library StEP-D (1-second annealing)				
1.00	0.93	0.88	0.56	Mode = 6.95	3.35	1.16	9.84	1.81	Mode = 9.84
3.19	1.02	3.61	1.78		1.65	1.41	0.77	0.66	
5.59	0.69	2.83	2.13		1.01	0.65	2.94	3.85	
4.37	4.40	1.31	2.84		1.06	5.17	5.28	8.23	
6.12	6.95	0.94	0.79		8.55	0.99	9.15	3.26	
6.86	1.03	0.56	6.41		8.33	1.56	1.77	2.93	

Table 2.3: Comparing the relative average activities of 24 mutants from each of Libraries StEP-C and StEP-D.

The highest relative activity detected among 24 variants from Library StEP-C was 6.95 $\mu mol^{-1}min^{-1}$, compared to 9.84 $\mu mol^{-1}min^{-1}$ from Library StEP-D. Although this was a marginal difference between the two values, we could still infer that the recombination efficiency was improved with a shorter annealing/extension step and this could possibly lead to the discovery of more potent variants with new combinations of mutations during screening. DNA sequencing analysis of a variant from Library StEP-D showed that mutations G259R and C269A were successfully recombined (Figure 2.14).

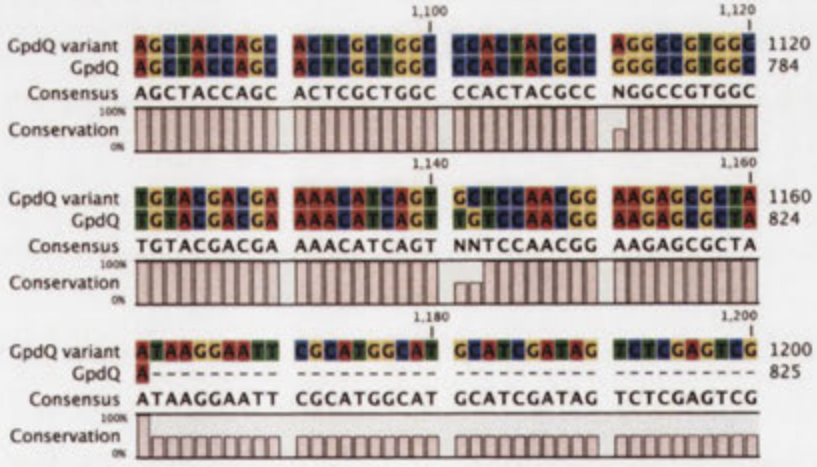


Figure 2.14: DNA sequence alignment of an eighth round variant from Library StEP-D against wild-type GpdQ, showing mutations G259R and C269A.
Nucleotides 775-777 and 805-807 encode for G259 and C269 respectively in the wild-type; whereas nucleotides 1111-1113 and 1141-1143 in the GpdQ variant encode for R259 and A269 respectively.

Hence, the library produced from the StEP-D protocol was used for subsequent eighth round screening. The size of Library StEP-D (3 x 10^4) was significantly smaller than that of Library StEP-C (1 x 10^5). However, this should not be immediately regarded as a disadvantage of using a shorter annealing/extension step, as it was possible that Library StEP-C contained numerous copies of the same small set of poorly recombined mutants. Moreover, the *in vitro* method used to screen the libraries severely limits the number of clones that can be practically examined due to its laboriousness.

The remaining cloning procedures were the same as for the first generation library. The whole process of preparing a library of *gpd*Q mutants is summarised in Figure 2.8.

2.4.2 Developing methods to screen libraries for improved variants

In vivo screening is quick and less laborious compared to *in vitro* random screening, enabling a library as large as 10^6 to be screened. Two *in vivo* agar plate-based methods for the preliminary screening of GpdQ mutant libraries were tested with little success. GpdQ mutant libraries were eventually subject to *in vitro* random screening in the 96-well microtiter plate format.

2.4.2.1 *In vivo* agar plate-based screenings with phosphotriester and phosphodiesters

McLoughlin's growth assay method with phosphodiesters dimethyl phosphate (DMP) and diethyl phosphate (DEP)

A growth-based strategy for screening OpdA mutants with improved MPO activity was devised by Dr Sean Yu-McLoughlin during his PhD candidature in the Ollis laboratory [4, 31]. A catabolic pathway was designed within *E. coli* cells that co-expressed a phosphotriesterase (OpdA), phosphodiesterase (GpdQ) and alkaline phosphatase that would enable the bacteria to utilise triesters as the sole phosphorus source.

For the purpose of screening GpdQ mutant libraries in this project, McLoughlin's plate assay screening was modified slightly. DH5α *E. coli* cells that were transformed with recombinant GpdQ mutants expressed only the phosphohydrolase and not the other transporter proteins of the putative *Enterobacter aerogenes ugp* operon (Figure 1.7). The *E. coli* transformants were grown on modified M9 or potassium morpholinopropane sulfonate (MOPS) [32] minimal medium agar plates containing 50 μg/mL ampicillin and 1 mM DMP or DEP as the sole phosphorus source. There was no cell growth after 2 days of incubation at 37°C, suggesting that the there was not sufficient phosphate from the catalytic breakdown of the alkyl diesters to support growth.

Cho's top agar assay method with phosphotriester methyl paraoxon (MPO)
Cho and colleagues developed a solid phase top agar assay based on the formation of *p*-nitrophenol for rapid preliminary screening of potential OPH variants with enhanced methyl parathion (MPS) activity [33]. Single colonies of transformed DH5α cells were spread onto M9 minimal medium plates supplemented with 0.1% Casamino acids and 50 μg/mL ampicillin and 1 mM $MnCl_2$ to promote growth and GpdQ activity. The minimal medium was used as it would not increase the background colour. After a 48-hour incubation at 37°C, 5 mL of 0.7% agarose containing 50 mM CHES buffer (pH 9.0) and 1 mM methyl paraoxon (solubilised in 10% methanol) was laid over the colonies on the plates. The agarose-substrate solution was pre-boiled and cooled to 55°C before it was poured over the colonies.

Colonies were selected based on the intensity of the yellow colour from the hydrolysed product *p*-nitrophenol, where any clones that appeared to have a larger yellow halo than the wild-type GpdQ would be selected for rescreening using a 96-well plate reader. However, after five hours of incubation at room temperature, no formation of *p*-nitrophenol product could be observed.

2.4.2.2 *In vitro* random screening with bis (*p*-nitrophenyl) phosphate

After the *in vivo* screening experiments were unsuccessful, the GpdQ mutant library was screened *in vitro* with substrate b*p*NPP. Instead of growing the library on LBA selection plates overnight and manually inoculating LBA liquid medium with single colonies, a less laborious method that was first developed by Dr Bradley Stevenson in our laboratory was used [34, 35]. The mutant library was grown in liquid media and screened in 96-well plates. The 1 mL YENB recovery culture (with 10% glycerol) from section 2.4.1.5 was diluted with LBA medium and the dilution was dispensed into 96-well round bottom microtiter plates to an average density of 2 cells per well. 200 µL of the diluted library was pipetted into each well. This method assumes that the enhanced activity of a mutant clone can be detected even when it is grown with other clones in the same well. It also assumes that the cells do not aggregate and divide during dispensation and are evenly distributed in the library, the initial number of viable cells placed in each well will obey the Poisson distribution:

$$P(r) = \frac{e^{-\mu}\mu^{r}}{r!} \qquad \text{(Equation 2.2)}$$

Where P(r) is the probability of a well having r number of cells when the mean number of cells per well is µ. If $\mu = 2$ and $r = 0$, then P(r) would be 0.135, meaning the expected number of sterile wells in each plate (no cell growth) when all the wells contain an average of 2 cells is approximately 13 (0.135 X 96 wells).

The 96-well plates were grown in a 37°C shaker overnight. The next day, the number of sterile wells was counted to confirm the actual mean number of cells in each well. 20 µL of the 200 µL culture in each well was aliquoted into 96-well flat bottom assay plates, lysed with a solution containing 1X BugBuster and 0.5 mM $MnCl_2$ (final metal concentration in the assay reaction), and assayed with b*p*NPP in 50 mM CHES pH 9.0. Production of

p-nitrophenol from the catalytic hydrolysis of b*p*NPP was measured spectrophotometrically in a 96-well plate reader at 405 nm, room temperature for up to 10 minutes. With this *in vitro* screening method, only about 10^4 cells could be screened in each round within a reasonable period of one week.

The culture from wells displaying enhanced activity was streaked on LBA plates and incubated overnight at 37°C. Four single colonies from each streaking was then picked to each inoculate 200 µL fresh LBA medium and grown overnight at 37°C. Secondary screening of the single-colony cultures was conducted using exactly the same methodology described above for preliminary screening, with the wild-type as the control. The activity observed here was typically higher than during preliminary screening as there was no more competition among the different variants within the same well.

5 µL of each of the fittest mutants from every round was used to inoculate 5 mL LBA medium. The cultures were grown overnight in a 37°C shaker. The plasmid DNA was isolated, analysed via an agarose gel electrophoresis after *Nde*I/*Eco*RI digestion and sequenced. Up to 12 of the best mutants that carried different combinations of mutations were subject to recombination using the StEP as described in 2.4.1.6. It has been shown that an efficient pooling strategy greatly enhances the screening throughput and the number of active variants being identified [36]

For the first four rounds, GpdQ was evolved with b*p*NPP using two strategies separately: towards higher affinity with low b*p*NPP concentrations and higher catalytic affinity with high b*p*NPP concentrations. With the first three rounds of high affinity evolution, the libraries were screened with 1 mM, 0.5 mM and 0.25 mM b*p*NPP respectively. From the fifth round onwards, GpdQ libraries were screened with a standard b*p*NPP

concentration of 1 mM. An example of kinetic progress curves from secondary screening is shown in Figure 2.15.

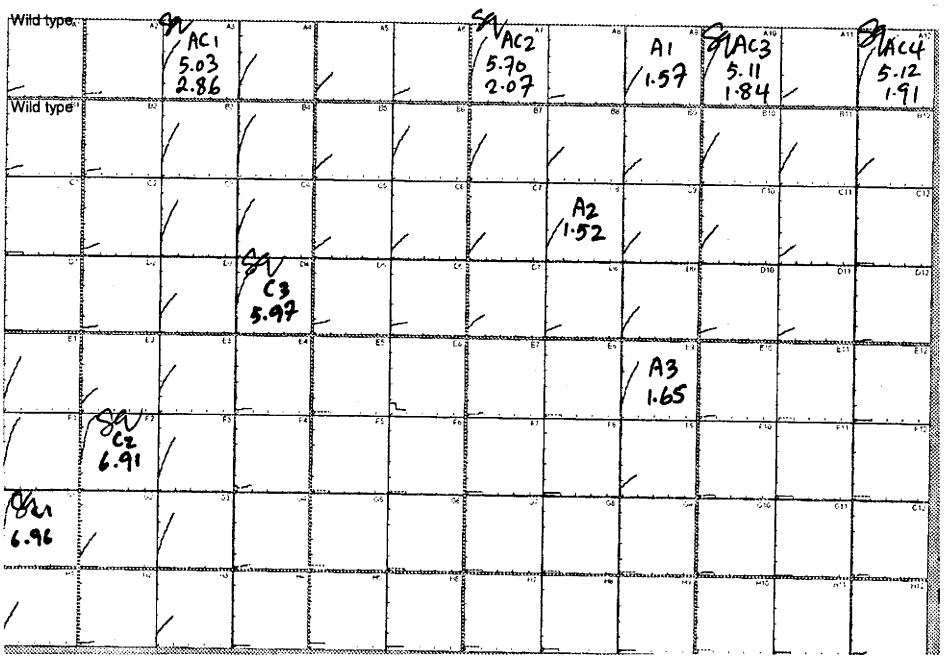


Figure 2.15: An example of a secondary screening during the sixth round of directed evolution.
The two wells A1 and B1 contain wild-type GpdQ. The numbers that were handwritten denote how times the average activity of a particular mutant is higher than that of the wild type. For example, mutant AC1 in well A3 displays average activities that are 5.03 and 2.86 higher than those of the wild-type at b*p*NPP concentrations of 1 mM and 0.1 mM respectively.

2.5 PROTEIN EXPRESSION AND PURIFICATION

2.5.1 Protein expression

pCY76-*gpd*Q plasmid DNA samples, purified from the Qiaprep® Miniprep Kit, were diluted 50 times with mQH_2O. 1 µL of the diluted pDNA was used to transform DH5α cells, which were recovered with 1 mL YENB medium immediately after electroporation. 20 µL of the recovery culture was streaked onto an LBA plate that was then incubated at 37°C overnight. The following day, 2 single colonies were picked to inoculate 5 mL of TBA (Terrific Broth-ampicillin) medium in a culture flask. It was very important

to use single colonies from a fresh transformation to ensure optimal protein expression.

The 5 mL starter culture was grown in a 37°C shaker for 5-8 hours (differing growth rates due to mutations), until the OD_{595} reached about 0.20. 500 mL TBA medium in a 1 L baffled culture flask was then inoculated with the starter culture and grown at 37°C overnight for 16 hours. Cells were then harvested by centrifugation with an SLA-3000 rotor at 4,000 x g for 15 minutes. A 500 mL culture would typically yield 4-5 g of *E. coli* cells. Cells were stored at -80°C.

With the leaky nature of *lac*-controlled constitutive expression, it was sometimes difficult to express recombinant proteins at high levels. pCY76 is a high copy number plasmid whose replication takes place in a relaxed fashion [6]. If the starter culture entered the log phase, it was very possible that the cells would divide vigorously without the correct segregation of plasmids to the daughter cells [37], although the *par* locus was included in the pCY76 vector to circumvent this negative phenomenon. This could potentially pose a threat to the expression of GpdQ. For this reason, the starter culture was not incubated overnight because it was important to arrest the cells at the lag phase during their growth. A starter culture of cell density > 0.2 would significantly reduce the protein expression level during the exponential phase (500 mL culture).

LBA (1 L) was used as the expression media but was later switched to TBA, a richer medium that gives greater cell mass and higher recombinant protein expression levels [38, 39]. Cells were also initially grown at 30°C. Expressing proteins at higher temperatures can sometimes result in production of non-refoldable, inactive aggregates or inclusion bodies [40, 41]. However, this did not dramatically decrease the solubility of GpdQ proteins. Expressing GpdQ at 37°C saved a lot of time as this is the optimal temperature for cell growth. Moreover, the use of TB medium helps to prevent cell death [37].

An SDS-PAGE gel comparing GpdQ expression levels in TBA and LBA media is given in Figure 2.16.

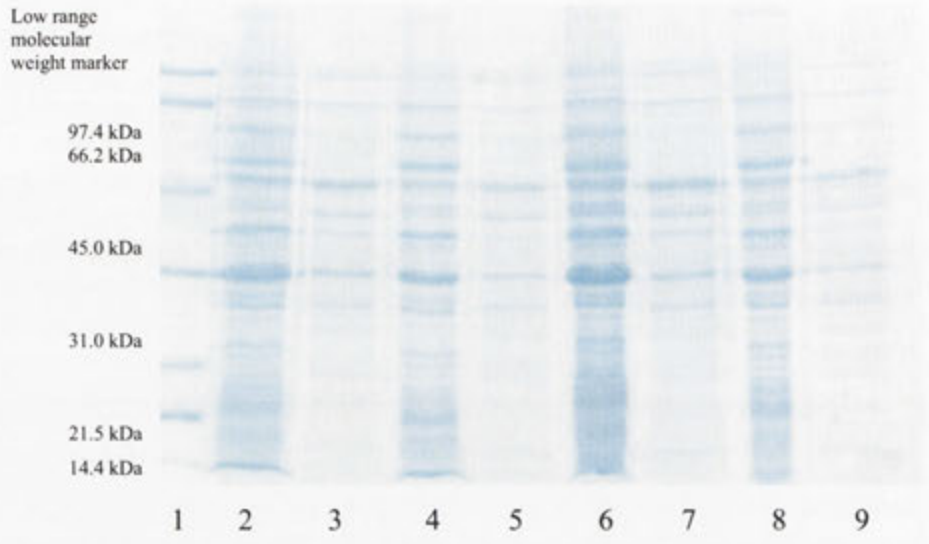


Figure 2.16: A comparison of GpdQ proteins expressed in TBA and LBA media.
Lane 1 is low range molecular weight marker, with the sizes (in kDa) of the protein standards given on the left. The GpdQ monomer (~30.8 kDa) is aligned with the 31.0 kDa carbonic anhydrase band in the marker. Lanes 2, 4, 6, 8: wild-type GpdQ, N53D, C269S and C269A expressed in TBA. Lanes 3, 5, 7, 9: The same proteins expressed in LBA.

2.5.2 Protein purification

ProtParam (http://au.expasy.org/tools/protparam.html) is a tool that allows the computation of various physical and chemical properties of a protein from its sequence. Prior to purification, the theoretical values of molecular weight (monomer), pI and extinction coefficient of every GpdQ protein were calculated based on DNA sequencing of each mutant.

All purification steps were carried out at 4°C to minimise loss due to denaturation. Three protocols were used to purify GpdQ in this study. All buffers and mQH_2O used in purification were filtered with vacuum driven 0.45 µm nitrocellulose membrane filters. An AKTA™ FPLC system was used to automatically control elutions.

2.5.2.1 Protocol A – McLoughlin *et al.* (modified)

Cell pellets were resuspended with 50 mM HEPES, pH 8.0 and lysed with a French® pressure cell press operated at 14,000 lb/in^2. The lysate was then centrifuged at 40,000 x g with an SS-34 rotor. The supernatant was loaded onto a DEAE anion exchange column that was eluted with a 0-0.6 M NaCl gradient applied over 6 column volumes (The column volume was 80 mL) at a flow rate of 1.5 mL/min. GpdQ enzymes eluted between 0.3M and 0.4 M NaCl, depending on the mutation(s) that had been made. It was noted that many mutations changed the pI and the elution point of the proteins.

The eluted fractions (8 mL each) were assayed for phosphodiesterase activity with b*p*NPP as the substrate (Figure 2.17). Fractions that displayed the highest activity were analysed by SDS-PAGE. 10 µL of each fraction was mixed with 10 µL Tris-HCl buffer pH 6.8 that also contained bromophenol blue (BPB), sodium dodecyl sulfate (SDS), glycerol and 5% β-mercaptoethanol. The samples were heated at 100°C for 5 minutes for a complete denaturation before they were loaded into the polyacrylamide gels. 10 µL of Low Range Molecular Weight marker was normally injected into the left most well of the gel. Electrophoresis was run in the Tris-HCl-glycine buffer system using a Hoefer Mighty Small II gel tank at 300 V and 30 mA with tap water cooling for about 40 minutes or until the BPB had migrated to the end of the gel. Recipes for 'cracking' buffer, stacking and resolving polyacrylamide gels, as well as the Tris-HCl-glycine buffer are given in detail in Appendix B. The gel was then detached from the tank, destained with mQH_2O for 30 minutes to remove SDS and later stained with GelCode® blue stain reagent to visualise the protein bands.

Up to 3 fractions with the highest amounts of GpdQ protein were chosen for the next purification step. 3 M ammonium sulphate (AmS) solution was added to equilibrate the fractions up to a final concentration of 1 M. The protein solution was then centrifuged at 30,000 x g for 40 minutes. The

supernatant was loaded onto a Phenyl Sepharose hydrophobic interaction column. A linear AmS gradient (0.6 to 0 M) was applied over 6 column volumes to elute bound proteins at 1.5 mL/min. Each column volume was 80 mL. GpdQ enzymes typically eluted between 0.3 and 0.2 M.

The main disadvantage of this protocol was that the high AmS concentration caused GpdQ to precipitate, leading to a significant loss of the protein during the purification process.

2.5.2.2 Protocol B – Ghanem *et al.* (modified)

Due to the inefficiency of the DEAE-Phenyl Sepharose step, a method based on the one devised by Ghanem and colleagues was used [42]. In brief: cells were resuspended in 50 mM Tris-HCl buffer, pH 8.0. The soluble fraction, obtained after cell disruption and centrifugation, was treated with AmS at 60% saturation. The saturation point of AmS is 3.93 M at 4°C. The protein-AmS mixture was equilibrated at 4°C for an hour before it was centrifuged at 10,000 x g for 30 minutes. The pellet was resuspended with 1 mL elution buffer (50 mM Tris-HCl, pH 8.0) and loaded slowly onto a Superdex 200 gel filtration column with a peristaltic pump. The sizing column had a column volume of 120 mL. The proteins were eluted at 0.5 mL/min to ensure good resolution. Once again, the eluted fractions were assayed for hydrolysis of 5.0 mM b*p*NPP. The existence of the protein of the required molecular weight in fractions with the highest activity was confirmed via SDS-PAGE. Confirmed fractions were pooled and loaded onto a Q Sepharose anion exchange column. Elution was run at 2 mL/min with a 0-0.6 M NaCl gradient applied over four column volumes (Column volume was 80 mL). The ~70% purity of GpdQ derived from this purification method was unsatisfactory.

2.5.2.3 Protocol C – Ghanem's method, 'reversed'

Cells were resuspended with 20 mL 50 mM Tris-HCl buffer (pH 8.0), disrupted and centrifuged as before. The soluble fraction was loaded onto the Q Sepharose column where a 0-0.6 M NaCl gradient was applied over three column volumes. During optimisation of the Q Sepharose elution, it was found that a steeper NaCl gradient gave a better resolution and separation of proteins. GpdQ enzymes could elute anywhere between 0.32 M and 0.50 M NaCl, depending on how radically the mutations change the pI of the proteins. The native protein eluted at about 0.4 M NaCl (Figure 2.18).

The two fractions with the highest GpdQ concentration and least impurities were selected for the next purification step. 6.61 g of ammonium sulfate salt was added to the two protein fractions (16 mL) and topped up to 20 mL with mQH_2O so that the final AmS concentration was 2.5 M. The proteins were equilibrated for an hour at 4°C, then centrifuged as described before. The pellet was resuspended with 1 mL elution buffer and loaded onto the Superdex 200 gel filtration column. The chromatography was run at 0.5 mL/min for the first column volume (elution of proteins) and 1.0 mL/min subsequently (re-equilibration of column). GpdQ eluted at ~60 mL (hexamer), ~72 mL (dimer) and ~78 mL (monomer) of the elution volume. The Superdex 200 column had been calibrated with protein standards (details in section 2.7). A Superdex 200 elution profile of wild-type GpdQ is given in Figure 2.19.

SDS-PAGE indicated that the proteins were at least 95% pure, based on the estimation of the intensity of the protein bands using Adobe Photoshop (Figure 2.20). The purest fractions were dialysed against storage buffer 50 mM Tris, pH 7.5 overnight or for at least four hours. Purified GpdQ proteins were stored at 4°C.

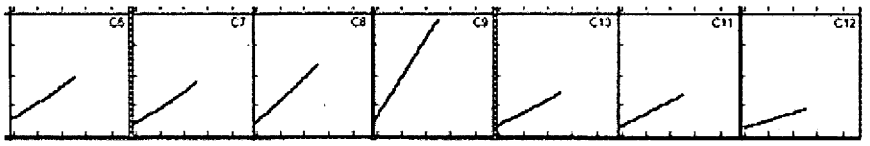


Figure 2.17: Drop assay of fractions eluted from Q Sepharose column. Grids C6-C12 correspond to fractions 8-14 of the elution profile in Figure 2.17.

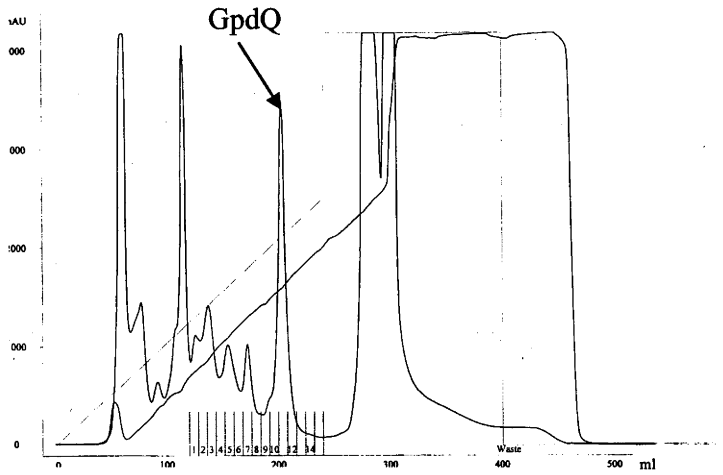


Figure 2.18: An elution profile of the soluble fraction of crude lysate eluted from the Q Sepharose column.
Elution of proteins was detected at 280 nm (blue line). The green line represents the NaCl gradient while the brown line represents the ionic conductivity. Wild-type GpdQ eluted at about 0.42 M (fractions 8-14). Fractions 11 and 12 were selected for further purification with the Superdex 200 column (Figure 2.19).

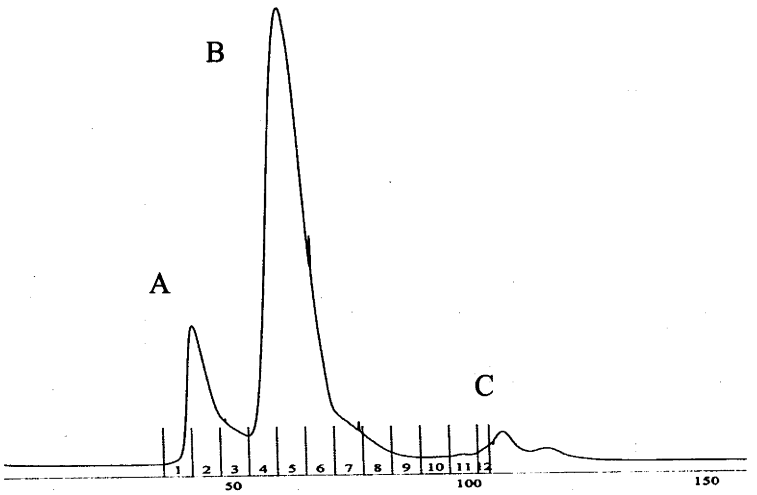


Figure 2.19: A profile of wild-type GpdQ eluted from the Superdex 200 column.

Peak A: Unresolved protein aggregates were the first to elute. Peak B: The ~180 kDa hexamer eluted in fractions 4-6. Peak C: Salts eluted towards the end of the first column volume (120 mL).

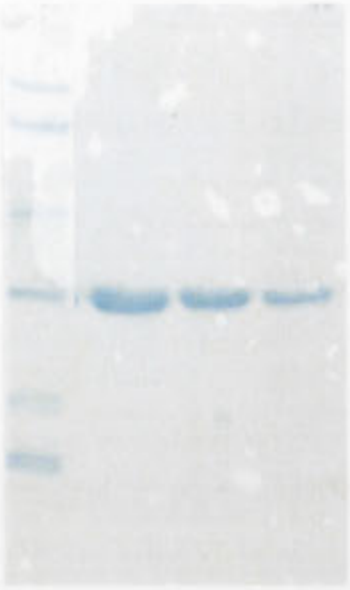


Figure 2.20: Purified GpdQ visualised in a 17.5% SDS-PAGE gel.
The low-range molecular weight protein marker from Bio-Rad was used and visualised on the left most lane of the gel.

Proteolytic activation of GpdQ

GpdQ had been previously reported to undergo proteolytic activation [4, 43]. Purified GpdQ proteins, when visualised in an SDS-PAGE gel (Figure 2.17) were associated with the 31.0 kDa band of the marker. There were other visible faint bands of sizes smaller than the 30.8 kDa GpdQ monomer that could not be removed through any further purification step. These may be GpdQ proteolytic fragments.

The two proteolytic cleavages were predicted to occur between D168-P169 and R227-P228 [44]. Mutagenesis experiments, based on the predictions, were conducted in this study to investigate this proteolysis phenomenon further. Results and discussions are found in Appendix E.

2.6 KINETIC ANALYSIS OF GpdQ PROTEINS

Purified and dialysed GpdQ proteins were assayed for phosphodiesterase activity with b*p*NPP as the substrate. Mutants N80D and N80A were also

assayed for their phosphomonoesterase activity with *p*-nitrophenol phosphate (*p*NPP) that was a product from the hydrolysis of b*p*NPP. Figure 2.21 shows the degradation of b*p*NPP and *p*NPP.

Both b*p*NPP and *p*NPP were obtained from Sigma. b*p*NPP (FW = 362.16) in the form of a sodium salt was ≥99% pure. *p*NPP disodium salt hexahydrate (FW = 371.14) was in the form of 20 mg tablets.

The production rate of *p*-nitrophenolate by wild-type GpdQ and mutants was monitored spectrophotometrically at 405 nm (ε_{405} = 17,000 $M^{-1}cm^{-1}$) using various concentrations of b*p*NPP or *p*NPP. b*p*NPP concentrations ranged from 50 μM-20 mM while *p*NPP concentrations ranged from 0.5-20 mM.

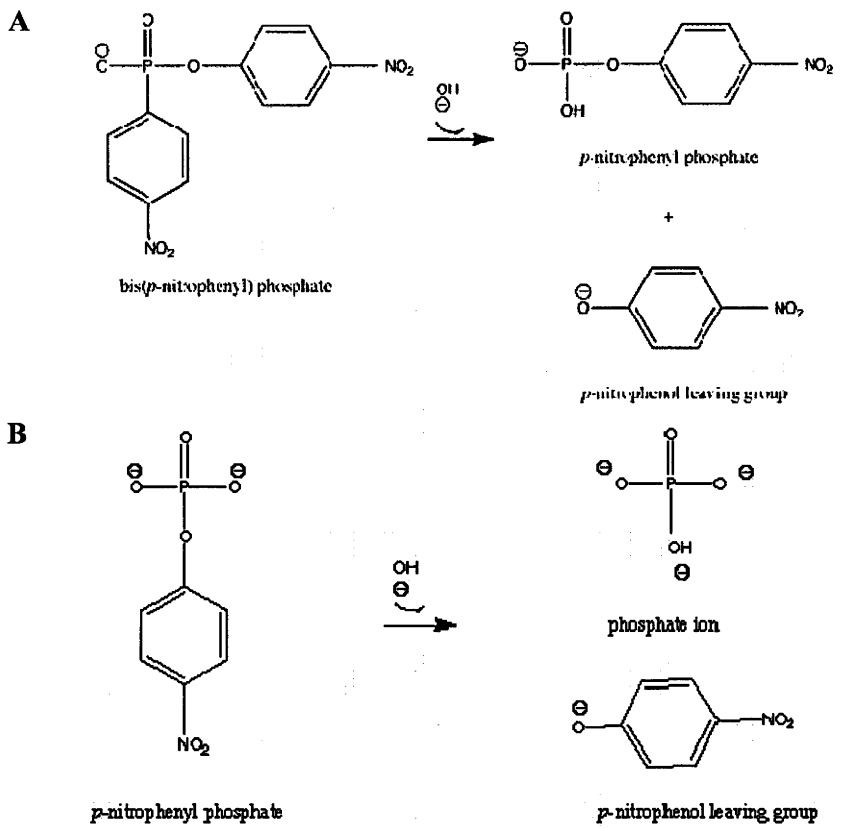


Figure 2.21: Hydrolysis of b*p*NPP and *p*NPP.
Diagram A shows the hydrolysis of b*p*NPP into *p*-nitrophenyl phosphate and *p*-nitrophenol. Diagram B shows the production of phosphate and *p*-nitrophenol from the hydrolysis of *p*NPP.

The most commonly used assay conditions were 50 mM CHES (FW = 207.29) pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL bovine serum albumin (BSA) and 20°C. The $MnCl_2$ was used as an exogenous supply of enzyme co-factors. BSA was used to help stabilise the enzyme. All buffers, substrate and metal solutions were filtered with syringe driven 0.20 μm filter units to reduce the noise due to background during kinetic measurements.

The assays with b*p*NPP were set at an alkaline pH so that the measurements would not be contaminated with phosphomonoesterase activity. It had been shown that GpdQ's monoesterase activity is higher at lower pH values [4, 44, 45]. In some other assays, Mn^{2+} concentrations were varied to investigate the effect on enzymatic activity. To study the role of residue H81, assays were also carried out at pH 6.0 using 50 mM MES (FW = 213.2) as the buffer.

Non-cooperative GpdQ enzymes

For GpdQ enzymes that showed no cooperativity, the kinetic parameters k_{cat} and K_M were determined by fitting the initial velocity (v) data collected at different subtrate concentrations to the classic Michaelis-Menten equation:

$$v = \frac{V_{max}[S]}{K_M + [S]} \qquad \text{(Equation 2.3)}$$

Where v is initial velocity, K_M is the Michaelis constant and [S] is the substrate concentration. The value of k_{cat} is derived from the following equation:

$$k_{cat} = V_{max}/[E] \qquad \text{(Equation 2.4)}$$

And [E] is the enzyme concentration.

The software KaleidaGraph was used for Michaelis-Menten curve fitting purposes.

Negatively cooperative GpdQ enzymes

For GpdQ enzymes that showed negative cooperativity, the Eadie-Hofstee derivative of the Michaelis-Menten was plotted:

$$v = - K_M (v/[S]) + V_{max} \qquad \text{(Equation 2.5)}$$

How apparent values of V_{max} and K_M were obtained from the Eadie-Hofstee plots are detailed in Chapter 3. Microsoft Excel was used for Eadie-Hofstee curve fitting.

Determining cooperativity in GpdQ enzymes

To ascertain cooperativity in GpdQ enzymes, the value of n in Equation 2.5 was determined by using values of V_{max} and K_M as obtained from Equations 2.2 or 2.4. Equation 2.6 is a simplified, linearised version of the Hill equation:

$$\log [v/(V_{max} - v)] = n \log [S] + \log K_d \qquad \text{(Equation 2.6)}$$

where v is initial velocity; V_{max}, maximum velocity; n, Hill coefficient; [S], substrate concentration and K_d, dissociation constant.

2.7 DETERMINATION OF PROTEIN MOLECULAR WEIGHT

All size exclusion columns used in this study were pre-calibrated with a set of six protein standards whose molecular weights were known (Figure 2.22). The linear graph of logarithm of protein sizes in Dalton (Da) versus their elution volumes was plotted and fitted using linear regression (Figure 2.23). GpdQ protein sizes were subsequently calculated based on the equation.

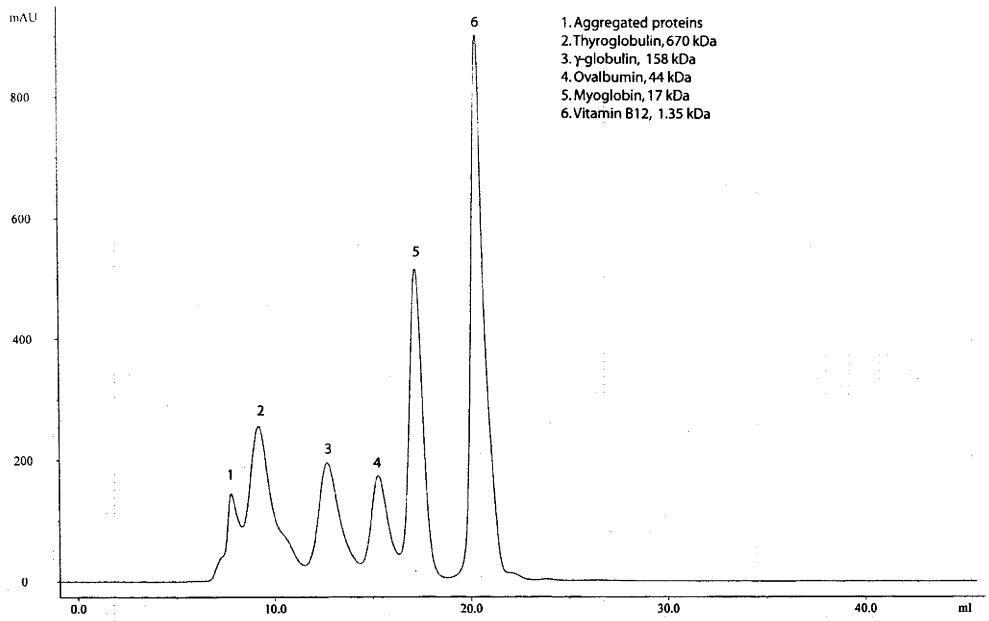


Figure 2.22: Elution profile of molecular weight markers from Superdex 200 size exclusion column.

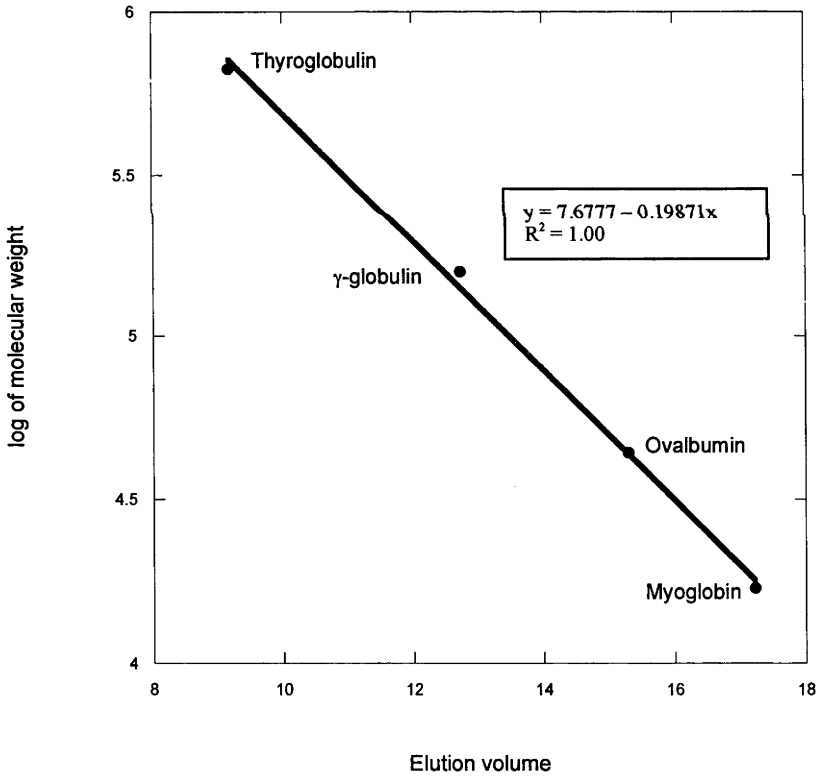


Figure 2.23: Calibration curve obtained with molecular weight markers eluted from Superdex 200 size exclusion column.

The linear equation derived from the curve fit is given in the box at the right hand corner. Vitamin B_{12} has been excluded in order to give a good fit.

2.8 PROTEIN CRYSTALLOGRAPHY

2.8.1 Vapour diffusion of hanging drops

GpdQ proteins that were to be crystallised were subject to a second gel filtration elution to enhance their purity. The proteins were then concentrated with Amicon Ultra-15 centrifugal filter units (30 or 10 kDa MWCO) that were spun at 3,000 x g in the 11133 rotor. The proteins were adjusted to a final concentration of 10-20 mg/mL with mQH_2O and filtered with a syringe driven Millex®-GV 0.22 μm filter unit.

For wild-type GpdQ and other mutants whose hexameric structures were intact, crystallisation conditions were 60% (v/v) Tacsimate, 0.1 M Bis-Tris propane, pH 7.5, with or without 1 mM $MnCl_2$ or $CoCl_2$, and 18°C. GpdQ was crystallised via vapour deposition of hanging drops. 1 mL of the 60% Tacsimate solution was laid into the reservoirs of a 24-well crystallisation tray. The sides of the wells were greased with petroleum jelly before use. 1 μL of the reservoir solution was mixed with 2 μL of protein on a plastic coverslip. If metals ware used, 1 μL of 4 mM $MnCl_2$ or $CoCl_2$ was also added into the mixture. The coverslip was then overturned to cover the wells.

Crystals grew within three days but were left to grow for at least a further four weeks before they were mounted, with the assistance of Drs Paul Carr and Jee-Loon Foo. Crystals were scooped up with a nylon loop, soaked in a cryoprotectant solution (80% Tacsimate) and flash-frozen in liquid nitrogen. The mounted crystals were then irradiated with monochromatic X-rays to test their ability to diffract well. Irradiation experiments, data collection and analyses were all done by Dr Paul Carr.

A crystal of wild-type GpdQ that diffracted to a resolution of 1.9 Å was obtained in this study [44]. This structure (PDB ID 3D03) was in a different

space group ($P2_12_12_1$) from those previously solved by Jackson *et al.* ($P2_13$) [46]. The 3D03 crystal was orthorhombic in shape while those reported by Jackson *et al.* had cubic morphology.

2.8.2 Crystallisation screens

New crystallisation conditions were sought for GpdQ dimer proteins like the F21K mutant using Hampton Research (Index, Crystal HT, PEG, PEG/Ion) and Jena Bioscience HT screens. A Cartesian dispensing robot was used to set up all the screening trays. 100 µL of each screening solution was dispensed into the reservoir wells while 0.2 µL of it was taken out to mix with the ~20 mg/mL F21K purified protein. These crystallisation screens were set up by Miss Tracy Murray.

2.8.3 Computer software programs

Analyses of crystal structures of GpdQ and other structurally and/or functionally proteins were performed using the software Coot Auto Opener with files downloaded from Protein Data Bank (http://www.rscb.org/pdb/). Structural diagrams presented in this thesis were generated using the MacPymol program.

CHAPTER TWO REFERENCES

1. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) High-efficiency transformation of *Escherichia coli* by high-voltage electroporation. *Nucleic Acids Research 16*, 6127-6145.
2. Ymer, S. (1991) Heat Inactivation of DNA ligase prior to electroporation increases transformation efficiency. *Nucleic Acids Research 19*, 6960-6960.
3. Nishimura, A., Morita, M., Nishimura, Y., and Sugino, Y. (1990) A Rapid and highly efficient method for preparation of competent *Escherichia coli* cells. *Nucleic Acids Research 18*, 6169-6169.
4. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of Escherichia coli coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Applied and Environmental Microbiology 70*, 404-412.
5. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *Journal of Bacteriology 190*, 2597-2606.
6. Jeffries, C. (1998) Enzyme engineering: Active site modifications of dielelactone hydrolase. *The Australian National University.*
7. Love, C. A., Lilley, P. E., and Dixon, N. E. (1996) Stable high-copy-number bacteriophage lambda promoter vectors for overproduction of proteins in Escherichia coli. *Gene 176*, 49-53.
8. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology 185*, 60-89.
9. Chambers, S. P., Prior, S. E., Barstow, D. A., and Minton, N. P. (1988) The Pmtl Nic-Cloning Vectors .1. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene 68*, 139-149.
10. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research 32*, -.
11. Liu, H. T., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology 8*, -.
12. Chakrabarti, R., and Schutt, C. E. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene 274*, 293-298.
13. Kureishi, A., and Bryan, L. E. (1992) Pre-boiling high GC content, mixed primers with 3' complementation allows the successful PCR amplification of *Pseudomonas aeruginosa* DNA. *Nucleic Acids Research 20*, 1155-1155.

14. Parikh, M. R., and Matsumura, I. (2005) Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. *Journal of Molecular Biology 352*, 621-628.
15. Leung, D. W., Chen, E. and Goeddel, D. V. (1989) A method for random mutagenesis of a defined DNA fragment using a modified polymerase chain reaction. *Technique 1,* 11-15.
16. Cadwell, R. C., and Joyce, G. F. (1994) Mutagenic PCR. *PCR-Methods and Applications 3*, S136-S140.
17. Wilson, D. S., and Keefe, A. D. (2001) Random mutagenesis by PCR. *Current Protocol in Molecular Biology Chapter 8*, Unit 8.3.
18. Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods and Applications 2*, 28-33.
19. Cirino, P. C., Mayer, K. M., and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods in Molecular Biology 231,* 3-9.
20. Wei, H., Therrien, C., Blanchard, A., Guan, S. X., and Zhu, Z. Y. (2008) The fidelity index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Research 36*, -.
21. Cariello, N. F., Keohavong, P., Sanderson, B. J. S., and Thilly, W. G. (1988) DNA damage produced by ethidium bromide staining and exposure to ultraviolet light. *Nucleic Acids Research 16*, 4157-4157.
22. Zimmermann, M., Veeck, J., and Wolf, K. (1998) Minimizing the exposure to UV light when extracting DNA from agarose gels. *Biotechniques 25*, 586-586.
23. Hengen, P. N. (1997) Methods and reagents - Protecting vector DNA from UV light. *Trends in Biochemical Sciences 22*, 182-183.
24. Grundemann, D., and Schomig, E. (1996) Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *Biotechniques 21*, 898-903.
25. Stemmer, W. P. C. (1994) DNA Shuffling by random fragmentation and reassembly - in-vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America 91*, 10747-10751.
26. Patten, P. A., Howard, R. J., and Stemmer, W. P. C. (1997) Applications of DNA shuffling to pharmaceuticals and vaccines. *Current Opinion on Biotechnology 8,* 724-733.
27. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature 391*, 288-291.
28. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology 16*, 258-261.
29. Zhao, H. M. (2004) Staggered extension process in vitro DNA recombination. *Protein Engineering 388*, 42-49.

30. Zhao, H. M., and Zha, W. J. (2006) In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature Protocols 1*, 1865-1871.
31. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. (2005) Increased expression of a bacterial phosphotriesterase in *Escherichia coli* through directed evolution. *Protein Expression and Purification 41*, 433-440.
32. Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) Culture Medium for Enterobacteria. *Journal of Bacteriology 119*, 736-747.
33. Cho, C. M. H., Mulchandani, A., and Chen, W. (2002) Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. *Applied and Environmental Microbiology 68*, 2026-2030.
34. Stevenson, B. J. (2006) Directed evolution of pyruvate decarboxylase for in vitro glycolysis. *The Australian National University Chapter 6,* 117-189.
35. Stevenson, B. J., Liu, J. W., and Ollis, D. L. (2008) Directed evolution of yeast pyruvate decarboxylase 1 for attenuated regulation and increased stability. *Biochemistry 47*, 3013-3025.
36. Polizzi, K. M., Parikh, M., Spencer, C. U., Matsumura, I., Lee, J. H., Realff, M. J., and Bommarius, A. S. (2006) Pooling for improved screening of combinatorial libraries for directed evolution. *Biotechnology Progress 22*, 961-967.
37. Feinbaum, R. (2001) Introduction to plasmid biology. *Current Protocols in Molecular Biology Chapter 1*, Unit 1.5.
38. Tartof, K.D., and Hobbs, C.A. (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Research Laboratories. Focus 9*, 12.
39. Sambrook, J., and Russell, D. W. (2001) Molecular cloning: A laboratory manual, 3rd edition. *Cold Spring Harbor Laboratory Press.*
40. Al-Samarrai, T. H., Kirk, C. A., Jones, W. T., Harvey, D., and Sun, X. L. (2007) Expression in *Escherichia coli* and in vitro refolding of the plant transcription factor *Arabidopsis thaliana* RGL3. *Protein Expression and Purification 53*, 289-292.
41. Sorensen, H. P., and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli. Journal of Biotechnology 115*, 113-128.
42. Ghanem, E., Li, Y. C., Xu, C. F., and Raushel, F. M. (2007) Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry 46*, 9032-9040.
43. Jackson, C. J. (2006) Enzymatic catalysis of phosphate ester hydrolysis. *The Australian National University.*

44. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of the American Chemical Society 130*, 14129-14138.
45. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural flexibility enhances the reactivity of the bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.
46. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.

3

MUTATIONAL ANALYSIS OF ACTIVE SITE RESIDUES AND MECHANISTIC STUDIES OF GpdQ

CHAPTER THREE: MUTATIONAL ANALYSIS OF ACTIVE SITE RESIDUES AND MECHANISTIC STUDIES OF GpdQ

3.1 INTRODUCTION

This chapter chronicles experiments that were carried out with the purpose of investigating the following:

- The roles and functions of active site residues, in particular the ligands coordinating the metal ions.
- The catalytic mechanism of GpdQ .

Site-directed mutagenesis was the main technique employed to produce active site mutants. These mutants are listed in Table 3.1 below.

Residue	Substitutions	Significance of residue
D8	N, A	A ligand of α-metal.
H10	A	A ligand of α-metal.
D50	N, A	The bridging ligand of α- and β-metal.
N80	D, A	A ligand of β-metal. Mutant 3-25-N80A was also made to study cooperativity between the two metal binding sites.
H81	Q, A	Conserved active site residue with putative catalytic roles.
H156	Q, A	A ligand of β-metal.
H195	Q, N, A	A ligand of β-metal.

Table 3.1: List of active site mutants made, characterised and discussed in this study.

Jackson *et al.* first detected the possible difference in the metal-binding affinities at the GpdQ binuclear active site [1]. Using the anomalous data collected at the Zn K-edge, Jackson *et al.* generated a Bijvoet-difference

Fourier map of the active site of the Zn^{2+}-GpdQ holoenzyme. The map demonstrated the respective nearly full (96%) and partial (80%) occupancies of the α- and β-sites with Zn^{2+} ions, which indicated the difference in metal binding affinities between these two sites. This observation agreed well with the ligand coordination geometries at the GpdQ active site: both α- and β-metal ions are coordinated by two histidine ligands each and jointly, the bridging D50 ligand, but the α-metal has another strong acidic ligand in asparatate (D8) whereas the β-metal is additionally coordinated by an asparagine residue (N80) (Figures 1.13 and 3.1).

One of the key research objectives of this thesis was to characterise the binding affinities of the binuclear active site. To this end, mutants D8N/A, D50N/A, N80D/A were constructed. D → N mutations neutralise the negative charge of the acid residue and are expected to reduce but not abolish the ability of the residue to bind metal ions altogether, since asparagine is capable of forming hydrogen bonds. N → D mutations have the opposite effects and increase the metal binding affinity of a residue. D → A mutations abolish the ability of the residue to coordinate metals completely.

Histidine residues commonly act as ligands in metalloenzymes, especially zinc enzymes. The zinc ions can be substituted by other metal ions like Mn^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+} and others. Site-directed mutagenesis is a common tool employed to identify metal ligands and such mutations often lead to a huge loss in the catalytic activity of the enzyme [2-5]. Mutants H10A, H156Q/A and H156Q/N/A were made to investigate the importance of these histidine residues as metal ion ligands. With these ligands, the rationale behind making the H → Q/N mutations was to deprive the residue of the imidazole ring that coordinates metal ions. The chemical structures of glutamine and asparagine are the closest to histidine among the 20 amino acids as each of them carries a nitrogen atom in the side chain. Therefore, H → Q/N mutations were expected to cause the least disruption to the

conformation of the enzyme, especially the active site. Once again, no metal ion coordination and hydrogen bond formation would be possible when a histidine was substituted by an alanine (H → A).

H81 does not coordinate any metal ions directly but is an active site ligand that is highly conserved in several structurally related phosphodiesterase and phosphatases. Therefore H81Q/A mutants were made to study the role(s) of this residue in catalysis/substrate binding. Such mutations also eliminated the ability of the residue to donate and accept protons. Residues homologous to H81 in other related enzymes have been postulated to participate in general acid/base catalysis.

Figure 3.1 shows a diagram of the active site of GpdQ and all the active site mutants made in this study.

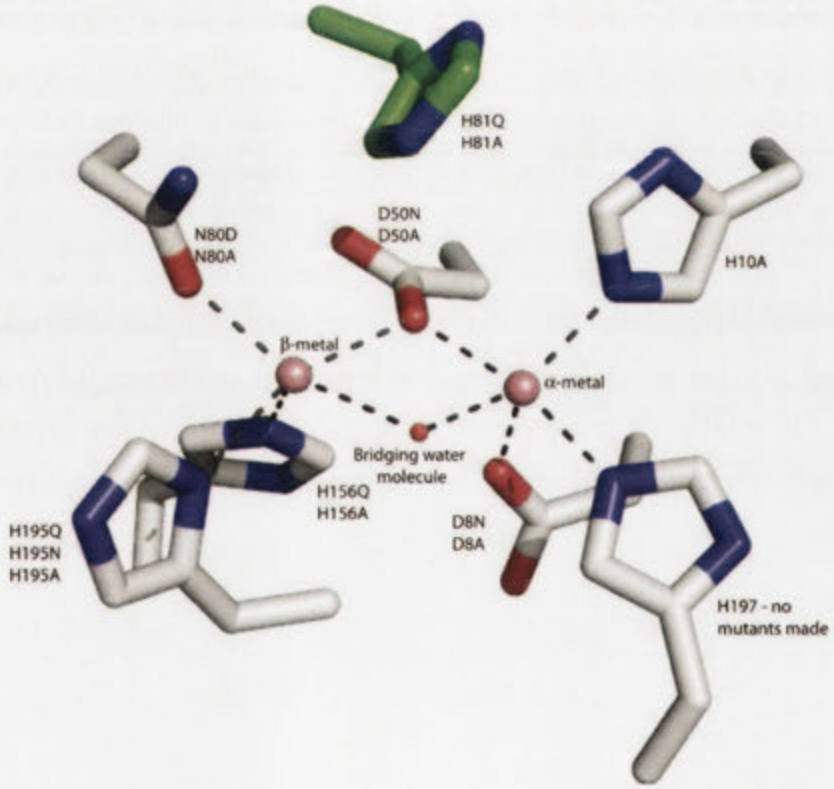


Figure 3.1: The active site of GpdQ and active site mutants made in this study.

3.2 A SUMMARY OF MUTAGENESIS STUDIES OF GpdQ ACTIVE SITE RESIDUES

All active site mutants discussed in this chapter were made, expressed and purified using methods described in Chapter 2. Mutants N80D/A, D8N/A, D50N/A and H81Q/A were made in years 2006-2007. At that time, Protocol B as described in section 2.5.2 of Chapter 2 was used to purify these proteins. The purified proteins were then preliminarily assayed for their bis (*p*-nitrophenyl) phosphate (b*p*NPP) activity at pH 9.0, 20°C and in the presence of 0.5 mM $MnCl_2$ and 0.1 mg/mL BSA. In 2009, mutagenesis studies of GpdQ active residues resumed with the production of new mutants H10A, H156Q/A and H195Q/N/A. These proteins were purified using Protocol C (section 2.5.2, Chapter 2). The mutants made in 2006-2007 were also expressed again and purified using Protocol C. Kinetic analyses of the old mutants was repeated with b*p*NPP at the same assay conditions and indicated no major differences from the kinetic data obtained two years ago. The biphasic behaviour of N80A kinetics was studied further in 2009 by varying assay conditions and the H81 mutants were also assayed at pH 6.0.

Table 3.2 is a summary of results from kinetic assays of GpdQ active site mutants with substrate b*p*NPP under the default assay conditions of 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$. 0.1 mg/mL bovine serum albumin (BSA) at 20°C. Discussions of these results will be presented systematically throughout this chapter. Kinetic data of assays performed under different conditions are not included in Table 3.2.

Mutant	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
Wild type	7.50 ± 0.1	2.10 ± 0.08	3.57×10^3
N80D	0.03 ± 0.002	1.59 ± 0.20	1.84×10^1
N80A	2.28	38.62	5.90×10^1
	0.17	0.39	4.26×10^2
H156Q	0.57 ± 0.01	5.03 ± 0.34	1.14×10^2
H156A	1.51	34.28	4.41×10^1
	0.19	1.62	1.16×10^2
H195Q	0.99 ± 0.04	4.03 ± 0.61	2.45×10^2
H195N	3.57 ± 0.008	3.04 ± 0.26	1.18×10^3
H195A	1.72 ± 0.05	6.75 ± 0.53	2.55×10^2
D8N	0.01 ± 0.0004	1.85 ± 0.37	7.73×10^0
D8A	0.002 ± 0.00005	1.78 ± 0.22	1.13×10^0
H10A	5.35 ± 0.03	2.58 ± 0.60	2.08×10^3
D50N	52.45 ± 1.3	6.70 ± 0.45	7.60×10^3
D50A	4.83 ± 0.14	10.03 ± 0.68	4.81×10^2
H81Q	0.92 ± 0.01	0.96 ± 0.06	9.60×10^2
H81A	4.83 ± 0.14	3.39 ± 0.44	2.17×10^2

Table 3.2: Kinetic parameters of GpdQ active site mutants.
The enzymes were assayed under the conditions of 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$. 0.1 mg/mL bovine serum albumin (BSA) at 20ºC.

The kinetic data on Table 3.2 shows in addition to the N80D mutant, the two D8 mutants and the D50N mutant, most mutagenesis of the active site residues reduced the activity of the GpdQ enzyme but did not completely abolish it. D50N was the only active site mutant whose k_{cat} was enhanced by mutagenesis by more than three-fold. As for the K_M values, some mutations improved substrate affinity while others reduced it.

Two mutants, namely N80A and H156A, exhibited activities that deviated from Michaelis-Menten kinetics. The kinetic data of these two mutants was treated differently to derive two sets of apparent values for k_{cat} and K_m. Results and discussions on N80A and H156A are also presented separately from the other β-metal ligand mutants in section 3.3.4.

3.3 MUTATIONAL ANALYSIS OF β-METAL LIGANDS

3.3.1 N80D

As pointed out in Chapter 1, the loose metal binding affinity of the β-site could be somewhat predicted through the ligand coordination geometries at the active site of GpdQ. The α-metal is coordinated by two histidines and an aspartate. Its β-counterpart also has two histidines but the third ligand is a polar but uncharged asparagine, which is N80. Therefore, N80D and N80A were the first active site mutants constructed in this project.

N80 was mutated to an aspartate with the objective of increasing the metal binding affinity at the β-site. This mutation caused GpdQ to lose its phosphodiesterase activity, with a 250-fold drop in the catalytic turnover (Table 3.2). Another research group working on the structurally related mammalian protein phosphatase 1 (PP-1) made a similar mutation (N124D) and that resulted in a 12.5-fold loss in catalytic activity (k_{cat}) [6]. Although N80D increased the substrate affinity (by reducing K_M) in GpdQ, an increase in K_M was observed in N124D of PP-1. The slight improvement in K_M in N80D could be explained by the notion that the β-metal is associated with the substrate molecule. If the product also leaves together with β-metal, this would justify the need for the metal binding at this site to be flexible.

N80A was catalytically active but deviated from Michaelis-Menten kinetics. This behaviour was different to the equivalent N97A mutant of the Rv0805 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* that was found to retain ~1% of the wild-type activity with b*p*NPP [2].

The biphasic kinetics of N80A is discussed in section 3.3.4.

3.3.2 H156Q, H195Q, H195N and H195A

3.3.2.1 H156Q

H156 is one of the two histidine residues that coordinate the β-metal. When H156 was mutated to a glutamine, the catalytic turnover was more than 10 times lower than that of the wild type, while the K_m suffered from a 2.4-fold increase (Table 3.2). As glutamine is a polar residue it should still be able to coordinate the β-metal. However, glutamine is a neutral residue as opposed to histidine, a basic residue. Such a difference might have resulted in detrimental changes to the active site conformation.

H156 was also mutated to an alanine. As H156A displayed biphasic kinetics, it will be discussed together with N80A in section 3.3.4.

3.3.2.2 H195 – H195Q, H195N and H195A

H195 is the second of the two histidine residues coordinating the β-metal. H195 was mutated to a glutamine (H195Q), asparagine (H195N) and alanine (H195A). Results of kinetic assays performed under conditions previously described for H156 mutants are given below in Table 3.2.

The kinetic results of H156Q, H195Q, H195N and H195A mutants can be explained if we assume that they decreased in the metal binding affinity at β-site in the mutants that led to a decrease in the catalytic rate (k_{cat}) and an increase in K_M relative to the wild type.

Shenoy *et al.* produced a H207A mutant of the Rv0805 protein that is homologous and has a similar effect to GpdQ's H195A. H207A also suffered from drastic reductions in both the catalytic rate and the substrate affinity [2].

3.3.3 Magnetic circular dichroism (MCD) studies on N80 mutants

Magnetic circular dichroism (MCD) experiments were performed on the wild-type GpdQ and two N80 mutants (N80D and N80A) by collaborators at the University of Queensland. These experiments were carried out with the purpose of characterising the α- and β-metal binding sites of GpdQ. With the wild-type GpdQ, the addition of two equivalents of Co^{2+} to the apoenzyme gave rise to a fully occupied α-site while to populate the β-site, an additional 48 equivalents of Co^{2+} were added.

Up to 100 equivalents of Co^{2+} were added gradually to N80A apoenzyme. The α-site was fully occupied with the metal ions but not the β-site. However, it was interesting to find that after the addition of 10 equivalents of inorganic phosphate P_i, the formation of a binuclear active site was observed. This observation was in agreement with the kinetic data in Table 3.6, where the presence of P_i appeared to promote metal binding at the β-site.

In contrast, a fully occupied binuclear metal center was formed in N80D after the addition of only two equivalents of Co^{2+} to the apoenzyme, in the absence of phosphate. This indicated an increased binding affinity for this mutant, particularly for the metal ion in the β-site [7].

3.3.4 N80A and H156A – Biphasic kinetics and negative cooperativity

It was found that N80A and H156A displayed biphasic, non-Michaelis-Menten kinetics, which is characterised by continuously increasing velocity (v) at high substrate concentrations [S], as shown in Figure 3.2 below.

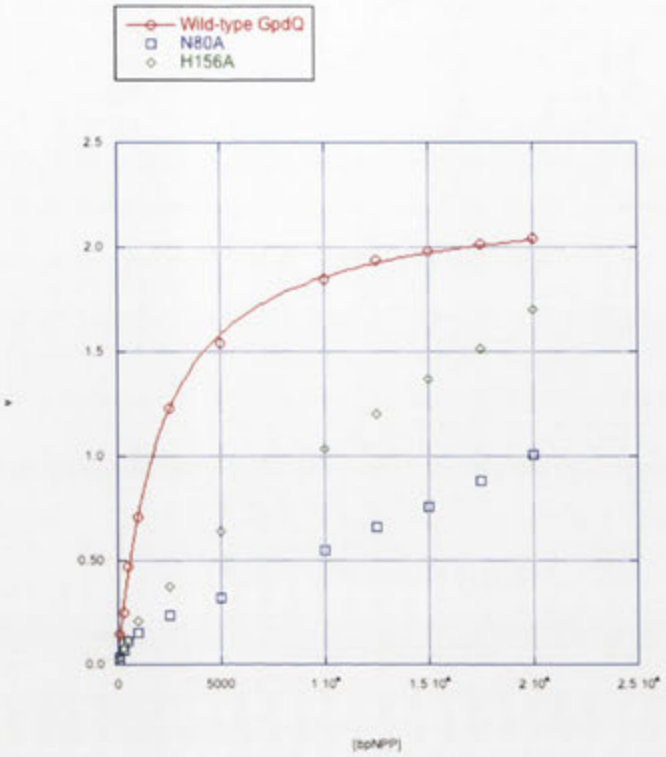


Figure 3.2: Michaelis-Menten plots of wild-type GpdQ, N80A and H156A.
The enzymes were assayed with b*p*NPP in 50 mM CHES (pH 9.0), 0.5 mM $MnCl_2$, 0.1 mg/mL BSA at 20°C.

3.3.4.1 Obtaining apparent values of kinetic parameters (K_M and V_{max}) using the Eadie-Hofstee curve

Biphasic kinetics can be indicative of allostery (negative cooperativity) and/or a low-affinity second substrate site [8]. A more conventional way to depict biphasic kinetics is through the Eadie-Holstee derivative of the Michaelis-Menten equation:

$$v = - K_M (v/[S]) + V_{max} \qquad \text{(Equation 3.1)}$$

The Eadie-Hofstee plot is a way of plotting kinetic data in order to yield a straight line for enzymatic reactions that obey Michaelis-Menten kinetics (*eg.* wild-type GpdQ in Figure 3.3). This is achieved by plotting velocity (v)

on the y-axis versus velocity/substrate concentration (v/[S]) on the x-axis. The slope of the line is $-K_M$ while the y-intercept is V_{max}.

All three linear derivative plots of the Michaelis-Menten equation, *ie.* Lineweaver-Burk, Eadie-Hofstee and Hanes-Woolf, have their respective advantages and disadvantages. The Eadie-Hofstee plot is ideal in highlighting even slight deviations from the Michaelis-Menten kinetics [9, 10].

Using the same sets of data to create the Michaelis-Menten plots in Figure 3.2, the Eadie-Hofstee plots of b*p*NPP activities of the wild-type GpdQ, N80A and H156A are as shown below.

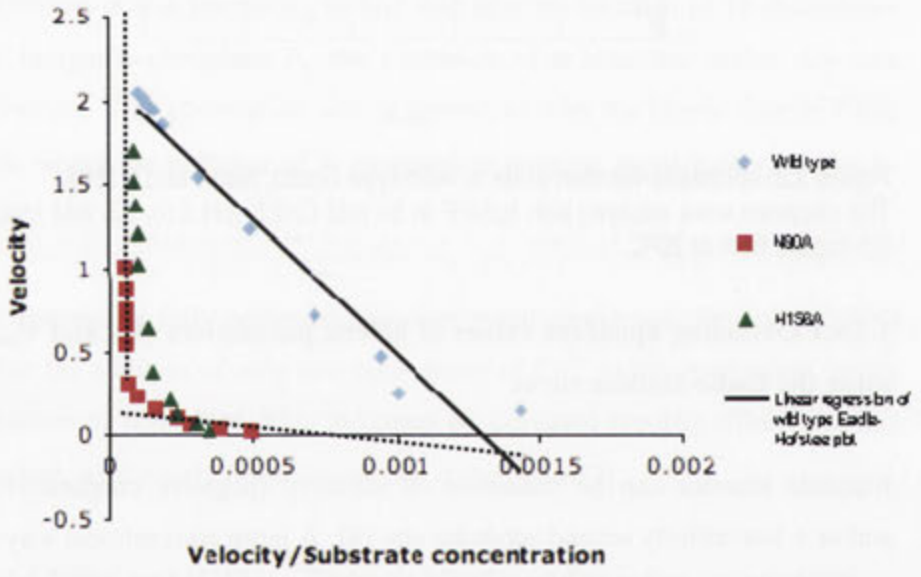


Figure 3.3: Eadie-Hofstee plots of wild-type GpdQ, N80A and H156A.

Apparent values of kinetic parameters of N80A and H156A were extrapolated from the two straight segments of the Eadie-Hofstee plots in Figure 3.3 (straight lines fit to four or five points from both segments, as indicated by dotted lines on the N80A curve). This method to parameterise k_{cat} and K_M in biphasic enzymes was adopted from literature [11, 12].

Two sets of values for k_{cat} and K_M were derived as shown in Table 3.3 for N80A and H156A, respectively (also previously shown in Table 3.2). Assay conditions were 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL BSA at 20°C. As these values were obtained using Microsoft Excel instead of KaleidaGraph, errors were not calculated.

Enzyme	Activity	App.* k_{cat} (s^{-1})	App. K_M (mM)	App. k_{cat}/K_M ($M^{-1}s^{-1}$)
N80A	High activity, low affinity (At high substrate concentrations)	2.28	38.62	5.90×10^1
	Low activity, high affinity (At low substrate concentrations)	0.17	0.39	4.26×10^2
H156A	High activity, low affinity (At high substrate concentrations)	1.51	34.28	4.41×10^1
	Low activity, high affinity (At low substrate concentrations)	0.19	1.62	1.16×10^2

App. – apparent

Table 3.3: Apparent values of k_{cat}, K_M and k_{cat}/K_M of N80A and H156A.

There have been many other kinetic models that analyse biphasic kinetics that rely on computer simulations to parameterise several other kinetic constants apart from the basic k_{cat} and K_m [8, 11, 13].

3.3.4.2 Assessing negative cooperativity in N80A and H156A using the Hill equation

Biphasic kinetics quite often is diagnostic of negative cooperativity, that is, in the case of homotropic allostery, the binding of the first ligand molecule makes it more difficult for the second one to bind [14].

The Michaelis-Menten equation describes the kinetics of a simple enzyme system that shows no cooperativity. However in reality many enzymes exhibit cooperativity and deviate from Michaelis-Menten kinetics. Cooperativity in proteins and enzymes can be defined as interactions between subunits or subsites where the binding of the first ligand molecule

increases (positive cooperativity) or decreases (negative cooperativity) the binding affinity of the other(s) [8]. An equation describing cooperative binding of ligands to macromolecules was proposed by Hill in 1910, based on his observation of the oxygen binding nature of haemoglobin [9, 15]:

$$v = \frac{V_{max}[S]^n}{K_d + [S]^n} \qquad \text{(Equation 3.2)}$$

where v is initial velocity; V_{max}, maximum velocity; n, Hill coefficient; [S], substrate concentration and K_d, dissociation constant. When $n = 1$, Equation 3.2 is exactly the same as the Michaelis-Menten equation and $K_d = K_M$.

A simplified, linearised Hill equation is given below:

$$\log [v/(V_{max} - v)] = n \log [S] + \log K_d \qquad \text{(Equation 3.3)}$$

Determining the Hill coefficient, n, is the easiest and most straightforward method of assessing the cooperativity of an enzyme. If $n = 1$, there is no cooperativity in the enzyme activity. $n \leq 1$ indicates negative cooperativity while $n \geq 1$ indicates positive cooperativity [9]. However, in practice, the Hill equation is more useful in showing positive cooperativity in enzymes. Proving negative cooperativity using the Hill equation can be a rather ambiguous process [16]. The Hill equation is more easily applied to ligand binding studies in receptors and cytochrome P450, where the binding capacity, which is equivalent to V_{max}, is known [17-19]. In the case of N80A biphasic kinetic studies, the Hill constants were calculated from the region of the maximum slope, by using the apparent V_{max} values in the high activity, low affinity component (Figure 3.4). This method in determining the Hill constant in negatively cooperative enzyme systems was proposed by Tipton [20, 21].

Figure 3.4 shows the Hill plots of wild-type GpdQ, N80A and H156A assayed with b*p*NPP in 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL

BSA at 20°C. Table 3.4 shows the values of the Hill coefficient (*n*) from plots in Figure 3.4.

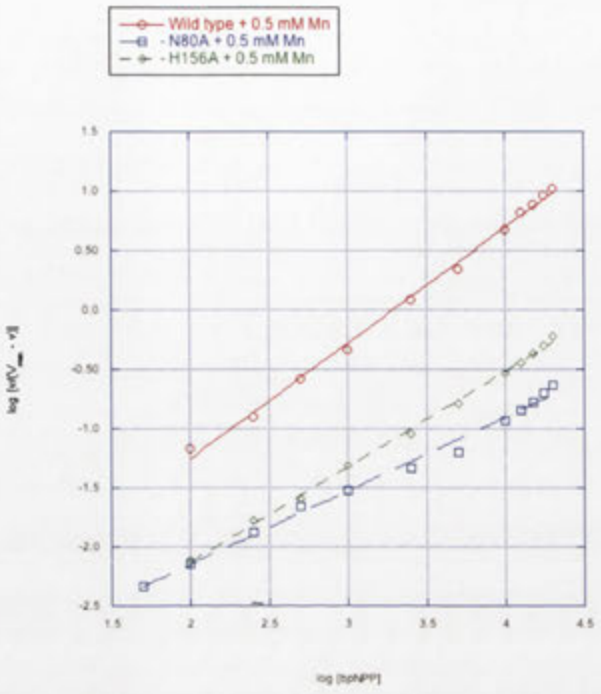


Figure 3.4: Hill plots of wild-type GpdQ N80A and H156A.
The enzymes were assayed with b*p*NPP in 50 mM CHES (pH 9.0), 0.5 mM $MnCl_2$, 0.1 mg/mL BSA at 20°C.

Enzyme	*n*	R^2
Wild-type GpdQ	1.0	1.00
N80A	0.6	0.99
H156A	0.6	1.00

Table 3.4: Values of the Hill constant (*n*) of wild-type GpdQ, N80A and H156A as derived from the plots in Figure 3.4 and their corresponding correlation coefficient (R^2) values.

The Hill constant values of N80A and H156A (both 0.6, *ie.* < 1.0) suggest that these two GpdQ enzymes are negatively cooperative.

A series of experiments was conducted to investigate the factors that might influence the negative cooperativity in N80A kinetics. The results and discussions of these experiments are found in Appendix F.

3.4 MUTATIONAL ANALYSIS OF α-METAL LIGANDS

3.4.1 D8 – D8N and D8A

D8 is a primary ligand to the α-metal ion and is responsible for the high metal binding affinity at the α-site. Mutants D8N and D8A were made to explore how the reduced metal binding affinities at the α-site would affect the enzyme activity. Results of kinetic assays conducted at conditions as described before are presented in Table 3.2.

By weakening the metal binding affinity at the α-site, both D8N and D8A mutations inactivated the enzyme. The k_{cat} values of D8N and D8A were 1000 and more 3000-times lower than that of the wild type, respectively, as the interaction of the residues with the α-metal became increasingly weak.

The D8 mutations did not have an adverse effect on the substrate affinity (K_M), thus strongly suggesting that the α-metal is involved in catalysis but not substrate binding or orientation. Another inference that could be made from the kinetic results in Table 3.11 is that a strong metal binding affinity at the α-site is essential for enzyme activity. The observed effects of D8A on GpdQ were similar to those of its equivalent mutant, D21A in Rv0805 [29].

3.4.2 H10 – H10A

H10 is one of the two histidine ligands to the α-metal; the other one being H197. H10 was replaced by an alanine and the effects of the mutation were investigated through b*p*NPP assays (Table 3.2).

The H10A mutation did not cause any great reduction in GpdQ activity, which is surprising because it seems unlikely that an alanine would be capable of coordinating the α-metal. H10A was an intrinsically active

enzyme. The slight decrease in k_{cat} and the minor increase in K_M are consistent with the observations found on H23A of Rv0805. Unlike the drastic effects the H156Q and H159Q/N/A mutations had on the β-metal site, H10A probably reduced the binding affinity at the α-metal site, but only slightly as it was compensated for by the assay conditions that contained an excess of Mn^{2+}.

3.5 MUTATIONAL ANALYSIS OF THE BRIDGING LIGAND, D50

The GpdQ active site has a μ-bridging ligand, D50, where one oxygen atom from the carboxylic group coordinates both the α- and β-metal. Together with the two metal ions, a water molecule that is coordinated equatorially, D50 is responsible for the octahedral coordination geometry of the binuclear metal center of GpdQ (Figures 1.13 and 3.1).

During purification, D50A was found to be mostly insoluble (Figure 3.5).

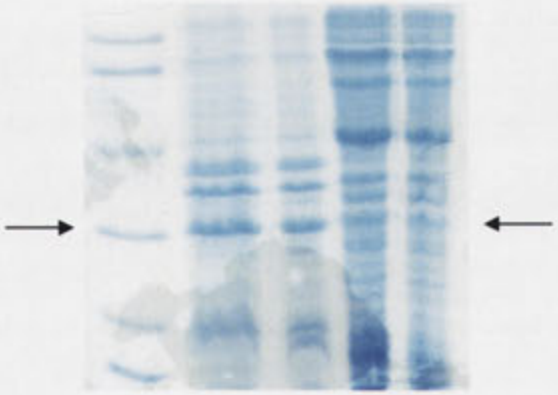


Figure 3.5: SDS-PAGE gel showing D50A after lysis using the French press.
Lane 1 shows the marker; Lanes 2 and 3 show duplicates of the insoluble fraction of D50A; Lanes 4 and 5 are duplicates of the soluble fraction. The arrows indicate the position of the D50A protein bands.

Kinetic assays of D50N and D50A were performed with b*p*NPP at a range of different concentrations at in 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL bovine serum albumin (BSA) at 20°C. Kinetic results are summarised in Table 3.2.

Weakened metal binding affinities in D50N and D50A mutants resulted in a progressive increase in K_M values. This again supports the hypothesis that the β-metal is associated with the substrate. With N50 and A50 having weaker binding affinities to the β-metal, the substrate affinities were also subsequently reduced.

With regard to the catalytic turnover, k_{cat}, D50N exhibited a 7-fold increase whereas D50A suffered from a dramatic decrease. If the substrate molecule is associated with the β-metal, the improved k_{cat} in D50N was likely due to the fact that the asparagine substitution facilitated the easy departure of β-metal and hydrolysis products from the active site. However, the polarity of the residue at this site had to be maintained for the enzyme to be catalytically active, as proven by the D50N mutant.

3.6 MUTAGENESIS OF H81, THE CONSERVED HISTIDINE

3.6.1 Introduction and background

The large and functionally diverse superfamily of calcineurin-like dimetallo-phosphoesterases contains a conserved histidine within the GNHD/E sequence motif. In GpdQ, this histidine is H81 [22].

The conserved histidine is always within 5 Å of the binuclear metal center although it does not directly coordinate either of the metal ions [23]. The backbone of this residue is stabilised via an extensive hydrogen-bonding network, including a hydrogen bond to the backbone carbonyl of the adjacent asparagine residue (N80A in GpdQ), which is a β-metal ligand [24] (Figure 3.6).

The nitrogen atoms of the imidazole group of H81 are hydrogen bonded to the backbone carboxyl of N53 and the amide group of N80 respectively. N53, on the other hand, interacts with H10, an α-metal ligand as well as

S268 of another chain (not shown in Figure 3.10). Additionally, the backbone carboxyl of H81 is hydrogen bonded to G55.

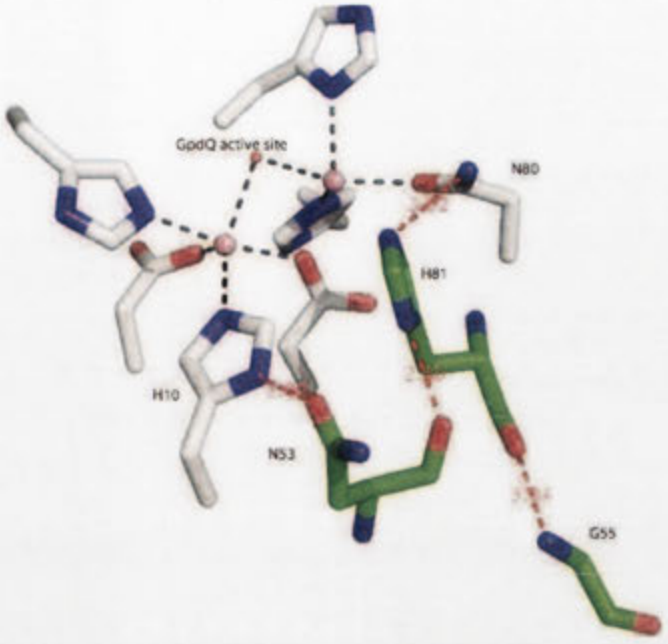


Figure 3.6: The extensive hydrogen-bonding network of H81.
The carbon chains of non-metal ligand residues are coloured in green while those belonging to metal ligand residues are in white.

Research groups working on different enzymes that contain the conserved histidine have extensively studied the residue. Proposed roles for this conserved histidine include the following:

i. Substrate binding and orientation.
ii. Activation of nucleophile that attacks the phosphodiester bond of substrates.
iii. General acid catalysis involving the protonation of the leaving group. The protonation of the leaving group of poor substrates in the transition state is vital for physiologically relevant substrates that typically have high pK_a values [23].
iv. General base catalysis by deprotonation of an iron-coordinated solvent molecule [24].

Results from mutagenesis studies of the conserved histidine have not been consistent and conclusive enough to conclusively distinguish the correct hypothesis. Table 3.5 summarises the kinetic findings on the conserved

histidine residues of related metallophosphatases over the last two decades by four research groups.

Research Group	Enzyme and mutants	k_{cat}	K_M	Conclusions
Zhuo *et al.* [25]	λ protein phosphatase H76N	10^5 lower	Not affected in the absence of Mn^{2+} but 40 times higher when the metal was added	H76 contributes to metal binding and catalysis.
Funhoff *et al.* [24]	Purple acid phosphatase H92N	3 times higher at the optimal pH value	5 times higher	H92 is involved in positioning either the nucleophile or the substrate, rather than directly in acid or base catalysis.
	H92A	10 times lower at the optimal pH value	9 times higher	
Mertz *et al.* [23]	Calcineurin H151Q			H151 and H76 are not required for protonation of the leaving group. Loss of activity is attributed to the perturbation of the ligand environment, possibly by disruption of a hydrogen bond between the histidine and a metal-coordinated solvent.
	*p*NPP hydrolysis	460-fold lower	10 times lower	
	[P]-R_{II} hydrolysis	1300-fold lower	Similar	
	λ protein phosphatase H76N			
	*p*NPP hydrolysis	590-fold lower	Similar	
	Phenyl phosphate hydrolysis	530-fold lower	Similar	
Hoff *et al.* [26]	λ protein phosphatase H76N	Lower, difference is more significant in higher pH values.	Higher at low pH values but not affected at high pH values.	Substrate binding is assisted when H76 is protonated.

Table 3.5: Kinetic findings from mutagenesis studies of the conserved histidine residues in various structurally related dimetallophosphatases.

3.6.2 Results

H81Q and H81A mutants of GpdQ were made. The purified enzymes were assayed at pH 9.0 and pH 6.0. Assay results are summarised in Tables 3.6 and 3.7.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_m	k_{cat}/K_M ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
H81Q	0.92 ± 0.01	0.12	0.96 ± 0.06	0.46	9.60×10^2	0.27
H81A	4.83 ± 0.14	0.64	3.39 ± 0.44	1.61	2.17×10^2	0.06

Table 3.6: Results of kinetic assays of H81 mutants at pH 9.0.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_m	k_{cat}/K_M ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	0.08 ± 0.003	1.00	0.51 ± 0.11	1.00	1.53×10^2	1.00
H81Q	0.03 ± 0.0004	0.38	0.64 ± 0.07	1.25	4.84×10^1	0.32
H81A	0.03 ± 0.01	0.38	2.31 ± 0.41	4.53	1.45×10^1	0.09

Table 3.7: Results of kinetic assays of H81 mutants at pH 6.0.

3.6.3 Discussion

The uniform reduction in catalytic activity (k_{cat}) in both H81Q and H81A mutants, at both pH 9.0 and pH 6.0, strongly demonstrated that this histidine residue might play a role in the catalytic mechanism of GpdQ.

At high pH values (pH 9.0), H81 was unprotonated, hence making protonation of any leaving group impossible. H81 was therefore not involved in general acid catalysis at alkaline pH.

However, at pH 9.0, an unprotonated H81 could possibly participate in general basic catalysis and activate the nucleophile. This proposition

generates support from the poor catalytic activities of both H81Q and H81A.

H81 mutations had a similar effect on GpdQ's phosphodiesterase activity at pH 6.0, although, like λ protein phosphatase, the decrease was more profound at higher pH values. At pH 6.0, H81 was protonated and so its role in protonating the leaving group was possible. As the pK_a of the *p*-nitrophenolate leaving group is 9.6, it seems unlikely that the hydrolysis of b*p*NPP would require protonation.

An alternative explanation for the loss of activities in the mutants was that in the absence of H81, the active site environment had been disrupted. This is a reasonable argument considering how H81 is involved in an extensive hydrogen bonding network, including an interaction with N80.

At pH 9.0, H81Q exhibited an enhanced affinity for b*p*npp while H81A experienced otherwise. At pH 6.0, the K_M of H81Q was similar to the wild-type while H81A's increased more significantly than at pH 9.0. Comparing the two K_M values of the wild-type enzyme at pH 9.0 and pH 6.0, it could be inferred that substrate binding was assisted when H81 was protonated (low K_M at pH 6.0). This may explain why the K_M value of H81Q was not affected at pH 6.0, as a glutamine was capable of interacting with the substrate due to the polarity of its side chain.

The improved K_M of H81Q has parallels. A H151Q mutant of calcineurin experienced a 10-fold drop in K_M when assayed with *p*npp. Mutation of histidine to a glutamine resulted in an increase in the magnitude of the negative charge. This may present a more conducive environment at the active site for the formation of a complete binuclear metal center [23].

3.6.4 Conclusions on H81

Extensive mutational analysis has been performed on the conserved histidine in the GNHD/E sequence motif of the α/β sandwich superfamily of dimetallo-phosphohydrolases. Experimental results from these mutagenesis studies are conflicting (Table 3.5). Therefore though this histidine residue is conserved across all the members of this family of enzymes, it may play different roles in different enzymes.

Based on the kinetic assays done in this study, the following conclusions are made on GpdQ's H81:

i. H81 has a catalytic role in GpdQ:
   - Whether or not the protonated H81 is involved in general acid catalysis cannot be ascertained with substrate b*p*npp as it has a good leaving group.
   - The unprotonated H81 could possibly act as a general base to activate the nucleophile.
   - Loss of activity in both H81Q and H81A is attributed to the perturbation of the binuclear metal center environment, possibly by disruption of a hydrogen bond between H81 with the carbonyl oxygen of N80 backbone. Thus, H81 also has a structural function in stabilising the folded structure of the active site.

ii. H81 contributes to substrate binding, especially when it is protonated at low pH values.

iii. H81 may also assist in metal binding.

3.7 CONCLUSIONS

The catalytic mechanism of GpdQ remains unclear. The catalytic mechanisms of other structurally related enzymes such as purple acid phosphatases, Ser/Thr protein phosphatases, 3',5'-exonucleases and 5'-

nucleotidases have been more or less established [27]. The metal ligands of all these enzymes, including GpdQ, are conserved with only slight variations of one or two residues. However, though structurally similar, these enzymes catalyse the hydrolysis of different substrates that are mainly diesters and monoesters of phosphoric acid, including ribo- and deoxyribonucleoside 5'-monophosphates [28].

At the end of this study, two questions surrounding the catalytic mechanism of GpdQ remain unaddressed: how the catalytic water molecule or the nucleophile is activated and whether the substrate is bound in a monodentate or bidentate mode. The biggest success to have come out from the mechanistic studies in this research project is the characterisation of the binuclear metal center with primary focus on the two non-identical metal binding sites.

Non-identical binuclear active site and implications on the catalytic mechanism

Anomalous scattering analysis [1] and magnetic circular dichroism (MCD) experiments [7] have both unequivocally confirmed the difference in metal binding affinities between the two α- and β-sites.

The high metal binding affinity at the α-site is largely attributed to the acid ligand, D8. Decreasing and abolishing altogether the ability of this residue to coordinate the α-metal via mutants D8N and D8A respectively caused significant drops in k_{cat}. The mutations did not affect the K_m adversely. The inference drawn from these kinetic results is that the α-metal plays a pivotal role in catalysis.

The ligand responsible for the low metal binding affinity at the β-site is N80. Enhancing the binding affinity by converting the asparagine to its acid resulted in an inactive GpdQ mutant enzyme. N80A, on the other hand was

catalytically active. This demonstrated the importance of flexibility in metal binding at the β-site.

Results from the kinetic assays of N80D and N80A mutants suggested that the β-metal is associated with the substrate molecule. Initially, especially with the improved K_m in N80D, it was tempting to believe that the β-metal was responsible for substrate binding. However, EPR and stopped-flow fluorescence experiments later showed otherwise, that the binding of the β-metal was promoted by the substrate [29].

Biphasic kinetics and negative cooperativity

The dual phenomena of biphasic kinetics and negative cooperativity were observed in mutants N80A and H156A. Negative cooperativity makes a protein or enzyme less sensitive in response to the changes in ligand or substrate concentrations but enables the protein to work at a broader range of concentrations [14]. The nature of the biphasic kinetics/negative cooperativity in N80A was investigated further in Appendix F to establish factors that influenced the kinetic behaviour.

The non-Michaelis-Menten kinetics was prevalent only when then enzyme was assayed with b*p*NPP. It also appeared to be a function of metal binding, as the biphasic behaviour was not observed when N80A was assayed in the absence of exogenously added metal ions. The concentration and the type of metal ions as well as the oligomeric state of the enzyme did not affect greatly the extent of negative cooperativity in N80A, as indicated by the values of the Hill constant in Tables F.1 and F.3 in Appendix F.

Mutational analysis of histidine ligands

Mutating the histidine residues that coordinate the α- and β-metals, namely H10, H156 and H195, all led to reductions in enzyme activity, thus underlining the importance of histidines as ligands in metalloenzymes.

Activation of the nucleophile

As two forms of phosphodiesterase activity have been characterised with pH activity analyses [9, 33], it is highly possible for GpdQ to employ two different catalytic mechanisms at alkali and acidic pH values. The role of β-metal in substrate binding probably remains the same, but the nucleophilic activation, catalysis and substrate coordination roles played by the α-metal and H81 are flexible, subject to factors like pH and the pK_a of the leaving group of hydrolysis products.

CHAPTER THREE REFERENCES

1. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.
2. Shenoy, A. R., Sreenath, N., Podobnik, M., Kovacevic, M., and Visweswariah, S. S. (2005) The Rv0805 gene from *Mycobacterium tuberculosis* encodes a 3 ',5 '-cyclic nucleotide phosphodiesterase: Biochemical and mutational analysis. *Biochemistry 44*, 15695-15704.
3. Bertino, I., Drago, R. S., Luchinat, C. (1983) The coordination chemistry of metalloenzymes: The role of metals in reactions involving water, dihydrogen, and related species. *Proceedings of the NATO Advanced Study Institute.*
4. Kuo, J. M., and Raushel, F. M. (1994) Identification of the Histidine Ligands to the Binuclear Metal Center of Phosphotriesterase by Site-Directed Mutagenesis. *Biochemistry 33*, 4265-4272.
5. Sundberg, R. J., and Martin, R. B. (1974) Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. *Chemical Reviews 74*, 471-517.
6. Huang, H. B., Horiuchi, A., Goldberg, J., Greengard, P., and Nairn, A. C. (1997) Site-directed mutagenesis of amino acid residues of protein phosphatase 1 involved in catalysis and inhibitor binding. *Proceedings of the National Academy of Sciences of the United States of America 94*, 3530-3535.
7. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-Promoted Formation of a Catalytically Competent Binuclear Center and Regulation of Reactivity in a Glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of the American Chemical Society 130*, 14129-14138.
8. Atkins, W. M. (2005) Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annual Review of Pharmacology and Toxicology 45*, 291-310.
9. Cornish-Bowden, A. (2004) Fundamental of enzyme kinetics, 3rd edition. *Portland Press.*
10. Kennedy, L. M. (1989) Gymnemic acids - specificity and competitive inhibition. *Chemical Senses 14*, 853-858.
11. Martin-Nieto, J., Flores, E., and Herrero, A. (1992) Biphasic kinetic-behavior of nitrate reductase from Heterocystous, nitrogen-fixing Cyanobacteria. *Plant Physiology 100*, 157-163.
12. Junior, M. B., Siqueira, C. G., Antoniazi, S. A., and Ueta, J. (1996) Location of the beta-galactosidase of the yeast *Kluyveromyces marxianus* ATCC 10022. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 69*, 357-361.

13. Kurganov, B. I. (2000) Analysis of negative cooperativity for glutamate dehydrogenase. *Biophysical Chemistry 87*, 185-199.
14. Koshland, D. E. (1996) The structural basis of negative cooperativity: Receptors and enzymes. *Current Opinion in Structural Biology 6*, 757-761.
15. Hill, A. V. (1910) The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *Journal of Physiology 40*, 1-7.
16. Goodrich, J. A., and Kugel, J. F. (2007) Binding and kinetics for molecular biologists. *Cold Spring Harbor Laboratory Press.*
17. Sakai, S. (1994) Negative cooperativity in the prolactin-receptor interaction in the rabbit mammary gland - action of high prolactin concentration. *Journal of Dairy Science 77*, 433-438.
18. Henis, Y. I., and Levitzki, A. (1980) Mechanism of negative cooperativity in glyceraldehyde-3-phosphate dehydrogenase deduced from ligand competition experiments. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 77*, 5055-5059.
19. Harlow, G. R., and Halpert, J. R. (1998) Analysis of human cytochrome P450 3A4 cooperativity: Construction and characterization of a site-directed mutant that displays hyperbolic steroid hydroxylation kinetics. *Proceedings of the National Academy of Sciences of the United States of America 95*, 6636-6641.
20. Tipton, K. F. (2001) Principles of enzyme assays and kinetic studies in Enzyme assays, 2nd edition (revised). *Oxford University Press Chapter 1,* 1-47.
21. Tipton, K. F. (2002) Enzyme activities in Essential Cell Biology Volume 2: Cell Function A Practical Approach. *Oxford University Press Chapter 1,* 1-46.
22. Jackson, C. J. (2006) Enzymatic catalysis of phosphate ester hydrolysis. *The Australian National University.*
23. Mertz, P., Yu, L., Sikkink, R., and Rusnak, F. (1997) Kinetic and spectroscopic analyses of mutants of a conserved histidine in the metallophosphatases calcineurin and lambda protein phosphatase. *Journal of Biological Chemistry 272*, 30589-30589.
24. Funhoff, E. G., Wang, W. L., Andersson, G., and Averill, B. A. (2005) Substrate positioning by His92 is important in catalysis by purple acid phosphatase. *Febs Journal 272*, 2968-2977.
25. Zhuo, S. Q., Clemens, J. C., Stone, R. L., and Dixon, J. E. (1994) Mutational analysis of a Ser/Thr phosphatase - identification of residues important in phosphoesterase substrate binding and catalysis. *Journal of Biological Chemistry 269*, 26234-26238.
26. Hoff, R. H., Mertz, P., Rusnak, F., and Hengge, A. C. (1999) The transition state of the phosphoryl-transfer reaction catalyzed by the lambda Ser/Thr protein phosphatase. *Journal of the American Chemical Society 121*, 6382-6390.

27. Mitic, N., Smith, S. J., Neves, A., Guddat, L. W., Gahan, L. R., and Schenk, G. (2006) The catalytic mechanisms of binuclear metallohydrolases. *Chemical Reviews 106*, 3338-3363.
28. Hunsucker, S. A., Mitchell, B. S., and Spychala, J. (2005) The 5 '-nucleotidases as regulators of nucleotide and drug metabolism. *Pharmacology & Therapeutics 107*, 1-30.
29. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural flexibility enhances the reactivity of the bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.

4

DIRECTED EVOLUTION OF GpdQ – SCREENING METHODS AND PART ONE

CHAPTER FOUR: DIRECTED EVOLUTION OF GpdQ – SCREENING METHODS AND PART ONE

4.1 INTRODUCTION

Organophosphate triesters that are commonly used as pesticides and chemical warfare agents pose serious risks upon the environment and human health. Much research has been dedicated to the development of efficient and safe strategies to detoxify organophosphate triesters that have been dispersed into the ecosystem.

One of the objectives of this project, initially, was to further develop Dr Yu-McLoughlin's growth-based assay as a preliminary screening method for the directed evolution of GpdQ with methyl paraoxon, ethyl paraoxon, dimethyl phosphate and diethyl phosphate. Improved activity with any of these substrates would potentially lead to bioremediation applications [1-3].

The growth assay and top agar overlay agar plate-based methods of *in vivo* screening of GpdQ mutant libraries were pursued for 2-3 months with little success. Details of how and why these experiments failed are reported in section 4.2. The objective was then changed to evolving GpdQ's bis (*p*-nitrophenyl) phosphate (b*p*NPP) hydrolytic activity. An *in vitro* method was employed to screen the libraries for improved variants. At the end of this project, eight full rounds of directed evolution of GpdQ with b*p*NPP were completed.

4.2 *IN VIVO* AGAR PLATE-BASED SCREENING OF GpdQ MUTANT LIBRARIES

One very interesting feature of GpdQ is the unusually broad substrate specificity of the enzyme. The hydrolase has primarily diesterase activity, but since its first discovery, purification and characterisation more than

three decades ago, there have been reports of weak activities with monoesters of phosphoric acid [4-6]. Since the revival of research interest in the enzyme in recent years, GpdQ has been characterised more thoroughly. It was found that this highly promiscuous phosphohydrolase extends its specificity spectrum wide enough to even include organophosphate triesters [7, 8]. Table 4.1 shows the kinetic parameters for the hydrolysis of organophosphate triesters, diesters and monoester used in this study, as published by Ghanem and colleagues [8]. The last column is the values of k_{cat}/K_m relative to that of the best substrate, b*p*NPP.

	Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M ($M^{-1}s^{-1}$)
DIESTER	bis (*p*-nitrophenyl) phosphate	6.0 ± 0.2	0.9 ± 0.1	(6.7 ± 0.5) x 10^3	1
	Dimethyl phosphate	(5.0 ± 0.1) x 10^{-3}	0.06 ± 0.01	82 ± 14	1.2 x 10^{-2}
	Diethyl phosphate	(7.4 ± 0.5) x 10^{-3}	0.5 ± 0.06	15 ± 1	2.2 x 10^{-3}
TRIESTER	Methyl paraoxon	nd	nd	13 ± 1	1.9 x 10^{-3}
	Ethyl paraoxon	nd	nd	< 1	Nd
MONOESTER	*p*-nitrophenol phosphate	nd	nd	5 ± 1	7.5 x 10^{-4}

nd – not determined

Table 4.1: Kinetic parameters of GpdQ-catalysed hydrolysis of various organophosphate compounds used in this project, as published by Ghanem *et al.* [8].

4.2.1 McLoughlin's growth assay method [7]

Using this method, GpdQ mutant libraries were screened for heightened DMP or DEP hydrolysis activity. Experimental procedures are described in Chapter 2 (2.4.2.1). The idea behind this strategy was that, the more

efficiently a GpdQ mutant degraded the alkyl diester compounds, the bigger the size of its *E. coli* DH5α host colony, as it had a more abundant supply of inorganic phosphate (Figure 1.7) to stimulate growth. Other *E. coli* colonies bearing inferior GpdQ mutants would either grow very slowly or they might not survive at all due to the lack of a direct phosphorus source. In other words, the substrate for GpdQ was a source of nutrient for bacterial growth.

After two days of incubation at 37°C, no cell growth was observed on the modified MOPS minimal medium that was plated with GpdQ mutants. A set of control experiments was designed to investigate reasons behind this negative observation. The *E. coli* DH10β cells carrying the recombinant *E. aerogenes ugp* operon were prepared by Dr Sean Yu-McLoughlin in 2003 and stored at -80°C.

Strain	Phosphorus source (all at the concentration of 1 mM)			
	None	DMP	DEP	K_2HPO_4
DH5α (GpdQ$^+$)	-	-	-	++
DH10β (*ugp*$^+$)	-	+	+	++

- denotes no growth, + colonies of < 1 mM in diameter after 2 days, ++ colonies of > 1 mM after 2 days

Table 4.2: Results of the growth assays of *E. coli* strains phosphate-free MOPS minimal medium each supplemented with a different source of inorganic phosphate.

Discussion

Ghanem and colleagues successfully grew *E. coli* expressing GpdQ in liquid MOPS phosphate-free medium supplemented various organophosphate diesters and methyl phosphonate monoesters as an only source of inorganic phosphate. Based on their experiments, the agar-plate based growth assays in this study were expected to work.

There are three main reasons which may explain why in this study, DH5α *E. coli* that expressed recombinant GpdQ failed to grow on DMP or DEP as their sole phosphorus source:

1. *The complete putative ugp operon was vital for cell survival and growth*

The results summarised in Table 4.2 illustrate the inability of DH5α *E. coli* to utilise simple alkyl diesters like DMP and DEP as the sole source of phosphorus. On the other hand, these diesters supported the growth of DH10β cells, thus highlighting the importance of having the complete, functional putative *E. aerogenes ugp* operon for the survival of cells. As reported in section 1.3.2 in Chapter 1, the operon contains four other open reading frames that would presumably express permeases and transporter proteins, thus facilitating the uptake of diesters into the cytoplasm. McLoughlin *et al.* had previously shown that native DH10β cells were unable to survive on DMP plates, which suggests that *E. coli* does not express a phosphodiesterase capable of hydrolysing DMP under phosphate-limiting conditions [7]. Finally, transformed DH5α and DH10β cells grew equally well on plates fortified with a phosphate salt as a direct source of phosphate. This implied there was no apparent difference in growth rate between the two *E. coli* strains.

2. *Difference in media used (solid/liquid forms)*

Lower recoveries of stress-induced cells in agar media compared to liquid media have been reported in literature [9]. Reasons for poor recoveries include impurities found in the agar, cell density effect and limited diffusion of nutrients into the cells [10-13]. This might explain why in this study, *E. coli* cells expressing GpdQ could not survive on solid minimal media supplied with DMP or DEP as the sole phosphorus source. Ghanem *et al.,* using liquid media, obtained bacterial growth with similar experiments [8].

3. *Different properties of E. coli strains (protein expression level and growth rate)*

The BL21 (DE3) cells used by Ghanem *et al.* had a recombinant protein expression system that was superior to their DH5α counterpart, due to

the presence of the T7 promoter. BL21 (DE3) cells also exhibit a higher growth rate.

McLoughlin's growth assay is a high-throughput method for screening GpdQ mutant libraries. Especially for the first round of screening, a quick and least laborious process that enables a library of 10^6 variants to be screened is ideal. This elegant screening method associates GpdQ's phosphodiesterase activity with a phenotype that is easily recognisable - the size of the *E. coli* host colony. It is also extremely useful when the enzyme is evolved for higher catalytic activity with non-chromogenic substrates like DMP and DEP.

However, the workability of this method was complicated by the fact that GpdQ was originally part of an operon functioning unit that encodes other proteins to regulate the system. Discovery of improved mutants may be compromised when the recombinant *gpd*Q gene has been segregated from its sister genes in the putative *E. aerogenes ugp* operon, and the support of their encoded proteins.

The cells were not grown on the screening minimal medium for more than 48 hours as long incubation periods (at 37°C) might deplete the ampicillin selectivity of the medium and cause an increased background of satellite colonies.

4.2.2 Cho's top agar method

This screening technique, applicable only to chromogenic substrates, was tried with paraoxon only [14]. In our experiment, MPO was used as GpdQ reportedly displayed a preference for methyl-substituted triesters [8]. A mutant's potency in degrading MPO was associated with the intensity the yellow halo (formation rate of *p*-nitrophenol hydrolysis product) surrounding its colony. There was no apparent formation of yellow colour

around the cell colonies after they had been subject to 5 hours of incubation at 37°C with the MPO-containing agarose top solid layer.

Another positive control experiment was set up where DH5α cells were transformed with recombinant *opd*A (courtesy of Dr Jee-Loon Foo) which expresses the eponymous phosphotriesterase. Cells that grew overnight on M9 minimal medium with ampicillin selection were then layered with the agarose solution containing 1 mM MPO. Faint yellow halos formed around the cell colonies after 1 hour of 37°C incubation.

The positive control experiments suggested that GpdQ's poor phosphotriesterase activity was the reason behind the failure of the MPO screen. The success of the screen was further inhibited by the cell membrane barrier which limited the diffusion of organophosphate esters into the cytoplasm.

Discussion

The top agar overlay facilitated screening faced a major dilemma: When GpdQ was evolved with a slow substrate, it was preferable to have the enzyme expressed on the surface of *E. coli*, like Cho and co-workers did while evolving the activity of organophosphate hydrolase with methyl parathion [14]. Otherwise, the poor catalytic activity of the enzyme and limited accessibility of the subtrate into the cell would both hinder the success of the screen. A longer waiting time, for example overnight, was not recommended due to the high level of toxicity of organophosphate triesters.

Conclusions on in vivo agar plate-based screening

The biggest challenge faced by most directed evolution projects is the availability of a suitable, high-throughput screening/selection system for the enzyme of interest.

Once established and optimised successfully, agar plate-based screens are cost and time effective. They also require minimal and simple equipment to operate. Above all, they enable huge libraries to be evaluated in each round of evolution [15]. However, both the methods tried in this study inflict cellular stress (phosphate starvation, incubation with toxic compounds) that might have sabotaged *E. coli* survival.

4.3 *IN VITRO* RANDOM SCREENING OF GpdQ MUTANT LIBRARIES

bis (*p*-nitrophenyl) phosphate (b*p*NPP) is a model phosphodiester substrate extensively used for the characterisation of phosphodiesterase activities in enzymes like nucleases, cyclic nucleotide phosphodiesterases (PDEs) and glycerophosphodiesterases (GPDs) as well as micelles [16-18]. The popularity of the substrate is attributed to its relatively convenient assay where the release of *p*-nitrophenol from catalysis can be monitored directly via spectrophotometry. The phosphodiester bond in b*p*NPP can also be broken down easily as the *p*-nitrophenol leaving group has a low pK_a value.

b*p*NPP is the fastest among the characterised non-physiological substrates of GpdQ [8]. During the earlier rounds (rounds 1-4), GpdQ was evolved towards two different directions, simultaneously but separately with b*p*NPP: higher substrate affinity (low K_m) and higher catalytic activity (high k_{cat}). Results and discussion will be presented in 4.5. Directed evolution from round 5 onwards, where different strategies had been applied will be discussed in section 4.6.

All techniques used in the *in vitro* random screening of GpdQ genetic libraries are explained in Chapter 2 (section 2.4.2.2).

In random screening, each variant of the library is subject to examination for a particular desired characteristic, which in the case of this project, is

enhanced activity with b*p*NPP. Individual clones are picked from an agar plate, either manually or robotically, to inoculate growth medium in a 96-well culture plate. Clearly, inoculation is a tedious process that drastically reduces the size of the libraries to be screened [15]. Therefore, the library to be screened was diluted with the growth medium instead before it was dispensed into the wells at a mean nascent concentration of ~2 cells per well. The size of the screened libraries in each round was maintained at approximately 10,000.

This 'diluted library' random preliminary screening method offered a direct and less ambiguous phenotype observation. The *in vivo* agar plate-based assay screening techniques described in 4.3 depend solely on visual judgements of colony size or formation of yellow colour around a colony to determine the catalytic rate of a variant, whereas in the random screening described hereby, catalytic activities were measured directly using UV spectrophotometry.

However, just like the *in vivo* screenings, there were several confounding factors including: cell density, protein expression level and solubility. A very good mutant's improvement might be hidden by its low cell density or poor protein expression.

4.4 DIRECTED EVOLUTION OF GpdQ (ROUNDS 1-4)

4.4.1 High affinity evolution

Under the assay conditions of 50 mM CHES pH 9.0, 0.1 mM $MnCl_2$, the native GpdQ was reported to have a K_m of 0.9 mM [19]. In this study, under the similar assay conditions, except where the concentration of $MnCl_2$ was 5 times higher at 0.5 mM, the K_m value was about 2.0 mM. To evolve GpdQ towards higher substrate affinity, the libraries were subject to kinetic assay screenings with decreasing amounts of b*p*NPP (1.0, 0.5 and 0.25 mM) each

successive round in the first three rounds, respectively to increase selection pressure. After each round of screening, the 10 mutants displaying the highest b*p*NPP activities (at substrate concentrations that were below K_m) were selected for StEP recombination to generate the library for the next round of evolution. After the third round of evolution, selected mutants were sequenced. Table 4.3 summarises the 20 best mutants obtained after three rounds of high affinity evolution and their genotypes.

Genotype	Number of mutants with same genotype	Percentage of mutants with same genotype (%)
C54G	3	15
C54G, H217R	1	5
C54G, Y221H	2	10
C54G, G259R	4	20
C54G, H217R, G259R	3	15
C54G, Y221H, G259R	7	35

Table 4.3: The 20 fittest mutants showing highest activities when assayed with 0.25 mM b*p*NPP during the third round of directed evolution.

Based on the sequencing results of the 20 best mutants listed in Table 4.3, the high affinity evolution of GpdQ appeared to have reached convergence with four repetitive mutations (C54G, H217R, Y221H and G259R) after only three rounds of genetic recombination. Convergence could be due to: (i) pressure from natural selection; (ii) fixation by random genetic or allele drift. (iii) It could be simply due to limitations of the PCR protocols (epPCR and StEP-A) used to generate the libraries, or (iv) a combination of these events. It was noteworthy that two of the iterative mutations, H217R and Y221H, did not occur together in any of the 20 selected variants. These mutations are close in sequence and their exclusiveness may arise from limitations of the StEP-A protocol.

Six GpdQ enzymes, representing each of the six different genotypes in Table 4.4, were expressed, purified to near homogeneity and characterised for b*p*NPP hydrolysis activity. The purification method used for these six

enzymes and all other enzymes characterised in this chapter was Protocol B, as described in section 2.6.2 of Chapter 2. Assay conditions were 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL bovine serum albumin (BSA) and at 20°C (section 2.7, Chapter 2).

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
3-4 (C54G)	13.74 ± 0.2	1.83	0.43 ± 0.04	0.20	3.21×10^4	8.99
3-3 (C54G, H217R)	15.70 ± 1.8	2.09	0.23 ± 0.02	0.11	6.78×10^4	18.99
3-16 (C54G, Y221H)	7.01 ± 0.1	0.93	0.29 ± 0.02	0.14	2.42×10^4	6.78
3-5 (C54G, G259R)	11.82 ± 1.9	1.58	0.29 ± 0.03	0.14	4.06×10^4	11.37
3-10 (C54G, H217R, G259R)	9.46 ± 0.2	1.26	0.37 ± 0.04	0.18	2.52×10^4	7.06
3-25* (C54G, Y221H, G259R)	91.45 ± 1.1	12.19	0.21 ± 0.02	0.10	4.26×10^5	119.33

* Hexamer-dimer equilibrium

Table 4.4: Kinetic parameters of the best variants acquired after the first three rounds of high affinity evolution of GpdQ with substrate b*p*NPP.
These six variants were essentially different combinations of the four repetitive mutations – C54G, H217R, Y221H and G259R.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
C54G	13.74 ± 0.2	1.83	0.43 ± 0.04	0.20	3.21×10^4	8.99
H217R	9.68 ± 0.2	1.29	0.71 ± 0.07	0.34	1.36×10^4	3.81
Y221H	8.20 ± 0.1	1.09	0.43 ± 0.04	0.20	1.89×10^4	5.29
G259R	11.25 ± 0.2	1.50	0.66 ± 0.07	0.31	1.70×10^4	4.76

Table 4.5: Kinetic parameters of single site mutants C54G, H217R, Y221H and G259R.
These mutants were generated via site-directed mutagenesis to examine the effects of the individual mutations on GpdQ phosphodiesterase activity.

The kinetic data in Table 4.5 emphasises the synergistic effects of the mutations or the importance of recombination as the individual mutations did not change the catalytic activity (k_{cat}) significantly. The K_m values saw a

3- to 5-fold improvement with the single mutations functioning by themselves.

Triple mutant 3-25 emerged as the fittest variant after three rounds of evolution. The dominance of this genotype as depicted in Table 4.6 (constituting 7 out of the 20 best variants) and again in Table 4.7 (8 out of 20 fourth round variants) was already a sign of its superiority. It was found that the catalytic turnover of 3-25 was about 12 times higher than the wild-type GpdQ, although it had been evolved towards higher substrate affinity, not higher catalytic activity. The K_m of mutant 3-25 was 10 times lower than that of the wild type.

Due to the early convergence of the mutations, high affinity evolution of GpdQ was not continued for a subsequent fourth round.

4.4.2 High activity evolution

For the initial four rounds of high activity evolution, GpdQ libraries were screened with 5.0 mM b*p*NPP in 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$, 0.1 mg/mL bovine serum albumin (BSA) and at 20°C. For these experiments, mutants chosen from each round of evolution were sequenced. Up to 10 variants were selected for StEP recombination in the next round of evolution. The list of the sequenced mutants from the first four rounds of high activity evolution is summarised in Tables 4.6 and 4.7.

Evolution of GpdQ towards enhanced b*p*NPP catalytic activity also assumed a convergent nature. The number of mutations that led to amino acid changes decreased progressively from 36 to 17 (first round → second round) to 11 (second round → third round) and finally to 4 (third round → fourth round). Single-site mutants were reproduced via site-directed mutagenesis to explore to what extent the individual mutations improved the catalytic activity of GpdQ (Tables 4.8 and 4.11). Kinetic analyses of these

mutants revealed that many of mutations might in fact be artefactual as a result of spurious observations during the screening process. Most of these single-site mutants had either minimal activity improvement or none at all. However, all these mutations did nonetheless contribute towards lowering the K_m. The reason for this false positive scenario was that the 5.0 mM concentration of b*p*NPP was too high for an accurate screening for improved catalysis.

Convergence drove the number of beneficial mutations to narrow down further to just 4 at the end of the fourth round of evolution. These four mutations, interestingly, were the very same mutations generated after three rounds of high affinity evolution. Once again, with high affinity evolution as the precedence, the dominance of triple mutant C54G-Y221-G259R (from now on referred to as 3-25) was manifest, with 8 out of the 20 best fourth round mutants bearing this genotype (mutant 4-5 in Table 4.7).

It should also be mentioned that among the selected third round and fourth round mutants lay a recurring silent mutation of residue Y250 (TAC $\rightarrow$ TAT).

	1 -1	1 -2	1 -3	1 -4	1 -5	1 -6	1 -7	1 -8	1 -9	1- 10	1- 11	1- 12	1- 13	1- 14	1- 15	1- 16
R12C			√													
R12L							√									
R12P															√	
F21L							√									
A26P				√												
D30G						√										
C54G		√								√						
C54Y													√	√		
N53D						√										
R64P		√														
V62G				√												
I77V						√										
G79L						√										
P96L		√							√							
A113T								√								
S114T		√														
R115L								√								
D120Y										√						
S122C	√															
S126T											√	√				
S127L			√													
A141T					√											
N164D			√													
Q166H				√												
Q206P																√
H217R																√
Y221H		√														
Y221N									√							
Y229stop	√														√	
S233A														√		
S253L	√															
G259R													√			
Y263C														√		
N266T					√											
I267T												√				
S268G											√					

Table 4.6: The genotypes of the 16 mutants selected from the first round of high activity evolution.

Mutation	R12C	F21L	C54G	C54Y	V62G	S122C	N164D	H217R	Y221H	Y221N	S233A	S253L	G259R	Y263C	N266T	I267T	S268G	Frequency of mutant
Second round (9 mutants)																		
2-1										√				√				nd
2-2					√													nd
2-3			√								√				√		√	nd
2-4				√					√				√					nd
2-5	√																	nd
2-6		√						√										nd
2-7						√	√				√							nd
2-8								√								√		nd
2-9									√			√						nd
Third round (30 mutants)																		
3-1									√				√					2
3-2								√										4
3-3									√									2
3-4							√	√										2
3-5										√				√				3
3-6						√	√											2
3-7						√	√				√							2
3-8								√								√		4
3-9									√			√						2
3-10								√			√							2
3-11			√						√				√					5
Fourth round (20 mutants)																		
4-1			√						√									3
4-2			√										√					4
4-3			√					√					√					2
4-4									√				√					3
4-5			√						√				√					8

nd – not determined.

Table 4.7: List of best mutants from second to fourth rounds of high activity evolution.
The grey shading highlights the retention of mutants and the convergent nature of the evolution.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
S122C	7.60 ± 0.1	1.01	1.50 ± 0.10	0.71	5.08×10^3	1.42
N164D	5.41 ± 0.1	0.72	1.08 ± 0.08	0.51	5.03×10^3	1.41
Y221N	4.32 ± 0.1	0.58	0.95 ± 0.12	0.45	4.52×10^3	1.27
S233A	2.14 ± 0.0	0.60	0.67 ± 0.04	0.32	3.20×10^3	0.90
S253L	3.92 ± 0.1	0.52	1.34 ± 0.08	0.64	2.93×10^3	0.82
Y263C	1.87 ± 0.0	0.25	0.57 ± 0.05	0.27	3.27×10^3	0.92
I267T	4.42 ± 0.1	0.59	1.39 ± 0.21	0.66	3.17×10^3	0.89

Table 4.8: Exploration of point mutations derived from the third round of high catalytic evolution of GpdQ.

4.5 DISSOCIATION OF OLIGOMERIC STRUCTURE IN MUTANT 3-25

4.5.1 Initial detection

Protocol B (section 2.6.2, Chapter 2) was used to purify all the GpdQ mutant enzymes listed in Tables 4.4, 4.5 and 4.8. As size exclusion chromatography was used as the first purification step in this protocol, the soluble fraction from the crude lysate was loaded onto the Superdex 200 column as sample. The proteins eluted with poor resolutions (Figure 4.1).

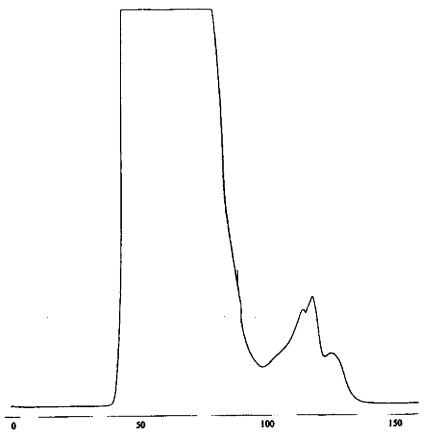


Figure 4.1: Size exclusion chromatography of GpdQ mutant 3-5.
A Superdex 200 elution profile of the soluble fraction of the crude lysate during the purification of mutant 3-5 of high affinity evolution using Protocol B.

Protocol B did not give proteins of sufficient purity for future studies. It was therefore replaced by Protocol C where the purification procedure was reversed, thus making the size exclusion chromatography the second and last purification step. Mutant 3-25 was purified again, this time using Protocol C (section 2.6.2, Chapter 2) so that the characterisation of this best mutant could be repeated and verified.

The Superdex 200 elution profile of mutant 3-25 revealed disruption to the quatenary structure of GpdQ, where the hexameric enzyme had dissociated into dimers (Figure 4.2).

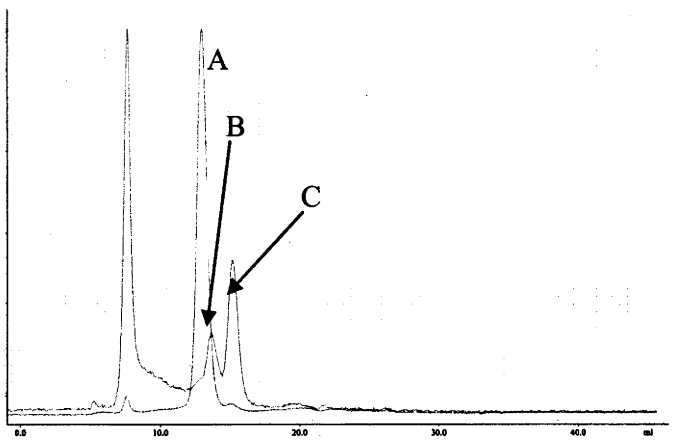


Figure 4.2: Oligomeric dissociation of GpdQ mutant 3-25.
In the 24-mL Superdex 200 column, wt-GpdQ (red) eluted in one single peak (A); mutant 3-25 (blue) eluted in two distinct peaks (B and C).

4.5.2 Determination of protein molecular weight in mutant 3-25

The Superdex 200 size exclusion column (120 mL) used to estimate the size of wt-GpdQ and mutant 3-25 proteins weas pre-calibrated with six protein standards purchased from Sigma (Figures 2.19 and 2.20, Chapter 2).

The size of the wild-type GpdQ hexamer was estimated at 171 kDa. The two forms of mutant 3-25 protein are estimated at 167 kDa and 56 kDa, which probably represented the hexamer and the dimer.

4.5.3 Discovery of two forms of activities in mutant 3-25

The finding of oligomeric dissociation in variant 3-25 suggested that the hexamer and dimer should be characterised separately. Two sets of activities corresponding to the two different forms of the enzyme were observed.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
3-25 dimer	76.06 ± 1.0	10.14	0.13 ± 0.01	0.06	5.76×10^5	161.34
3-25 hexamer	5.29 ± 0.1	0.71	0.96 ± 0.09	0.31	5.54×10^3	1.55

Table 4.9: Characterisation of 3-25 hexamer and dimer with *bp*NPP.

The kinetic data revealed that the 3-25 dimer was a more active form of the enzyme. The structural dissociation had a greater impact upon the catalytic activity than it did on substrate affinity, with a 10-fold k_{cat} increase compared to the wild-type value; while the k_{cat} of 3-25 hexamer was slightly lower than the wild-type's.

Several other GpdQ mutants also experienced a disruption to their oligomeric structures. The enhancement of activity by this phenomenon will be discussed further in Chapter 5.

4.6 STRUCTURE-FUNCTION STUDIES OF MUTATIONS C54G, H217R, Y221H AND G259R

GpdQ structural studies in this project were all performed using a solved structure of the wild-type protein, which can be found in Protein Data Bank (http://www.rcsb.org/pdb/) using the PDB ID of 3D03 [19]. The 3D03 structure (GpdQ-Co^{2+}, 1.9 Å) was crystallised in 60% v/v Tacsimate, 0.1 M Bis-Tris, pH 7.5 at 18°C using methods previously described in Chapter 2. All structural modelling and diagrams presented in this thesis were generated using or based on the same PDB file.

4.6.1 C54G

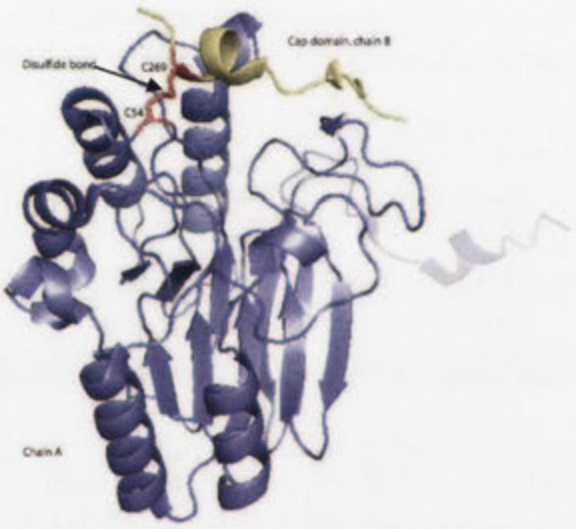


Figure 4.3: The inter-molecular disulfide bond in GpdQ.
A ribbon diagram showing the disulfide bond interaction (red stick) between residue C54 of one GpdQ subunit (chain A, blue) and residue C269 of another subunit (chain B, yellow).

One of the key features of the GpdQ structure is the presence of an inter-chain disulfide bond between residues C54 and C269. A description of the bond was provided in section 1.4.5 in Chapter 1.

A mutation of C54 into a glycine naturally abolished the disulfide bridge, which could result in three major consequences:

i. Reduced stability of the GpdQ dimer. However, since the GpdQ dimer is also held together firmly by the dimerisation domain, it is unlikely that the loss of the inter-chain disulfide bond would result in the dissociation of a dimer molecule into two monomers.

ii. A change in the orientation of the cap domain that 'flips open' and possibly disrupts the interface between two dimer units, thus the stability of the hexamer would also be compromised (Figure 4.4).

iii. Widened substrate entry into the active site that would enhance the activity of the enzyme, especially for non-physiological substrates. This was proven by the increment of the k_{cat}/K_m ratio by almost one magnitude of order. The 5-fold K_m improvement by a single C54G mutation was considerably significant (Table 4.4).

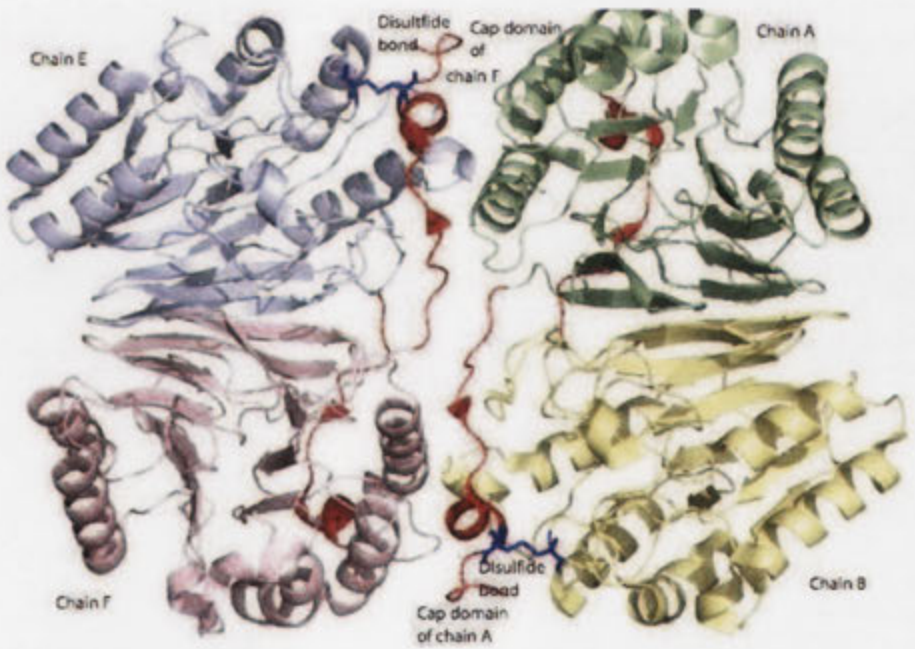


Figure 4.4: Locations of the cap domains at the interface between the dimeric units of GpdQ.
A diagram showing how the cap domains of chains A and F are located at the interface between A-B dimer with E-F dimer. In the absence of the A-B and E-F disulfide bonds (in blue) through the C54G, the interface and the stability of the GpdQ hexamer would be disrupted. In a similar manner, the cap domains of chains B and E would be located at the third C-D dimer (not shown here).

4.6.2 G259R

Glycine and proline residues have unusual conformational abilities and are commonly found in turns or bends in protein secondary and tertiary structures [20-22].

G259, together with its adjacent residue, P260, forms a bend in the cap domain of GpdQ (Figure 4.5A). A mutation at this point would almost certainly alter the orientation of the domain, especially when the side chain of the substitution, R259, would point towards W261 (Figure 4.5B). The effects of this mutation are likely to include destabilisation of the dimer and also the interface between dimeric units, as well as enhanced substrate binding through the changed conformation of the cap domain.

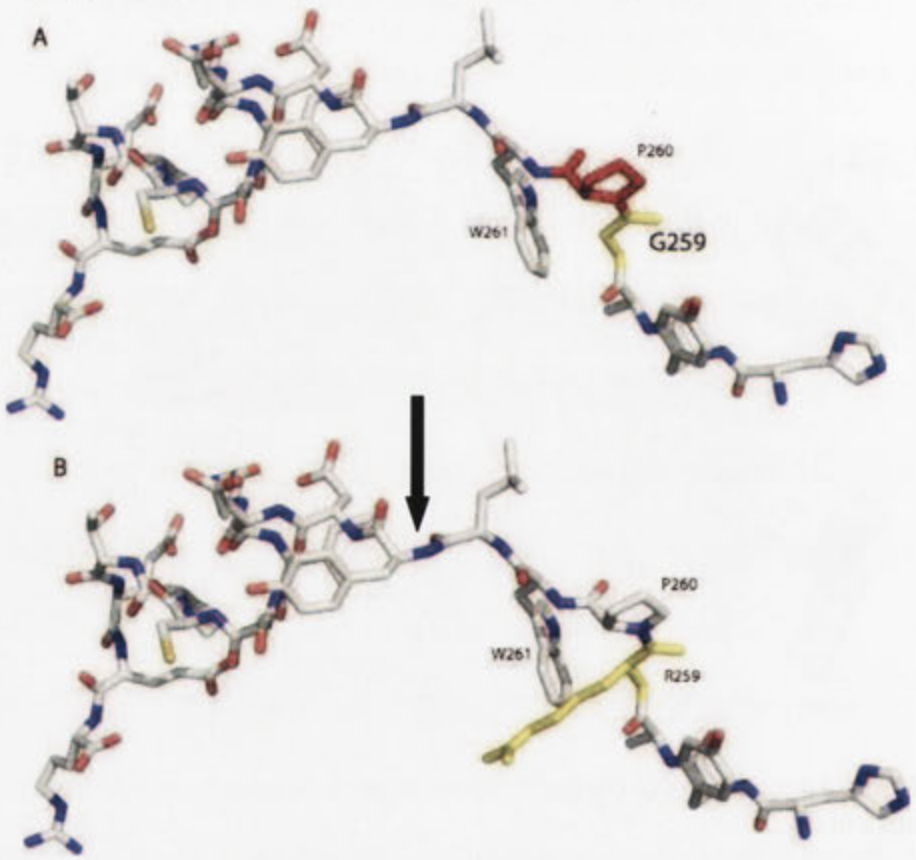


Figure 4.5: Structure-function studies of G259R.

An illustration of the G259-P260 bend on the cap domain of GpdQ in (**A**) and R259 as modelled using MacPyMOL in (**B**).

4.6.3 Y221H

Y221 is located in the outermost among the anti-parallel β-sheets of the dimerisation domain. It interacts directly with another tyrosine residue, Y250 of the other chain (Figure 4.6). The Y221H mutation directly affected the stability of the dimer, but since a full domain comprising about 60 residues (residues 196-255) has been dedicated to strengthen the dimeric structure, it would be unlikely for a single mutation to result in dissociation of a GpdQ dimer into monomers. Y221H possibly caused a change in the pH dependence of the protein or the conformation of the GpdQ dimer that was less conducive for the formation of a hexamer.

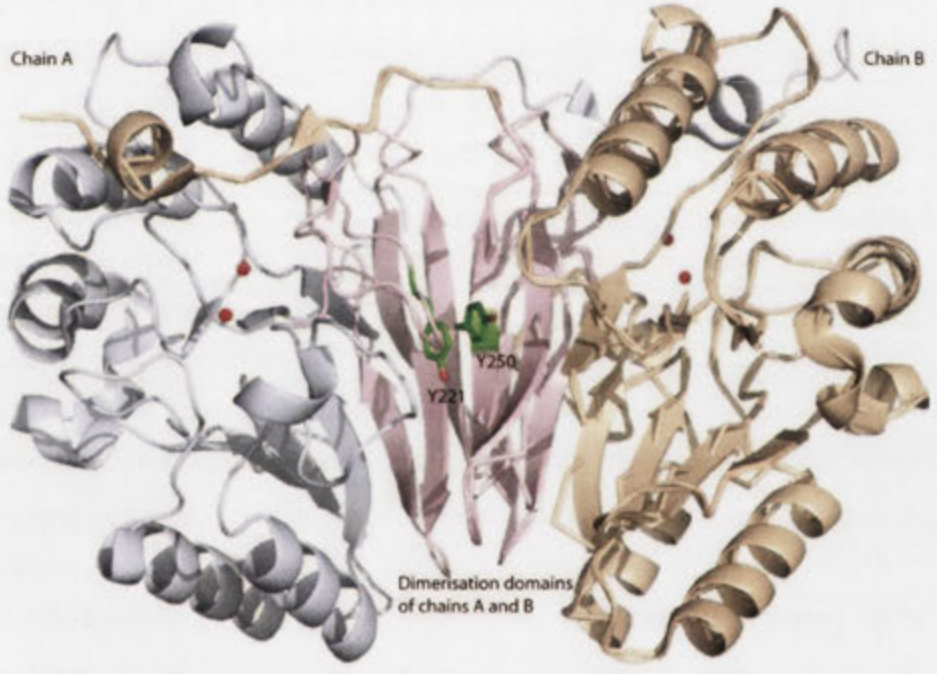


Figure 4.6: The Y221-Y250 interaction in GpdQ.
Y221 of chain A and Y250 of chain B are coloured in green to highlight their locations in the dimerisation domain (pink).

4.6.4 H217R

H217 is located close to the active site (within 5 Å) between H10 and H197 that both coordinate the α-metal at the binuclear metal center of GpdQ. Due to the orientation of the residue, it is unlikely for H217 to have aromatic interactions with the imidazole rings of the two histidine ligands. There are no obvious interactions formed by H217 except for a hydrogen bond that is 2.97 Å in length between its backbone carboxyl and H197, as shown in Figure 4.7(A). It remains uncertain if through the H217R mutation, the strong basic guanidino group of the arginine could act as a cation-π to H10 and H197. Figure 4.7(B) shows a possible conformation of the arginine where the residue is orientated away from the active site. This modelling was generated using the MacPyMOL software. The effects of H217R on GpdQ enzyme activity can only be determined by a crystal structure of the mutant.

Hadler *et al.* suggested that H217 is involved in an extensive hydrogen network within the vicinity of the active site [19]. To investigate the involvement of H217 in catalysis, they made a H217A mutant of GpdQ and discovered that the catalytic turnover was 3 times higher than the wild type while the substrate affinity was not much affected [23]. This result was rather different from the H217R mutation, which appeared to have a greater positive effect on the K_m than it did on k_{cat} (Table 4.5). Furthermore, Hadler *et al.* proposed that H217 is also involved in the binding of the β-metal, despite its distance from the β-site and lack of any important interaction with other residues related to the site.

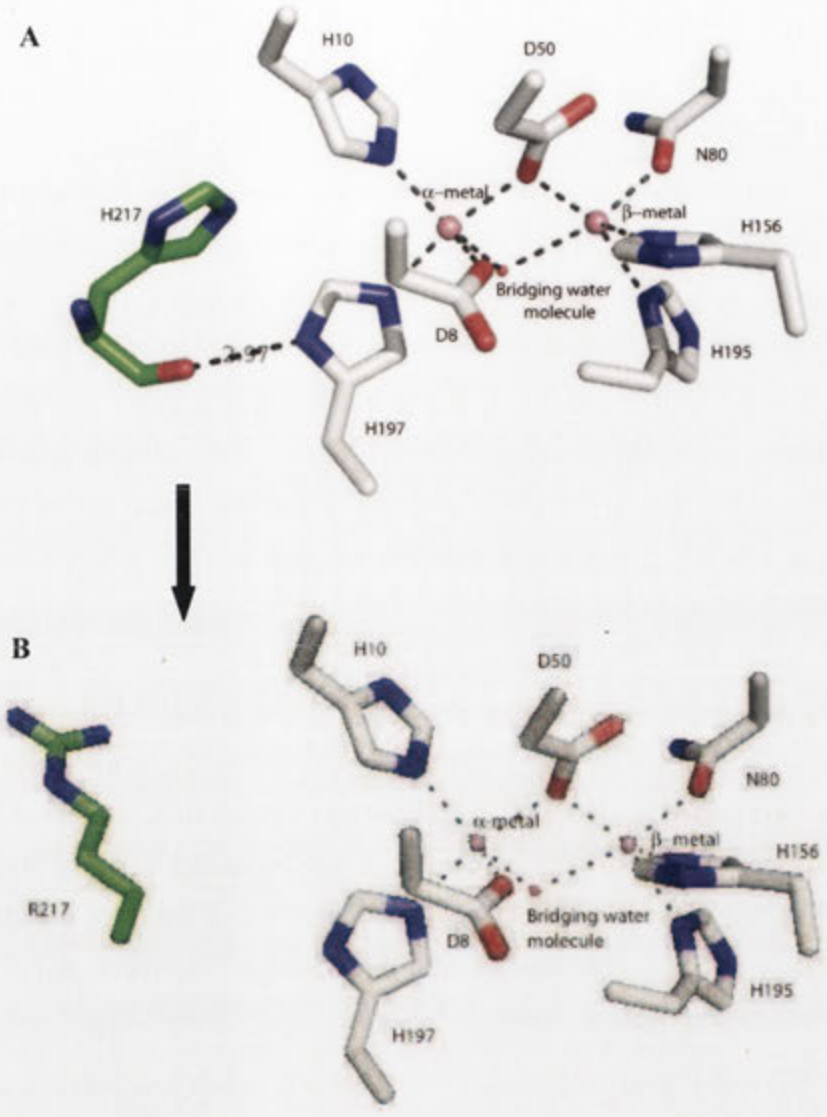


Figure 4.7: Structure-function studies of H217R.
Diagram A shows how H217 is located just outside the active site of GpdQ. Diagram B is H217R modelled using the MacPyMOL software.

4.7 CONCLUSIONS ON DIRECTED EVOLUTION OF GpdQ (PART ONE)

An overview

After experimenting with three screening methods (two *in vivo* agar plate based and *in vitro* random screening) and several phosphotriester (methyl paraoxon) and phosphodiester (dimethyl phosphate, diethyl phosphate and bis (*p*-nitrophenyl) phosphate) substrates, directed evolution of GpdQ was successfully carried out by evolving the enzyme's bis (*p*-nitrophenyl) phosphate (b*p*NPP) hydrolytic activity. The *in vitro* random screening method used for both the preliminary and secondary screenings of GpdQ genetic libraries was adapted from the original protocol devised by Dr Bradley Stevenson.

GpdQ was evolved towards high substrate affinity (low K_M) and high catalytic activity (high k_{cat}) separately. With the high affinity evolution, three rounds of evolution were completed and mutants of K_M values as low as 0.21 mM (mutant 3-25, Table 4.4), which produced a nearly 10-fold improvement. The best mutants derived at the end of the three rounds of high affinity evolution also exhibited improved catalytic activity, as indicated by their high k_{cat} values.

Four rounds of high activity evolution were successfully completed. Interestingly, the variants obtained after the four rounds of evolution carried the same genotypes as those acquired at the end of the third round of high affinity evolution. The best mutant from the two directed evolution projects was triple mutant C54G-Y221H-G259R, commonly referred to throughout this chapter as mutant 3-25.

Mutations and their implications

Both high affinity evolution (Rounds 1-3) and high activity evolution (Rounds 1-4) projects identified four major mutations: C54G, H217R, Y221H and G259R. Three of these four mutations had likely effects on the hexameric quaternary structure of GpdQ

Structure-function studies showed that both C54G and G259R mutations would alter the orientation of the cap domain, which would subsequently disrupt the interface between dimers and the stability of the GpdQ hexamer. The stability of the dimer was likely to be affected too since the cap domain is involved in domain swapping. However, the GpdQ dimer is also formed through the intertwining β-sheets of the dimerisation domain.

Y221H was a mutation in the dimerisation domain that would also have an effect upon the quaternary structure of GpdQ.

H217R was the only mutation whose location was near to the active site. The mutation was believed to have a stabilising effect on the active site, thus enhancing catalysis.

It is interesting to note that mutations of conserved residues, including active residues, were never observed.

Dissociation of the GpdQ hexamer – Dimer was significantly more reactive

Gel filtration analysis of mutant 3-25 showed that the enzyme had dissociated into a mixture of hexamers and dimers. This was not unexpected especially after learning the implications of the mutations through structure-function studies. The hexameric and the dimeric 3-25 enzymes were assayed separately and the high activity was found to be associated with the latter. The 3-25 dimer had an overall k_{cat}/K_M improvement of greater than 150 times compared to the wild type (Table 4.9).

Therefore, directed evolution unintentionally steered the GpdQ enzyme towards an additional direction apart from catalytic activity and substrate affinity improvements: dissociation of the hexameric structure. To paraphrase this more effectively, directed evolution of GpdQ led to a simultaneous optimisation of its b*p*NPP activity and quaternary structure. Such outcomes from directed evolution projects are not unprecedented [24].

Premature convergence

Premature convergence occurs through the loss of genetic variation within a population, which causes the inability of the population to produce offspring that are superior to their parents [25]. Premature convergence is a persistent problem in genetic and evolutionary algorithms, as well as in directed evolution studies [25-28].

At the end of the four rounds of high activity evolution and three rounds of high affinity evolution, only four mutations remained in the population of the screened mutants, suggesting convergence had been reached. Directed evolution experiments were then temporarily shelved while attempts were made to revise and optimise error-prone PCR (epPCR), staggered extension process (StEP) and random screening techniques. The GpdQ structure was also studied further in search of mutations that would form a stable GpdQ dimer.

Evolvability of K_M and k_{cat}

The GpdQ affinity for substrate b*p*NPP (K_M) appeared to have a higher evolvability than its catalytic activity (k_{cat}). Single mutations could result in significant K_M improvements (Tables 4.5 and 4.8) whereas great k_{cat} enhancements required synergy from at least three mutations, with 3-25 as a fine example.

CHAPTER FOUR REFERENCES

1. Ely, F., Foo, J.-L., Jackson, C. J., Gahan, L. R., and Ollis, D. (2007) Enzymatic bioremediation: organophosphate degradation by binuclear metallo-hydrolases. *Current Topics in Biochemical Research 9,* 63-78.
2. Raushel, F. M. (2002) Bacterial detoxification of organophosphate nerve agents. *Current Opinion in Microbiology 5*, 288-95.
3. Singh, S., Kang, S. H., Mulchandani, A., and Chen, W. (2008) Bioremediation: environmental clean-up through pathway engineering. *Current Opinion in Biotechnology 19*, 437-444.
4. Wolfenden, R., and Spence, G. (1967) Derepression of phosphomonoesterase and phosphodiesterase activities in *Aerobacter aerogenes. Biochimica et Biophysica Acta 146*, 296-8.
5. Gerlt, J. A., and Westheim, F. H. (1973) Phosphodiesterase from *Enterobacter aerogenes. Journal of the American Chemical Society 95*, 8166-8168.
6. Gerlt, J. A., and Whitman, G. J. R. (1975) Purification and properties of a phosphohydrolase from *Enterobacter aerogenes. Journal of Biological Chemistry 250*, 5053-5058.
7. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of Escherichia coli coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Applied and Environmental Microbiology 70*, 404-412.
8. Ghanem, E., Li, Y. C., Xu, C. F., and Raushel, F. M. (2007) Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry 46*, 9032-9040.
9. Wai, S. N., Mizunoe, Y., Takade, A. and Yoshida, S. (2000) A comparison of solid and liquid media for resuscitation of starvation- and low-temperature-induced nonculturable cells of *Aeromonas hydrophila. Archives of Microbiology 173*, 307-310.
10. Jacobs, S. E., and Harris, N. D. (1961) The effect of modifications in the counting medium on the viability and growth of bacteria damaged by phenols. *Journal of Applied Bacteriology 24*, 172-181.
11. Pellon, J. R., and Gomez, R. F. (1979) A note on the effect of different types of agar on the recovery of heated *Escherichia coli* K12. *Journal of Applied Bacteriology 46*, 331-334.
12. Payne, J. (1978) Damage and recovery in Staphylococci. In: Skinner, F. A., Quesenl LB (eds) Symposium series 7. *Streptococci Society for Applied Bacteriology. Academic Press, London*, 349-369.
13. Miller, R. J., and Davey, C. B. (1965) Influence of the microphysical structure of a system on the growth of bacteria. *Canadian Journal of Microbiology 11*, 761-763.

14. Cho, C. M. H., Mulchandani, A., and Chen, W. (2002) Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. *Applied and Environmental Microbiology 68*, 2026-2030.
15. Turner, N. J. (2006) Agar plate-based assays in Enzyme Assays: High-Throughput Screening, Genetic Selection and Fingerprinting. *Wiley-VCH Chapter 5*, 139-161.
16. Shenoy, A. R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S. S., and Podobnik, M. (2007) Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis*. *Journal of Molecular Biology 365*, 211-225.
17. Tehara, S. K., and Keasling, J. D. (2003) Gene cloning, purification, and characterization of a phosphodiesterase from *Delftia acidovorans*. *Applied and Environmental Microbiology 69*, 504-508.
18. Jiang, W. D., Xu, B., Zhong, J. B., Li, J. Z., and Liu, F. (2008) Hydrolytic cleavage of bis(p-nitrophenyl) phosphate by Schiff base Mn-III complexes containing morpholine pendants in Gemini 16-6-16 micelles. *Journal of Chemical Sciences 120*, 411-417.
19. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from Enterobacter aerogenes. *Journal of the American Chemical Society 130*, 14129-14138.
20. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics 23*, 566-579.
21. Richards, F. M., and Kundrot, C. E. (1988) Identification of structural motifs from protein coordinate data - secondary structure and 1st-level supersecondary structure. *Proteins-Structure Function and Genetics 3*, 71-84.
22. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577-2637.
23. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural Flexibility Enhances the Reactivity of the Bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.
24. Vamvaca, K., Butz, M., Walter, K. U., Taylor, S. V., and Hilvert, D. (2005) Simultaneous optimization of enzyme activity and quaternary structure by directed evolution. *Protein Science 14*, 2103-2114.
25. Liu, J., Cai, Z., and Liu, J. (2000) A novel genetic algorithm preventing premature convergence by chaos operator. *Journal of Central South University of Technology 7*, 100-103.

26. Andre, J., Slarry, P., and Dognon, T. (2001) An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization. *Advances in Engineering Software 32*, 49-60.
27. Harvey, I., and Thompson, A. (1997) Through the labyrinth evolution finds a way: A silicon ridge. *Lecture Notes in Computer Science 1259*, 406-422.
28. Wedge, D. C., Rowe, W., Kell, D. B., and Knowles, J. (2009) In silico modelling of directed evolution: Implications for experimental design and stepwise evolution. *Journal of Theoretical Biology 257*, 131-141.

5

OPTIMISATION OF GpdQ OLIGOMERIC STRUCTURE WITH RATIONAL DESIGN FOR BIS (*P*-NITROPHENYL) PHOSPHATE ACTIVITY

CHAPTER FIVE: OPTIMISATION OF GpdQ OLIGOMERIC STRUCTURE WITH RATIONAL DESIGN FOR BIS (*P*-NITROPHENYL) PHOSPHATE ACTIVITY

5.1 INTRODUCTION

One of the key objectives of this research project was to evolve GpdQ's weak and promiscuous activities. After experimenting with several screening methods and substrates, GpdQ was eventually evolved towards high activity with a non-physiological phosphodiester substrate, bis (*p*-nitrophenyl) phosphate (b*p*NPP).

The first part of the directed evolution studies in this project was made up of the rounds 1-4. Results and discussions of work from these four rounds were presented in Chapter 4. At the end of the four rounds, directed evolution experiments were discontinued for two reasons. Firstly, premature convergence significantly reduced the diversity of the population of mutations. Experiments and techniques, especially those related to generating diversity in the libraries, had to be redesigned.

Secondly, two of the four mutations in the best mutants screened from the first part of directed evolution had direct implications on the cap domain of GpdQ. It appeared that these mutations also broke down the oligomeric structure of the protein. These two mutations in question are C54G and G259R. The C54G mutation abolished the disulfide bond between the cap domain of one chain and the catalytic domain of a second chain (Figures 1.16, 4.3 and 4.4). G259R, on the other hand, was a mutation on the cap domain, which was likely to change the conformation of the domain. At first glance, these two mutations appeared to have a more direct effect on the stability of the GpdQ dimer, since the cap domain is involved in domain swapping, a known mechanism in the dimerisation of protein chains. However, GpdQ has an additional secondary domain, the dimerisation

domain, consisting of two pairs of anti-parallel β-sheets that increase the stability of the GpdQ dimer.

Instead, both C54G and G259R mutations likely affect the stability of the GpdQ hexamer. This hypothesis was made based on the location of the cap domain at the interface between the GpdQ dimers and strengthened by the results of the gel filtration analysis on mutant enzyme 3-25 (C54G, Y221H, G259R), which confirmed the hexamer → dimer dissociation (Figure 4.2). Assay studies also revealed that the activity of the dimeric 3-25 enzyme was remarkably higher than that of the 3-25 hexamer (Table 4.9).

The dissociation resulted in the disruption of the D_3 dihedral symmetry of the hexamer. The C_2 cyclic symmetry of the dimer remained intact.

The first part of directed evolution studies of GpdQ and structural analysis delivered an important message which would serve as the pointer for the rest of the experiments conducted in this project: The GpdQ dimer was optimal for the hydrolysis of b*p*NPP.

To this end, the structure of GpdQ was analysed again to identify mutations that could promote the dissociation of the hexamer. In other words, this chapter is a record of site-specific mutants and their studies that were all dedicated to the goal of producing a stable GpdQ dimer.

5.2 FACTORS AFFECTING THE DISSOCIATION EQUILIBRIUM

Dissociating GpdQ mutants did not migrate in two or more distinct peaks during size exclusion chromatography with the 120 mL Superdex 200 column, even when the sample volume and flow rate were reduced to less than 1% of the column volume (1% = 1.2 mL) and 0.25 mL per minute, respectively. The column was repacked at one point and re-equilibrated with the eluting buffer (50 mM Tris-HCl, pH 8.0) that had been filtered and

degassed (both with an air vacuum system) thoroughly to enhance its resolution. The column was also subject to regular maintenance cleaning with 0.5 M NaOH and mQH_2O. Notwithstanding the efforts made to minimise the contribution of column resolution as a factor in poor separation, dissociating GpdQ mutant proteins still eluted in overlapping peaks.

Dissociating mutant enzymes were initially characterised as hexamer-dimer mixtures. These two oligomeric forms of GpdQ were assayed separately. This was achieved by selecting elution fractions with pure hexamers (fraction 4, Figure 5.1) or dimers (fractions 6 and 7) to be assayed, while fractions containing a mixture like fraction 5 are eliminated. The vast difference in enzymatic activities implied that diluting the enzymes did not cause the hexamer to dissociate into dimers.

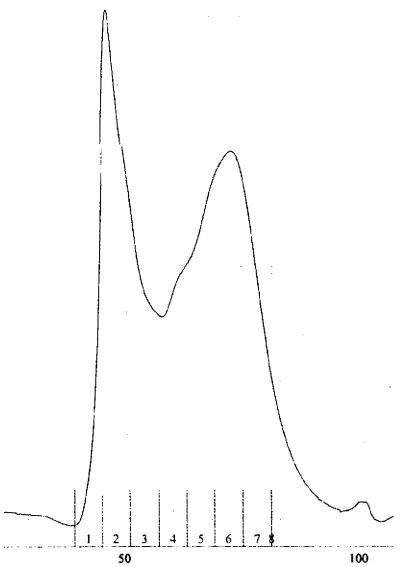


Figure 5.1: An excerpt of Superdex 200 migration profile of GpdQ triple mutant enzyme 3-25.

A series of gel filtration experiments were run to study the effects of enzyme concentration and presence of exogenous metal ions on the hexamer-dimer equilibrium.

Enzyme concentration

At high concentrations, hexamers were the dominant species of dissociating GpdQ enzymes. (Figure 5.2).

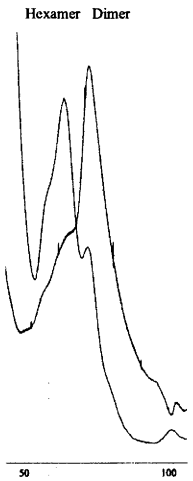


Figure 5.2: The effects of enzyme concentration on oligomeric dissociation in GpdQ.
Elutions of 9.35 mg/mL (blue) and 1.77 mg/mL (red) of mutant Y257stop respectively, from Superdex 200 column. The Y257stop mutant will be discussed in detail in section 5.5.1.

Metals

The presence of cofactor metals had minimal effect upon the re-association/dissociation of the enzyme (Figure 5.3).

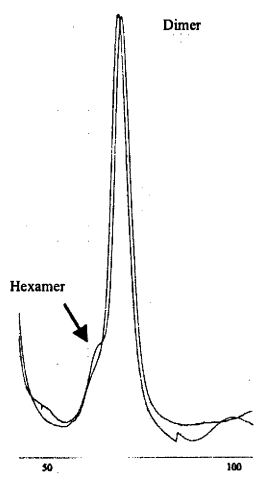


Figure 5.3: The effects of metal ions on oligomeric dissociation in GpdQ.
Elutions of mutant 3-25-C269A with 1 mM Mn^{2+} (blue) and without Mn^{2+} (violet).

5.3 C269 MUTANTS

Perhaps the most conspicuous mutation, detected since the first round of directed evolution and conserved all throughout the 20 fittest mutants after three rounds of high b*p*NPP affinity evolution and also the 20 best mutants after four rounds of high b*p*NPP activity evolution (Tables 4.3 and 4.7, Chapter 4), was C54G. Based on the information on the structure of GpdQ, it can be inferred, quite directly, that a mutation at position C54 would destroy the inter-chain disulfide bridge between the residue and C269. Consequently, C269S and C269A were the first rationally designed GpdQ mutants in this study. Site-directed mutagenesis was used to make C269 as well as all other mutants discussed in this chapter. Experimental procedures are described in section 2.3 of Chapter 2.

5.3.1 C269S and C269A

Both C269S and C269A single mutations caused GpdQ to dissociate into an equilibrium of hexamer-dimer, with majority of the enzyme remaining in assembly (Figure 5.4).

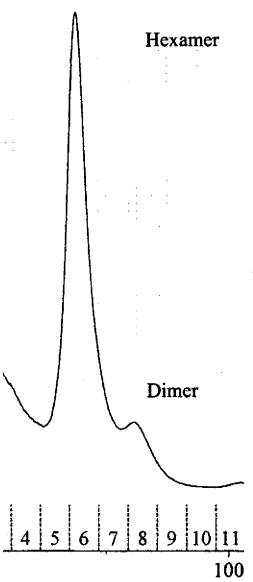


Figure 5.4: Gel filtration analysis of C269A.

C269S and C269A GpdQ enzymes as well as the rest of the enzymes discussed in this chapter were expressed and purified using Protocol C as described in Chapter 2 (section 2.6.2).

The kinetic properties of these two mutants are given in Table 5.1. The hexamer-dimer equilibrium enzyme mixtures were used for the characterisation. All assays reported in this chapter were performed in 50 mM CHES buffer pH 9.0, 0.5 mM Mn^{2+}, 0.1 mg/mL BSA at 20°C.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
C54G	13.74 ± 0.2	1.83	0.43 ± 0.04	0.20	3.21×10^4	8.99
C269S	47.68 ± 1.3	6.36	1.21 ± 0.17	0.58	3.93×10^4	11.01
C269A	36.02 ± 1.4	4.80	1.48 ± 0.27	0.70	2.43×10^4	6.81

Table 5.1: Characterisation of C269S and C269A.
Kinetic parameters of C54G are shown here again for comparison.

Kinetic data tabulated above show significant enhancements upon both the catalytic activity and substrate affinity of GpdQ, resulting in an overall ~7-fold and 11-fold increment in k_{cat}/K_M in C269A and C269S, respectively. All the three mutations of C54G, C269S and C269A underscored the importance of the absence of the inter-chain disulfide bond for enzyme activity. However, while C54G improved the K_M more remarkably than it did the k_{cat}, C269 mutations had a more profound effect on the activity.

Although these kinetic parameters indicated C269S as a superior mutant compared to C269A, it was the latter that was retained until the eighth and final round of directed evolution. During the sixth round of evolution, both C269S and C269A single site mutants were added into the DNA template mixture for the StEP PCR. C269A was the only genotype that was retained among the best mutants after screening (Tables 6.1 and 6.2). This retention highlights the importance of combining rational design into directed

evolution studies, as a mutation from a cysteine to an alanine (TGT → GCT), involving two nucleotide changes, could not have been achieved through the random mutagenesis approach used.

5.3.2 3-25-C269S and 3-25-C269A

Triple mutant 3-25 (C54G, Y221H, G259R) was the most catalytically efficient mutant obtained after four rounds of directed evolution. Mutants 3-25-C269S and 3-25-C269A were constructed via site-directed mutagenesis (SDM), using 3-25 as the template DNA for SDM PCR. Kinetic parameters of these two new mutants are summarised in Table 5.2. The hexameric and

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_m (mM)	Relative K_m	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_m
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
3-25 Hexamer	5.29 ± 0.1	0.71	0.96 ± 0.09	0.31	5.54×10^3	1.55
3-25 Dimer	76.06 ± 1.0	10.14	0.13 ± 0.01	0.06	5.76×10^5	161.34
3-25-C269S Hexamer	6.79 ± 0.1	0.91	0.59 ± 0.02	0.28	1.16×10^4	3.25
3-25-C269S Dimer	161.30 ± 1.4	21.51	0.26 ± 0.04	0.12	6.15×10^5	172.27
3-25-C269A Hexamer	14.13 ± 0.3	18.84	0.59 ± 0.07	0.28	2.38×10^4	6.67
3-25-C269A Dimer	151.05 ± 1.4	20.14	0.27 ± 0.02	0.13	5.68×10^4	159.10

Table 5.2: Assay results of hexamers and dimers of mutants 3-25-C269S and 3-25-C269A.
The kinetic parameters of mutant 3-25 are shown here again for easy reference and comparison.

Table 5.2 shows vividly how an additional C269S/A mutation doubled the k_{cat} value of 3-25 dimer. However, the improvement in k_{cat} was compromised by a commensurate increase in K_M. This further substantiates the proposition that C269 mutations improved the catalytic activity of GpdQ considerably more than substrate affinity. The k_{cat}/K_m ratio of the 3-25-C269S dimer was also slightly higher than that of its C269A counterpart, which once again confirmed serine as a better substitution for C269. Thus, it is unclear why directed evolution was preferential towards the C269A mutation instead.

Most importantly, Table 5.2 presents three instances of the GpdQ dimer as a more potent form of the enzyme, with approximately 20 and 10-fold enhancements on the k_{cat} and K_m, respectively compared to the native enzyme.

5.3.3 Hysteretic kinetics of C269 mutants

Figure 5.5 shows a typical example of a non-linear progress curve derived from kinetic assays of mutants C269S and C269A. Each assay reaction was run and measured over a period of 10 minutes to sufficiently allow the reaction to approach equilibrium or steady state. The values of initial velocity, *v*, were derived from the steady-state kinetics.

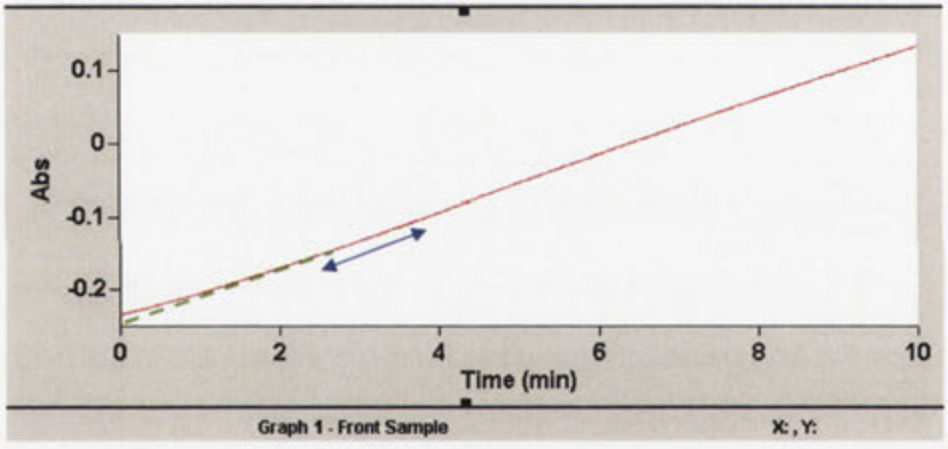


Figure 5.5: Hysteretic kinetics of GpdQ mutant C269S.
The red curve show the kinetic progress of 20 mM b*p*NPP hydrolysed with 0.625 nM of C269S for 10 minutes. The green dashed line was drawn to show the deviation of the curve from linearity. The blue arrow indicates the area of the curve (steady-state kinetics) from where the value of the initial velocity, *v*, was calculated.

Carl Frieden was responsible for introducing the term *hysteresis* to describe one form of non-linear kinetics, where the initial velocity displays a sigmoidal dependence on substrate concentration [1]. The term hysteresis was used because the behaviour of such enzymes was likened to the time lag taken by a body in reacting to exterior forces. Therefore, enzymes exhibiting

hysteresis undergo a time lag or a slow transition in response to a change in ligand concentration which in most cases, is represented by the addition of substrate to initiate a catalytic reaction. In other words, the pre-incubated enzyme is intrinsically inactive but goes through conformational changes upon substrate induction and becomes fully active when the kinetic progress curve becomes linear. Hysteric kinetics may also be indicative of quaternary structural fluctuations in oligomeric enzymes, where the lag time is the time taken for the enzyme to fully transform through association or dissociation into its active form [1-3].

However, as sigmoidal kinetic progress curves are diagnostic of both positive cooperativity and substrate-induced hysteretic kinetics, it is important to distinguish these two enzyme mechanisms. Not all hysteretic enzymes are cooperative, and vice versa. Neet and Ainslie proposed a system to divide hysteretic enzymes into four classes [4]. Enzymes that fall into Classes III and IV are hysteretic and cooperative simultaneously.

Class III that describes cooperativity in monomeric enzymes can be discounted immediately for C269S/A mutants as the enzymes did not dissociate into monomers.

The three progress curves presented in Figure 5.6 below would help to eliminate the possibility of cooperativity being prevalent in C269S/A mutants.

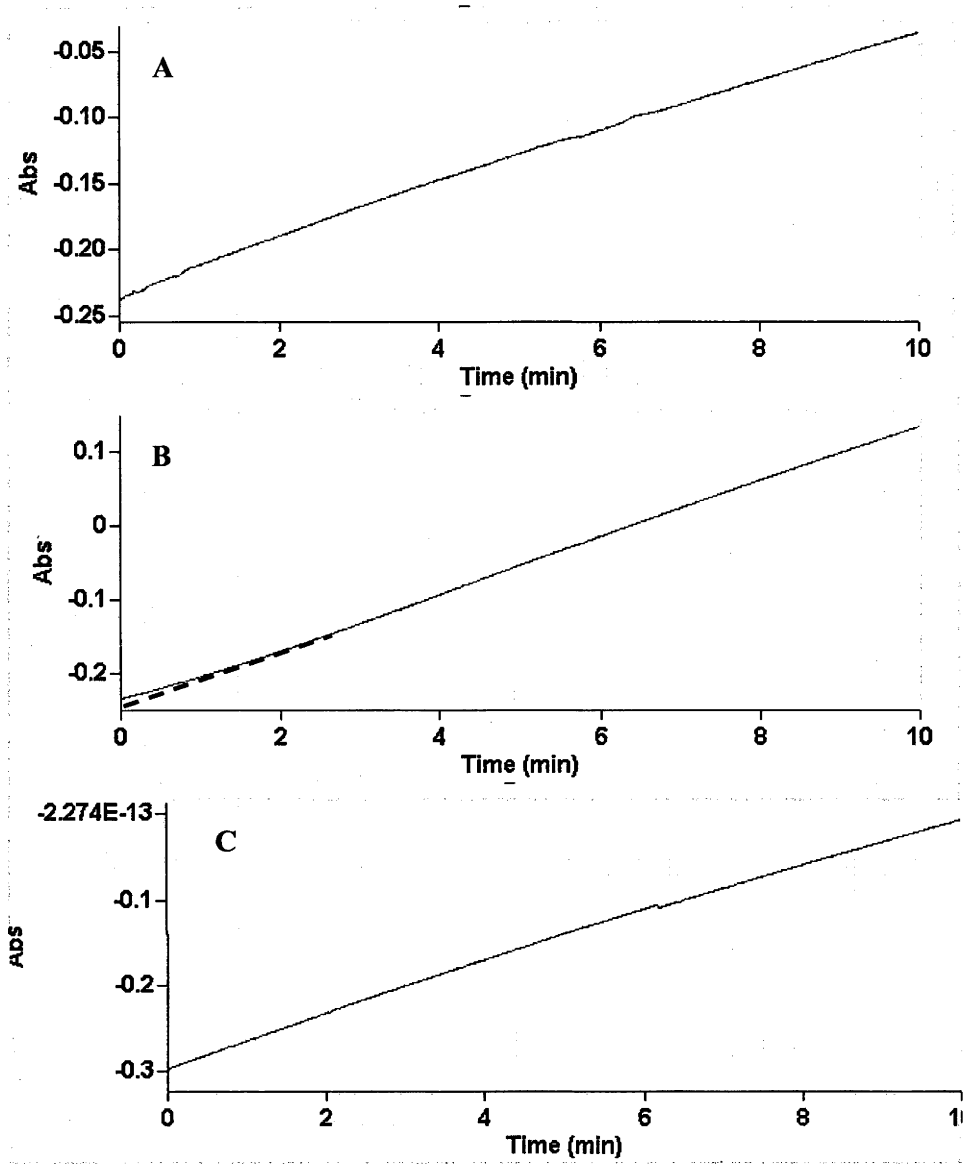


Figure 5.6: The kinetic progress curves of GpdQ C269 mutants.
10-minute hydrolysis of 20 mM b*p*NPP by **(A)** hexamer-dimer enzyme mixture of mutant 3-25-C269A; **(B)** 3-25 hexamer and **(C)** 3-25 dimer.

In Figure 5.6, hysteresis was detected only when 20 mM b*p*NPP was hydrolysed with the pure 3-25 hexamer (Experiment B). When the substrate was reacted with the pure 3-25 dimer and the 3-25 hexamer-dimer mixture, the classic hyperbolic progress curves were obtained (Experiments A and C).

In Experiment B, the hysteresis that was shown implies the partial inactivity of the pre-incubated 3-25-C269A hexamer which, upon the addition of b*p*NPP the substrate, underwent a ~2-minute transition time to convert into its active form. It is very tempting to assume that by 'active form' the dimeric state of the enzyme is meant, but the disparity between the catalytic reactivity of these two forms of enzyme, as shown in Table 5.2, does not reconcile with this notion. Since values of initial velocity were derived from the post-transitional part of the progress curves (indicated by the blue arrow in Figure 5.5), and if the lag is the time taken for hexamers to dissociate into dimers, the enzyme activity in Experiment B should be similar to both those attained from Experiments A and C. Hence, it is more reasonable to conclude that the exhibited lag period was the time required for the hexamer to undergo conformational changes to its tertiary fold, especially the active site to perfectly fit or envelope the structure of substrate molecules in order to catalyse the hydrolysis of the phosphodiester bond efficiently. Due to its low activity, it is unlikely that the dissociation of hexamers occurred at all.

A probable explanation for the lack of hysteresis in Experiment A is that the four mutations had caused most of the enzyme molecules to dissociate into dimers before any further induction by the substrate (Figure 5.3).

Determining positive cooperativity in C269 mutants

Two methods of analysis were used to determine the presence of positive cooperativity in C269 mutant enzymes: The Eadie-Hofstee plot and the Hill equation.

The Eadie-Hofstee plot

The Eadie-Hofstee plot is, as previously described in Chapter 3 (Equation 3.1), one of the several ways of linearising the Michaelis-Menten equation by plotting the reaction initial velocity v, as a function of initial velocity/substrate concentration ratio (v/[S]). Eadie-Hofstee plots of

standard Michaelis-Menten kinetic data would yield a straight line with a negative gradient (Figure 3.3).

An Eadie-Hofstee plot that is shaped like a curve convex to the right is diagnostic of positive cooperativity. An example of this kinetic pattern is shown in Figure 5.7.

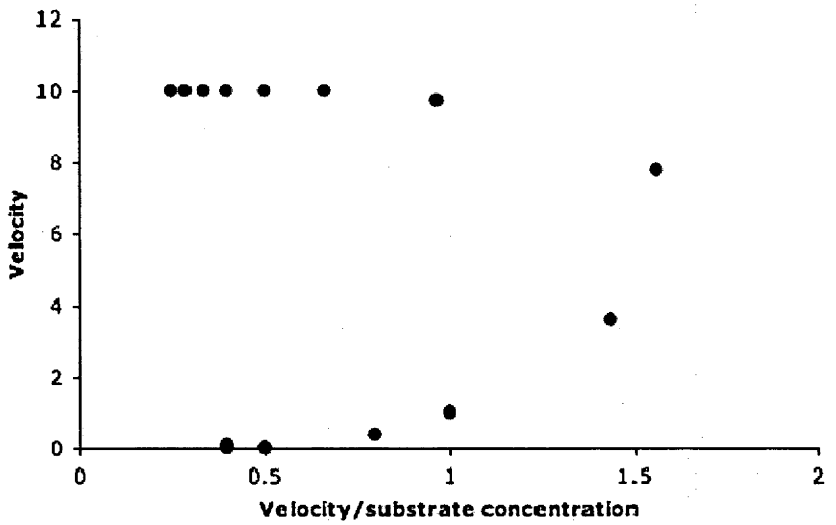


Figure 5.7: Positive cooperativity presented in an Eadie-Hofstee plot.

The Eadie-Hofstee plots of C269S/A and 3-25-C269S/A enzyme reactions, as shown in Figure 5.8, were not indicative of positive cooperativity.

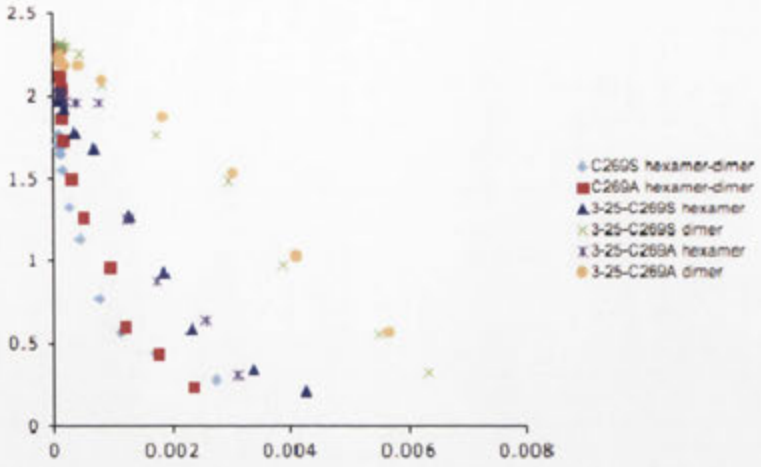


Figure 5.8: Eadie-Hofstee plots of C269 mutants.

The Hill equation

The Hill equation (Equation 3.2), its use, linearised version (Equation 3.3) and how the values of the Hill co-efficient, *n*, have been previously described in Chapter 3. The Hill plots of C269S/A, 3-25-C269S/A enzyme reactions with b*p*NPP are drawn using KaleidaGraph (Figure 5.9) and values of *n* are calculated (Table 5.3).

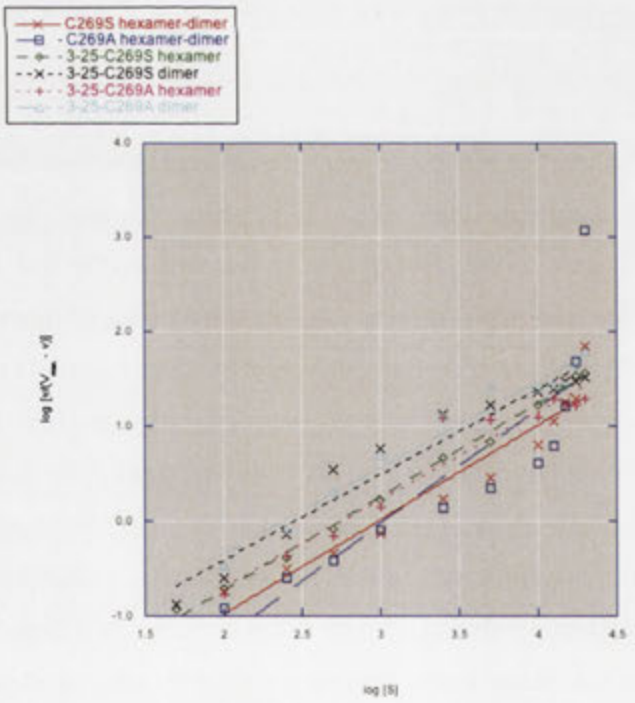


Figure 5.9: Hill plots of GpdQ C269 mutants.
Hill plots of C269S/A, 3-25-C269S/C269A catalytic hydrolysis of substrate b*p*NPP.

Enzyme	*n*	R^2
C269S hexamer-dimer equilibrium	1.0	0.96
C269A hexamer-dimer equilibrium	1.2	0.86
3-25-C269S hexamer	1.0	1.00
3-25-C269S dimer	0.9	0.98
3-25-C269A hexamer	0.9	0.97
3-25-C269A dimer	0.9	0.99

Table 5.3: Hill coefficient (*n*) and correlation coefficient (R^2) values of the catalytic reactions in Figure 5.9.

The calculated values of *n* in Table 5.4 are not very accurate as many of the corresponding R^2 values were low. But they nevertheless demonstrate the absolute ($n = 1$) and almost non-cooperativity ($n \leq 1$) of all the enzymes examined. An *n* value of 1.2 as portrayed by C269A hexamer-dimer cannot be considered as an indication of positive cooperativity as the Hill equation fits the kinetic data of this enzyme very poorly with an R^2 value of 0.86.

As previously discussed in Chapter, the Hill equation has many limitations despite its simplicity that appeals to many researchers. The true definition of the Hill coefficient (*n*) and what it signifies remains ambiguously interpreted and understood. For example, the Hill coefficient has been misused for a physical application to estimate the number of ligand binding sites [5]. Another fallacy commonly practised by some scientists is to draw unequivocal conclusions about the cooperativity of a protein based on kinetic data that are poorly fit to the Hill equation. Among the broad range of points plotted, only those in the middle would fit the equation exactly and they fall within the useful range of data for an accurate measurement of the Hill coefficient. Data points outside this range are often non-linear and will definitely affect the measurement accuracy when they are considered [6, 7]. One fine example from this study is the C269A hexamer-dimer (Figure 5.9)

where the Hill plot of the enzyme equilibrium clearly progressed into a non-linear curve as the substrate increased in concentration. If only data points ≤ 4 along the *x*-axis were used for the curve fit instead, the generated plot would be closely parallel to the other plots, which would then yield a lower *n* value.

Conclusions on the hysteretic kinetics of C269 mutants

All four C269 mutants, namely C269S, C269A, 3-25-C269S and 3-25-C269A exhibited a 2-minute lag in their kinetic progress curves (Absorbance *vs* time). This lag was found to be associated only with the hexamer (3-25-C269S and 3-25-C269A hexamers) and hexamer-dimer enzyme mixtures (C269S and C269A) but not the dimeric enzymes (3-25-C269S and 3-25-C269A dimers).

The best explanation for the hysteresis was that the C269 hexameric enzymes were initially inactive but went through conformational changes upon the addition of the substrate b*p*NPP into the reaction mixture.

Both Eadie-Hofstee plots and the calculation of the Hill coefficients (*n*) confirmed that all of the C269 mutants were non-cooperative enzymes.

5.4 N53 MUTANTS

N53 in GpdQ forms inter-chain hydrogen bond with S268 with the nitrogen in its amide group [7] (Figure 5.10). This interaction, together with two others (C54-C269 and R56-P270), is believed to help stabilise the hexameric and dimeric quaternary organisations of the protein.

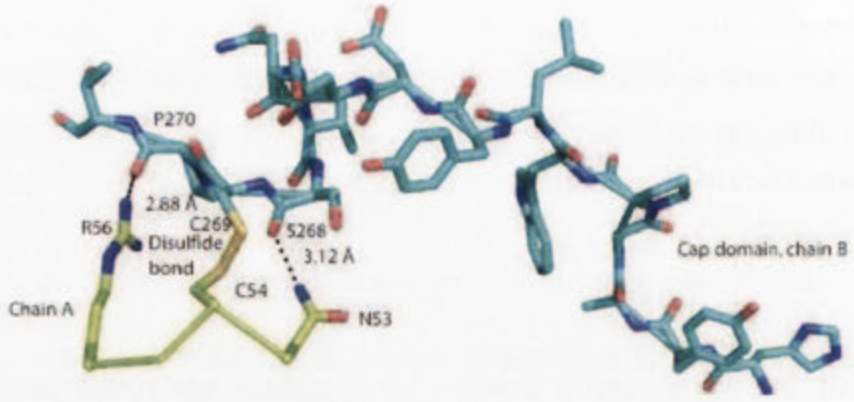


Figure 5.10: The three inter-chain bonds at the cap domain of GpdQ.

Furthermore, N53 is located at the periphery of the active site of GpdQ. The oxygen from the amide functional group of N53 is hydrogen bonded to H10, an α-metal ligand. The main chain carboxyl group of N53 is 2.80 Å away from H81, the conserved histidine in the GNHD motif of dimetallophosphatases, which is always within 5 Å from the active site [9]. N53 is homologous to D66 in the Rv0805 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* (Figure 5.11).

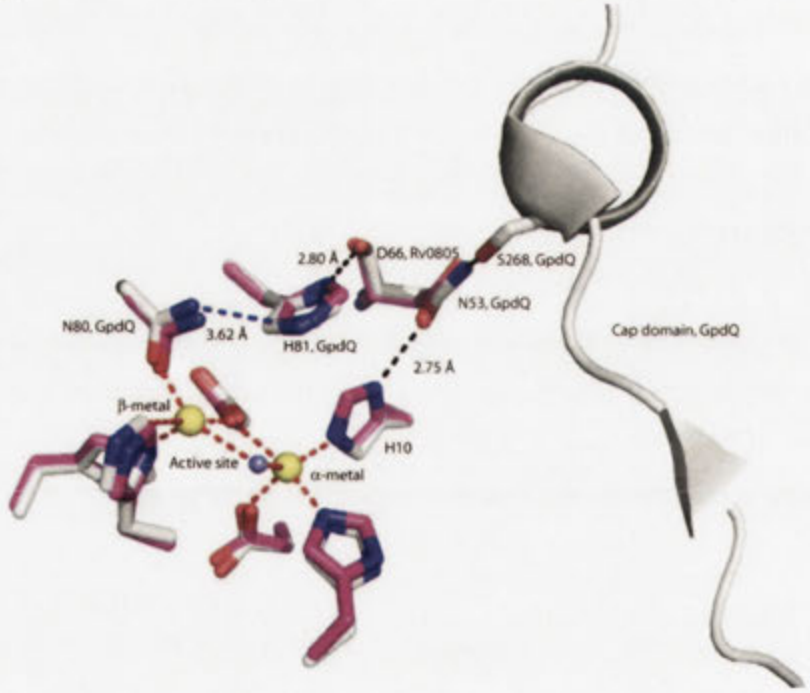


Figure 5.11: Structural alignment of the active sites of Rv0805 and GpdQ.
The backbones of Rv0805 and GpdQ residues are coloured in pink and white respectively.

5.4.1 N53D and N53A

Mutational analysis of N53 was carried out by making two mutants, where the residue was replaced by an aspartate and an alanine respectively. The dissociation phenomenon was also seen in both N53D and N53A, where the purified enzymes contained a mixture of hexamer and dimer. Kinetic results of the mixtures assayed with b*p*NPP are tabulated below (Table 5.4).

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
N53D	13.29 ± 0.3	1.77	1.39 ± 0.01	0.66	9.55×10^3	2.68
N53A	41.53 ± 2.3	5.54	2.28 ± 0.05	1.09	1.82×10^4	5.10

Table 5.4: Kinetic parameters of mutants N53D and N53A.

In N53D, the interaction with S268 was made impossible as the two carboxyl groups from S268 and D53 would repel each other. This means the stability of the quaternary organisation of GpdQ would be compromised, resulting in dissociation that was beneficial for the catalytic activity and substrate affinity of the enzyme.

An alanine substitution for N53, on the other hand, would annul all the hydrogen bond interactions with S268, and H10. Interaction with H81 would not be affected by any mutation as the hydrogen bond was established between H81 and the backbone carboxyl of any amino acid at position 53. The k_{cat} improvement for N53A was more remarkable than N53D's. However, the mutation caused N53A to lose its substrate affinity. The most likely explanation for the poor K_M was the loss of interactions with H10 and H81.

5.4.2 Comparison of GpdQ N53 mutants with Rv0805 D66A

As stated in section 1.5 in Chapter 1, the Rv0805 3',5'-cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis* is the enzyme with the closest in sequence (Figure 1.20) and structure (Figures 1.21 and 5.12) to GpdQ.

Shenoy *et al.* constructed the Rv0805 D66A mutant when the aspartic acid was shown to align with a tyrosine, Y167 in the kidney bean purple acid phosphatase (KBPAP) [10]. However, in the KBPAP, this tyrosine residue is a ligand to the Fe^{3+} ion and is responsible for imparting the purple colour to the protein. D66 does not coordinate any metal ion directly but interacts with H23, a metal ligand. The crystal structure of D66A revealed the loss of this hydrogen bond that consequently affected the positioning of H23 and the Fe^{3+} ultimately ion at the active site [11].

In Figure 5.12, it is shown that the backbone carboxyl of N53 is hydrogen bonded to the imidazole group of the conserved histidine in the GNHD motif, H81. In the same manner, the backbone carboxyl of D66 interacts with H98 in Rv0805. Mutagenesis of D66 or N53 to alanine should not affect the interaction due to the location of the carboxyl group in the main chain. This was confirmed by kinetic studies of Rv0805 D66A and GpdQ N53A enzymes, where no drop in activity was observed in both mutants. However, in other α/β sandwich metallophosphatases, the side chains of these two residues have been shown to interact, *eg.* D52 and H76 (homologous to N53 and H81 in GpdQ respectively) in bacteriophage λ Ser/Thr phosphatase (λ-PP) [11] (Figure 5.12). Mutagenesis of D52 in λ-PP led to a dramatic decrease in the enzyme activity [12].

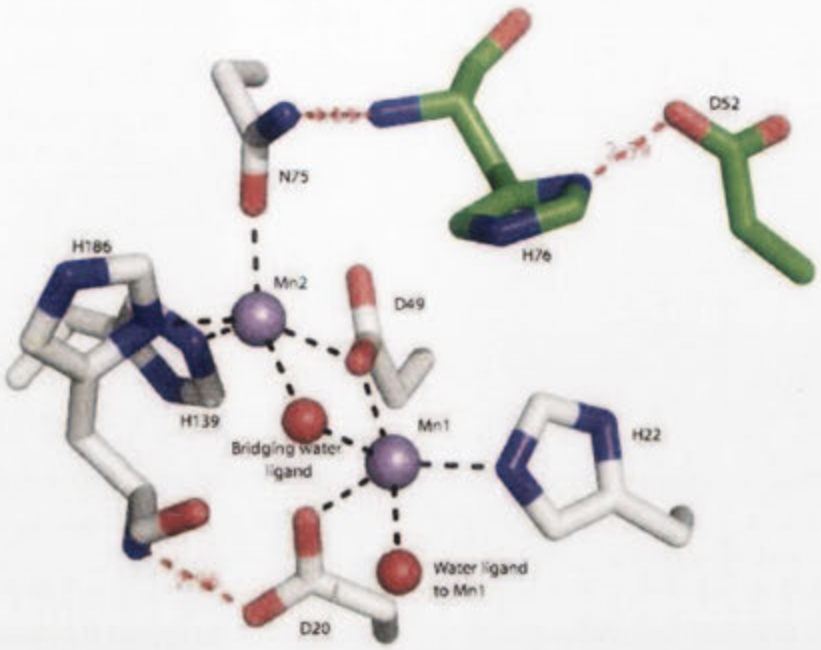


Figure 5.12: The active site of bacteriophage λ Ser/Thr protein phosphatase and the hydrogen network of H76 and D52.
The side chains of H76 and D52 are hydrogen bonded to each other; this interaction does not exist between the homologous residues in Rv0805 and GpdQ. This diagram was generated from the PDB structure 1G5B.

D66A displayed a mild increase in V_{max} but a significant increase in K_m when assayed with varying concentrations of Mn^{2+}. The sigmoidal kinetics observed indicated the role of D66 in the binding of metal ions at the active site of Rv0805 [11].

The activity improvements in both N53D and N53A mutants of GpdQ are likely due to the abolition of the interaction with S268, which promotes oligomeric dissociation. This N53-S268 inter-chain hydrogen bond does not exist in the Rv0805 protein.

5.5 TRUNCATION AT THE CAP DOMAIN OF GpdQ

Two of the mutations derived from directed evolution, namely C54G and G259R presumably resulted in a change in the orientation of the cap domain of GpdQ. We concluded that increasing the flexibility of this domain would

enhance both the catalytic activity and substrate affinity of the enzyme. The improvement would be manifested especially when working with non-natural substrates such as b*p*NPP, whose chemical structure is very different from glycerophosphodiesters, the putative physiological substrates.

5.5.1 Y257stop

To this end, the Y257stop mutant was designed by converting the tyrosine residue into a stop codon. This would produce a truncated GpdQ mutant without an expressed cap domain, with a calculated molecular weight of 28.7 kDa per monomer and an extinction coefficient of 31400 $M^{-1}cm^{-1}$. However, the difference in molecular weight could not be shown clearly in the SDS-PAGE gel below, probably due to the limitations of the resolving gel (Figure 5.13).

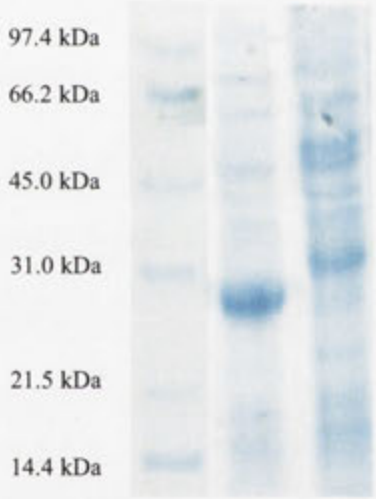


Figure 5.13: SDS-PAGE gel showing the expressed Y257stop compared to N53A.

A gel filtration analysis of Y257stop showed that the enzyme had dissociated into a mixture of hexamer and dimer (Figure 5.2). The assay results of the hexamer-dimer mixture are presented in Table 5.4.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
C54G	13.74 ± 0.2	1.83	0.43 ± 0.04	0.20	3.21×10^4	8.99
G259R	11.25 ± 0.2	1.50	0.66 ± 0.07	0.31	1.70×10^4	4.76
C269S	47.68 ± 1.3	6.36	1.21 ± 0.17	0.58	3.93×10^4	11.01
C269A	36.02 ± 1.4	4.80	1.48 ± 0.27	0.70	2.43×10^4	6.81
Y257stop	72.42 ± 1.8	9.66	1.45 ± 0.11	0.69	5.00×10^4	14.01

Table 5.5: Kinetic parameters of Y257stop.
Kinetic parameters of the wild type and four other single site mutants that sustained changes to the changes to the cap domain are also given for comparison.

Without the cap domain, GpdQ was not only still functional but the k_{cat} was almost 10-times higher. It is most likely that this remarkable activity improvement was attributed to the high level of dissociation of the enzyme (as shown in Figure 5.2) caused by the complete removal of the cap domain. Surprisingly though, the Y257stop mutation did not improve the substrate affinity greatly.

5.5.2 C54G-Y221H-Y257stop

A GpdQ variant with mutations C54G, Y221H and Y257stop was constructed via site-directed mutagenesis. This was achieved with designed Y257stop forward, reverse primers and the plasmid DNA of mutant 3-25 (C54G, Y221H and G259R) as template during SDM-PCR. The two mutations, Y257stop and G259R are mutually exclusive: the presence of Y257stop will not allow G259R to be expressed due to a premature end of translation; an expressed G259R would indicate the absence of Y257stop.

During the purification process, it was found that the combination of the three mutations had extended the dissociation of GpdQ further to a mixture of hexamers, dimers and monomers, as shown in Figure 5.14.

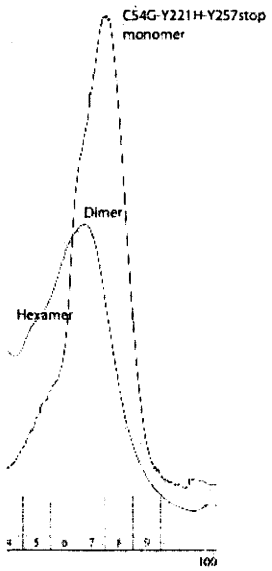


Figure 5.14: Dissociation of GpdQ into dimers and monomers.
Superdex 200 migration profiles of mutants C54G-Y221H-Y257stop (blue) and 3-25 (magenta).

Kinetic results of C54G-Y221H-Y257stop assayed with b*p*NPP are given in the table below.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
Hexamer	1.31 ± 0.1	0.18	0.31 ± 0.03	0.15	4.19×10^3	1.17
Dimer	65.62 ± 0.9	8.75	0.25 ± 0.02	0.12	2.66×10^5	74.51
Monomer	202.45 ± 3.8	26.99	0.29 ± 0.03	0.14	6.98×10^5	195.52

Table 5.6: Kinetic results of hexameric, dimeric and monomeric forms of C54G-Y221H-Y257stop.

As the C54G-Y221H-Y257stop enzyme dissociated from hexamer into dimer then monomer, the catalytic activity also increased accordingly. However, the different oligomeric forms did not appear to affect the K_m. The low K_M values of these three forms of the enzyme are also probably attributed to the other two mutations, C54G and Y221H as Table 5.5 showed that Y257stop alone did not improve the affinity much.

The k_{cat}/K_M ration of 3-25 dimer was about 160 (Table 5.3), which was more than twice higher than that of C54G-Y221H-Y257stop dimer. This would imply that the Y257stop mutation was not compatible with C54G and Y221H.

Surprisingly, the monomeric enzyme was the most catalytic reactive form of the enzyme. Based on the structure of the wild-type enzyme, where two domains contribute to dimerisation, one would expect the GpdQ dimer to be extremely stable and most catalytically efficient.

5.5.3 Conclusions on the cap domain of GpdQ

Exploring the role of the cap domain in GpdQ through Y257stop mutants led to several conclusions. Firstly, this short peptide serves as a mechanism for enhancing the substrate specificity of GpdQ by minimising the entrance space into the active site. In the absence of the cap domain in Y257stop, the active site was made more accessible, GpdQ was not only still functional but was significantly more active.

Secondly, the cap domain is also responsible for stabilising the oligomeric structure of GpdQ. Its role in holding two monomers together as a single dimeric unit can be seen and understood more directly. However, as the hexamer is very much less than the dimer, the single Y257stop mutation resulted in the hexamer → dimer dissociation but not dimer → monomer dissociation. The residues in the cap domain must have weak interactions with another dimer to stabilise the hexamer.

With the synergy of the three mutations in C54G-Y221H-Y257stop, the monomeric GpdQ was acquired. It cannot be ascertained though, whether a hexamer breaks down into monomers directly, or via dimer intermediates, or via both mechanisms.

5.6 TRUNCATION AT THE DIMERISATION DOMAIN OF GpdQ

Y204 was changed to a stop codon to truncate the protein prior to the dimerisation domain. This was done to investigate if GpdQ could function with only the catalytic domain.

Y204stop had a calculated molecular weight of 22.7 kDa. Y204stop had a very high level of expression and the expressed protein band was estimated to be ~22 kDa. However, the protein was insoluble and was challenging to purify. Most purification protocols of insoluble proteins involve the use of denaturing detergents like sodium dodecyl sulfate (SDS) [13]. The poor presentation of the Y204stop protein in the SDS-PAGE gel below (Figure 5.15) was due to the overloading of the samples.

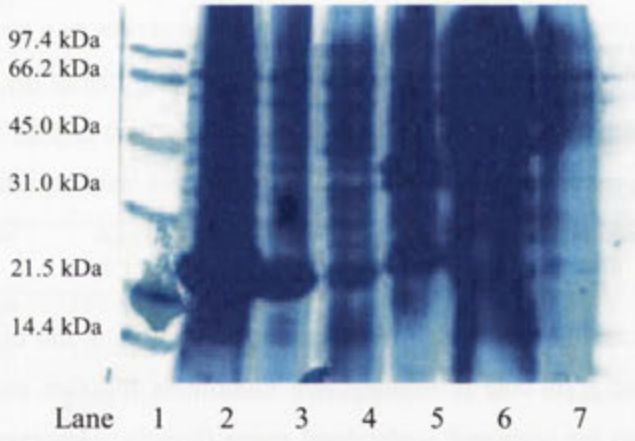


Figure 5.15: The insolubility of GpdQ mutant Y204stop.
Lane 1 shows the marker while lanes 2-7 show duplicates of Y204stop crude lysate (Lanes 2 and 3), insoluble fraction (Lanes 4 and 5) and soluble fraction (Lanes 6 and 7) respectively.

At the start of the purification, a drop assay with b*p*NPP was performed on the crude lysate, soluble and insoluble fractions of Y204stop and no activity was detected (Figure 5.16). Two attempts were made to purify the very little amount of Y204stop in the soluble fraction. However, due to the extremely low concentration, the exact fractions containing the eluted protein could not be determined. The inactivity and the insolubility of Y204stop indicate that the protein did not fold properly and therefore was catalytically inert.

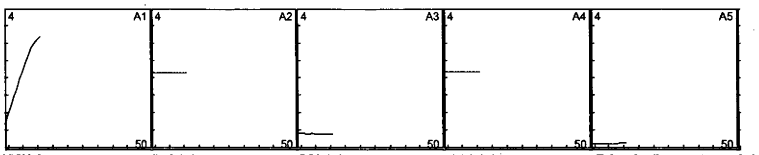


Figure 5.16: Drop assay of GpdQ mutant Y204stop.
Drop assay of Y204stop crude lysate, soluble and insoluble fractions with b*p*NPP, flanked by positive (wild-type GpdQ) and negative control experiments (b*p*NPP only).

5.7 F21

5.7.1 F21K

Directed evolution of GpdQ towards higher catalytic activity and substrate affinity simultaneously showed that dissociation of the enzyme was a key factor in reaching such a target, through mutations such as C54G, Y221H and G259R. However, the hexamer-dimer equilibrium made purification and kinetic assays more complicated and challenging. Wang proposed a method for determining kinetic parameters of dissociating enzyme systems that involved too many complex equations that are based on theoretical examples [14]. The very few citations of the publication would reflect the lack of user friendliness and the unpopularity of the mathematical equations. Moreover, the rapidly equilibrating system of GpdQ dissociating enzymes would make them nearly impossible to crystallise.

To this end, the GpdQ structure was analysed again by Dr Paul Carr to search for mutation(s) that would directly destabilise the hexameric interface so that a pure GpdQ dimer mutant enzyme could be acquired. As a result, two mutations, namely F21K and W261R were recommended.

F21 (of chain A) and W261 (of chain B) of the same dimer unit form hydrophobic contacts with the aliphatic parts of the side chains of H256 and D30 (of chain D that belongs to another dimer unit) that form a salt link with each other (Figures 5.17 and 5.18). In the absence of F21 and W261,

the side chains of H256 and D30 would be comfortable in a hydrophilic environment and this would favour the formation of a stable dimer.

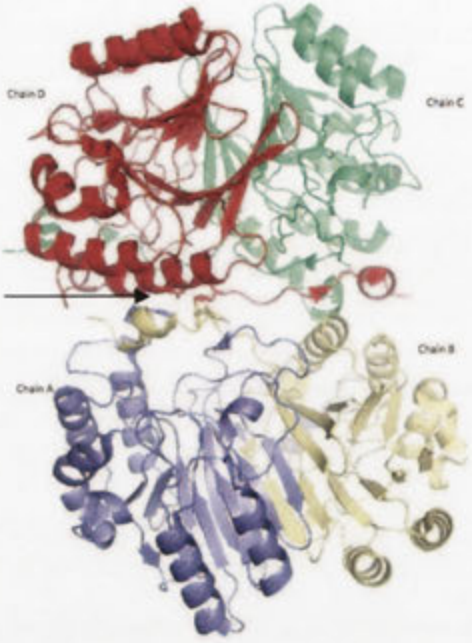


Figure 5.17: **The inter-chain hydrophobic patch in GpdQ.**
The hydrophobic patch, indicated by the black arrow is formed by interactions among F21, D30, H256 and W261 from different chains.

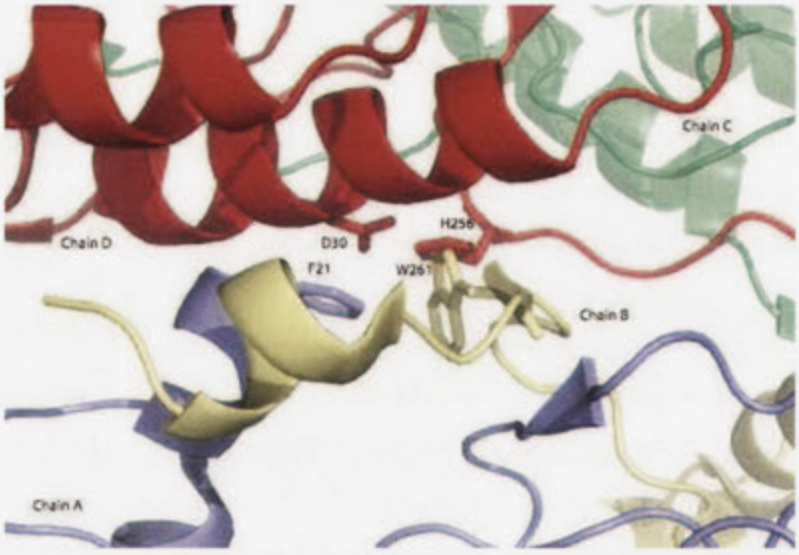


Figure 5.18: **Detailed view of the inter-chain hydrophobic patch in GpdQ.**
An up close view of the interactions among F21 (chain A, blue), W261 (chain B, yellow), H256 and D30 (chain D, red). Chains C and E have been omitted from this diagram for clarity reason.

The Superdex 200 elution profile of F21K in Figure 5.19 below shows that this mutant enzyme is about 90% dimeric.

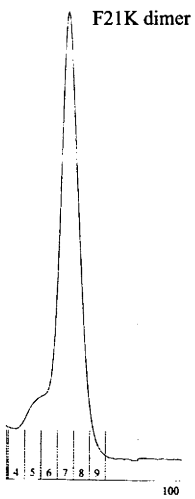


Figure 5.19: The F21K GpdQ dimer.
Superdex 200 migration profile showing the elution of dimeric F21K.

Table 5.7 presents the kinetic parameters of F21K.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
F21K	45.75 ± 0.1	6.10	0.33 ± 0.04	0.16	1.40×10^5	39.22

Table 5.7: Kinetic properties of F21K relative to the wild type GpdQ.

F21K emerged as the best of all the single mutants that had been characterised in this study. This once again underlined the importance of rational design to complement directed evolution in protein engineering, as a mutation from F21 to a lysine (TTT → AAG) involved three nucleotide changes that would not occur through random mutagenesis. It should be mentioned that an F12L mutation (involving one nucleotide change) was detected in one of the mutants after the second round of screening in directed evolution (Tables 4.6 and 4.7).

5.7.2 F21K-C54G-Y221H-Y257stop

The quadruple mutant was made via site-directed mutagenesis using F21K forward and reverse primers to introduce the mutation to the C54G-Y221H-Y257stop plasmid DNA template described in section 5.4.2.

The quadruple mutant dissociated into a mixture of hexamer, dimer and monomer. The hexamer, based on previous examples of dissociating GpdQ enzymes, was presumed to be inactive and was therefore not assayed. We focused on characterising the dimer and monomer separately to determine which form of the enzyme was more catalytically active. Assay results are given in Table 5.8.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
Dimer	13.29 ± 0.3	1.77	1.39 ± 0.01	0.66	9.55×10^3	2.68
Monomer	56.65 ± 1.0	7.55	0.18 ± 0.02	0.09	3.22×10^5	90.19

Table 5.8: Kinetic properties of the dimeric and monomeric forms of quadruple mutant F21K-C54G-Y221H-Y257stop.

Comparing the assay results presented in Tables 5.6 and 5.9, it can be inferred that F21K was not compatible with the other three mutations, as the k_{cat}/K_M ratio of the triple mutant, in both dimeric and monomeric forms, is significantly higher than that of the quadruple mutant with an additional F21K mutation. However, the two sets of kinetic results consistently showed that the monomeric enzymes displayed higher b*p*NPP activity than the dimers.

5.7.3 The preliminary crystallisation of F21K

Attempts were made to crystallise F21K. As the novel structure of the GpdQ dimer is different from the native hexamer, previous crystallisation

conditions as reported in section 5.3 of this chapter were no longer suitable. New crystallisation conditions were sought for the GpdQ dimer, as described in section 2.9.2 in Chapter 2.

Of a total of 576 conditions screened (96 from each of the six different screens named in Chapter 2), crystals were formed only under one condition: 0.5 M ammonium sulfate (salt), 0.1 M sodium citrate tribasic dihydrate pH 5.6 (buffer), 1.0 M lithium sulfate monohydrate (precipitant), which was reagent F3 of the Hampton Research Crystal Screen HT. Crystals are shown in Figure 5.20.

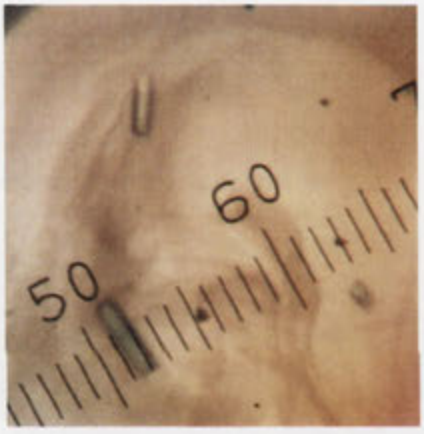


Figure 5.20: GpdQ F21K crystals.

Attempts to reproduce the crystals by using the hanging drop vapour diffusion method were unsuccessful. The crystallisation conditions were optimised by varying the pH values (5.0 to 6.5), and the lithium sulfate concentration (0.5 to 1.5 M).

5.8 W261R

The mutant W261R was constructed as another effort to acquire a pure GpdQ dimer. Interestingly, this mutant enzyme remained associated as a hexamer. The kinetic results of this mutant are given below.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
W261R	12.41 ± 0.2	1.65	0.71 ± 0.06	0.34	1.74×10^4	4.87

Table 5.9: Kinetic results of W261R.

The W261R mutation improved both the k_{cat} and K_m slightly but it did not cause GpdQ to dissociate as expected.

5.9 CONCLUSIONS

Rationally-designed GpdQ mutants with improved activity

A total of six single site mutants (C269S, C269A, N53D, N53A, F21K, W261R) and two truncation mutants (Y257stop and Y204stop) were designed in this chapter. Y204stop was an insoluble and completely inactive enzyme. Apart from Y204stop, the other mutants showed improved activity with substrate bis (*p*-nitrophenyl) phosphate (b*p*NPP). F21K was the best single site mutant to be produced in this thesis, with an almost 40-fold improvement in the overall k_{cat}/K_m ratio. Apart from W261R, all other mutations resulted in oligomeric dissociation in GpdQ.

The DNA samples of C269S, C269A, N53D, N53A, F21K and Y257stop were included as templates for *in vitro* gene recombination via staggered extension process (StEP) in the second part of directed evolution of GpdQ (Chapter 6).

High activity associated with simpler oligomeric forms of GpdQ

Four multiple site mutants of GpdQ were also designed in this chapter:

1. C54G-Y221H-G259R-C269S (3-25-C269S)
2. C54G-Y221H-G259R-C269A (3-25-C269A)
3. C54G-Y221H-Y257stop
4. F21K-C54G-Y221H-Y257stop

Mutants 3-25-C269S and 3-25-C269A had dissociated into mixtures of hexamers and dimers. When assayed separately, it was found that high activity was associated with the dimers of both enzymes.

Mutants C54G-Y221H-Y257stop and F21K-C54G-Y221H-Y257stop had dissociated further into hexamers (very small amounts), dimers and monomers. For both mutants, the monomers were the most catalytically active form of the enzymes. Therefore, these four multiple site enzymes showed a consistent pattern, where simpler oligomeric forms of GpdQ were favourable towards enzyme activity.

Rational design of proteins – fast and precise

Another highlight of this chapter is the ease and the precision of rational design in engineering very active and/or dissociating GpdQ enzymes. Many of the mutations made in this chapter involved two or three nucleotide changes within a codon, which were made possible only through site-directed mutagenesis.

CHAPTER FIVE REFERENCES

1. Frieden, C. (1967) Treatment of enzyme kinetic data II. The multisite case - comparison of allosteric models and a possible new mechanism. *Journal of Biological Chemistry 242*, 4045-4052.
2. Frieden, C. (1970) Kinetic aspects of regulation of metabolic processes - hysteretic enzyme concept. *Journal of Biological Chemistry 245*, 5788-5799.
3. Frieden, C. (1979) Slow transitions and hysteretic behavior in enzymes. *Annual Review of Biochemistry 48*, 471-489.
4. Neet, K. E., and Ainslie, G. R., Jr. (1980) Hysteretic enzymes. *Methods in Enzymology 64*, 192-226.
5. Weiss, J. N. (1997) The Hill equation revisited: uses and misuses. *Faseb Journal 11*, 835-841.
6. Cornish-Bowden, A. (2004) Fundamentals of enzyme kinetics, 3rd edition. *Portland Press.*
7. Goodrich, J. A., and Kugel, J. F. (2007) Binding and kinetics for molecular biologists. *Cold Spring Harbor Laboratory Press.*
8. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.
9. Mertz, P., Yu, L., Sikkink, R., and Rusnak, F. (1997) Kinetic and spectroscopic analyses of mutants of a conserved histidine in the metallophosphatases calcineurin and lambda protein phosphatase (vol 272, pg 21296, 1997). *Journal of Biological Chemistry 272*, 30589-30589.
10. Shenoy, A. R., Sreenath, N., Podobnik, M., Kovacevic, M., and Visweswariah, S. S. (2005) The Rv0805 gene from *Mycobacterium tuberculosis* encodes a 3',5 '-cyclic nucleotide phosphodiesterase: Biochemical and mutational analysis. *Biochemistry 44*, 15695-15704.
11. Shenoy, A. R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S. S., and Podobnik, M. (2007) Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis*. *Journal of Molecular Biology 365*, 211-225.
12. Zhuo, S., Clemens, J. C., Stone, R. L., and Dixon, J. E. (1994) Mutational analysis of a Ser/Thr phosphatase. Identification of residues important in phosphoesterase substrate binding and catalysis. *Journal of Biological Chemistry 269*, 26234-26238.
13. Scopes, R. K. (1994) Protein purification: Principles and methods, 3rd edition. *Springer-Verlag.*
14. Wang, Z. X. (1998) A novel method for determining kinetic parameters of dissociating enzyme systems. *Analytical Biochemistry 264*, 8-21.

6

DIRECTED EVOLUTION OF GpdQ (PART TWO)

CHAPTER SIX: DIRECTED EVOLUTION OF GpdQ (PART TWO)

6.1 INTRODUCTION

This chapter is a chronicle of experiments conducted as a continuation of the directed evolution studies presented and discussed in Chapter 4.

One of the two most important outcomes from the first four rounds of directed evolution in Chapter 4 was the emergence of 3-25 (C54G, Y221H, C259R) as the best mutant. The overall k_{cat}/K_m ratio of mutant 3-25 was two orders of magnitude higher than that of the wild-type GpdQ. The other major outcome was the discovery of oligomeric dissociation in mutant 3-25. However, due to the premature convergence encountered, directed evolution experiments were stopped temporarily after four rounds to review and revise the techniques that were being used. Several rationally designed GpdQ mutants were made with the aim of improving catalytic activity through the dissociation of the GpdQ hexamer (Chapter 5). Directed evolution now proceeded with the following strategies:

1. Random mutagenesis

To give the evolutionary work a fresh start, GpdQ was subject to random mutagenic amplification using the previously described epPCR method to generate a new 5th round library of mutants.

2. Changed StEP protocol

The StEP-A protocol was replaced by StEP-B (used for rounds 6-7) and later StEP-D (used for round 8) programs that were able to shuffle the homologous template genes better. Both StEP-B and StEP-D protocols were two-step PCR cycles that lacked the 72°C extension step.

3. The use of rational design to complement directed evolution

As discussed in Chapter 4, directed evolution of GpdQ led to the discovery of oligomeric dissociation in mutant 3-25. This phenomenon was unlikely to be artefactual as structure-function studies of C54G, Y221H and G259R supported the notion that these mutations would disrupt the interface between dimers within the GpdQ hexamer. Although the other improved mutants from the fourth round of evolution were not subject to gel filtration analysis to confirm dissociation, they were presumed to have also dissociated as their genotypes were made up of the same mutations as mutant 3-25, with the exception of H217R which was not present in 3-25. The structure of GpdQ was studied again to identify mutations that would stabilise the GpdQ dimer or at least, shift the hexamer-dimer equilibrium towards dissociation. As a result, several rationally designed GpdQ mutants were produced and they were included as DNA templates for the StEP PCR reactions. These mutants were F21K, N53D, N53A, Y257stop, C269S and C269A. The production of these site-specific mutants and their mutagenesis studies were presented in Chapter 5.

4. Changed substrate concentration during screening

The 5 mM substrate concentration used to screen the first four generations of mutants during high activity evolution proved to be too high, which ultimately led to the selection of many false-positive mutants, as shown in Table 4.8. On the other hand, while trying to improve b*p*NPP substrate affinity, concentrations ranging from 0.25-1 mM were used to screen the mutants. The effects of the four mutations responsible for enhancing the affinity, namely C54G, H217R, Y221H and G259R, were reproducible (Tables 4.4 and 4.5) and not artefactual when the proteins were expressed, purified and characterised individually.

Based on all the mutations obtained from both the high affinity and the high activity evolutions, it appeared that any improvement in activity would also automatically increase the affinity (Tables 4.4 and 4.5) but not vice versa

(Table 4.8). This would mean that screening with low concentrations of b*p*NPP is no longer necessary. Hence, the screening b*p*NPP concentration was fixed at 1 mM from round 5 onwards.

6.2 DIRECTED EVOLUTION OF GpdQ (ROUNDS 5-8)

6.2.1 Screening results

The selected, best mutants from rounds 5-8 and other templates included in StEP PCRs are listed in the Table 4.11. Table 4.12, shows detailed information about the genotypes of the 14 best mutants selected, and their activity improvements.

Table 6.1: Directed evolution of GpdQ (Part Two).
Up to 30 mutants were selected at each round and the best 15 were sequenced. The mutants listed above are not necessarily the very fittest mutants of their respective generations, but they best represent all the mutations that were retained in each round. The average specific activity improvements at 1 mM b*p*NPP, with reference to the wild type, are given in parentheses. The green shading means a particular mutation had been introduced into the mutant libraries via rational design at a specific generation (*eg.* C269A was introduced as one of the DNA templates for StEP at the sixth round). Only mutants that were retained through later generations are shown here. The yellow shading indicates a mutation occured via random mutagenesis during StEP.

	R12L	F21K	C54G	N103D	S127A	H217R	Y221H	Y257 stop	G259R	C269A	Activity Improvement
8-1				√	√	√			√		15.3
8-2	√			√		√				√	15.2
8-3					√	√				√	14.5
8-4						√				√	13.6
8-5			√	√	√	√			√		13.4
8-6					√		√			√	12.5
8-7		√	√			√				√	12.4
8-8						√		√			12.2
8-9		√				√			√		11.7
8-10			√	√	√	√			√		11.6
8-11		√	√			√			√		11.6
8-12	√			√						√	11.4
8-13			√	√	√	√			√		11.3
8-14		√				√			√		11.1
f	2	5	5	6	6	12	1	1	7	6	
Percentage of *f* (%)	14.3	35.7	35.7	42.9	42.9	85.7	7.1	7.1	50.0	42.9	

Table 6.2: An elucidation of the genotypes and activity improvement (relative to the wild-type) of the 14 best mutants from the 8th round of directed evolution.

The frequencies and relative frequencies of the mutations were also calculated. 15 mutants were sent for sequencing altogether. One sample did not yield sequencing results, thus explaining why only 14 mutants are presented in this table.

	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
8-1 Dimer	70.56 ± 0.9	9.41	0.76 ± 0.01	0.36	9.19×10^4	25.74
8-1 Monomer	111.61 ± 1.3	14.88	0.36 ± 0.09	0.17	3.08×10^5	86.27
8-2 Dimer	141.52 ± 0.5	18.87	0.24 ± 0.01	0.11	5.85×10^5	163.86
8-2 Monomer	70.09 ± 0.1	9.35	0.07 ± 0.001	0.03	1.02×10^6	285.71
8-3 Dimer	138.22 ± 0.7	18.43	0.07 ± 0.002	0.03	1.98×10^6	551.82
8-5 Dimer	162.60 ± 0.8	21.68	0.11 ± 0.005	0.05	1.47×10^6	411.76
8-7 Dimer	189.59 ± 2.1	25.28	0.11 ± 0.008	0.05	1.78×10^6	495.60

Table 6.3: Kinetic parameters of the five characterised eighth round mutants.

All the five GpdQ mutants from the eighth round of directed evolution experienced dissociation, at varying levels. There was still a negligible amount of the hexameric form in all the five enzymes, but they were presumed to be significantly less active than the other dissociated forms and therefore not characterised. Mutants 8-1 and 8-2 had dissociated further into monomers; the other three, namely 8-3, 8-5 and 8-7, had not.

With mutants 8-1 and 8-2, the monomers were more catalytically active than their dimers. However, the 8-3 dimer displayed the greatest k_{cat}/K_M improvement compared to the wild type. The relative values of k_{cat}/K_M of 8-5 and 8-7 dimers were also greater than those of 8-1 and 8-2 monomers. Based on the kinetic results of these five eighth round mutants, the GpdQ dimer appeared to be the most optimised form of the enzyme, displaying the highest values of both k_{cat} and K_M. The three mutant dimers, 8-3, 8-5 and 8-7, represent the most catalytically active GpdQ enzymes reported in this thesis, with k_{cat}/K_M values as high as 10^6 $M^{-1}s^{-1}$, an enhancement of three orders of magnitude upon the wild-type GpdQ.

R12L, N103D and S127A were three novel mutations that were not found in the first rounds of directed evolution. These three single site mutants were

made to assess each of the mutations individually. The kinetic results are given below.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_M (mM)	Relative K_M	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_M
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
R12L	32.21 ± 0.5	4.29	0.32 ± 0.03	0.15	1.00×10^5	28.01
N103D	5.29 ± 0.1	0.71	1.04 ± 0.09	0.50	5.10×10^3	1.43
S127A	5.12 ± 0.1	0.68	1.22 ± 0.07	0.58	4.21×10^3	1.18

Table 6.4: Kinetic results of single site mutants R12L, N103D and S127A compared to the wild type.

The kinetic results above show that with only the R12L mutation, the k_{cat}/K_M of GpdQ sustained a marked improvement. With N103D and S127A however, although both mutations gave slight K_M improvements, their k_{cat} values were lower than that of the wild type. As these two mutations 'survived' not one but at least three rounds of screening (from Round 5 onwards, table 6.2), it was unlikely that the conservation was artefactual. These two mutations probably had to combine with others to have a synergistic effect on the enzyme activity.

It is also noteworthy to mention that all the three enzymes in Table 6.4 appeared to remain associated as hexamers, as indicated by size exclusion chromatography and that none of the mutations seemed to have affected the oligomeric structure of GpdQ.

6.3 STRUCTURE-FUNCTION STUDIES OF MUTATIONS R12L, N103D AND S127A

6.3.1 R12L

R12L is a mutation that often co-existed with C269A, for example mutants 7-4 and 8-2 in Table 6.3. Residues 54-59, containing a glycine (G55) and a proline (P57), form a hydrophobic pocket as illustrated in Figure 6.1 below. In the presence of the mutation C269A, the free thiol group of C54 would increase the hydrophobicity of the region. An R12L mutation would have a stabilising effect over the hydrophobic area, as leucine is non-polar. This mutation would also have but probably less significance if the disulfide bond was abolished through the C54G mutation instead, since the hydrophobicity of a glycine would not be as strong as that of a free cysteine.

Additionally, as shown in Figure 6.1, R12 forms an inter-chain salt link with E273 in the cap domain (2.77 Å in distance). Jackson *et al.* did not report this salt link as the last three residues at the C-terminus (residues 271-274) were missing from the two crystal structures (2DXN and 2DXL) of GpdQ [6]. The structure 3D03 solved as part of this project had the complete 274 residues. An R12L mutation would abolish the salt bridge that would ultimately, like mutations C54G, G259R and C269S/A, disrupt the stability of the hexameric interface. However, the hexameric structure of the R12L enzyme seemed to be intact. Dissociation of the hexamer, if there was any, was minimal (Figure 6.2).

Another interaction formed by residue R12 is a 2.80-Å intra-chain salt link with E59. This interaction would be missing in the presence of the R12L mutation.

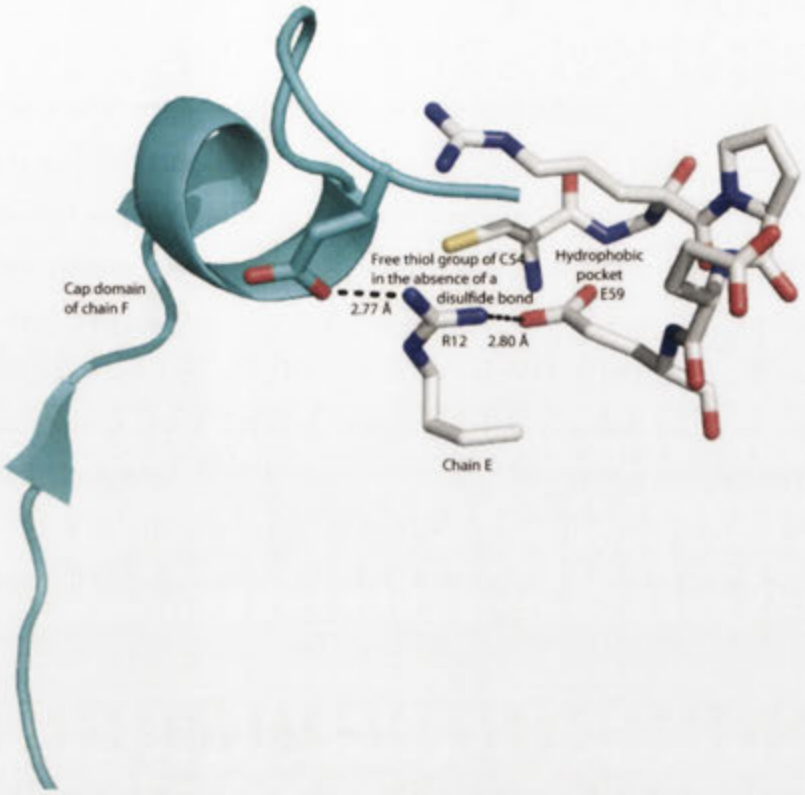


Figure 6.1: The hydrophobic pocket constituted by residues C54-E59 and the two salt link interactions formed by R12.

Figure 6.2: An excerpt of the Superdex 200 elution profile of R12L.
There was no obvious indication of dissociation of the hexamer in R12L.

6.3.2 N103D

N103 is a surface residue whose acyl group (of the amide functional group) interacts directly with the exogenous solvent environment, as shown in Figure 6.3. Although the nitrogen atom of the amide group is buried within the protein molecule, it does not appear to have any interactions with other residues. This nitrogen atom is 5.33 Å from K84 (Figure 6.4). When mutated to an aspartic acid, the residue could possibly form a salt link with K84, since the side chain of the lysine is flexible and could change its orientation and turn towards the acid residue (D103). Such a salt link could have a stabilising effect on the protein structure. The lack of activity improvement in the N103D mutant (Table 6.4) also supported this notion.

Figure 6.3: N103 – a surface residue in GpdQ.
A diagram with N103 coloured in blue to highlight their locations on the surface of the GpdQ hexamer molecule.

Figure 6.4: A close-up diagram of N103 and K84 on a GpdQ monomer.

6.3.3 S127A

The S127A mutation was retained after four rounds of screening (Rounds 5-8) (Table 6.2). At the end of the eighth round of evolution, the mutation was found to be present in 6 of the 14 best mutants were sequenced.

As illustrated in Figure 6.5, the side chain S127 is 3.62 Å away from N80 in chain A, while in other subunits the distance varies from 3.39-3.78 Å. N80 is the ligand responsible for the low metal binding affinity at the β-site. Despite the proximity, a hydrogen bond interaction between S127 and N80 was unlikely because of the conformation of S127, whose side chain hydroxyl group points away from the amide group of N80. However an alanine substitution at this position might allow N80 to move flexibly, thus enhancing the ability of the residue to bind and release the β-metal more easily. Kinetic assay of the S127A enzyme (Table 6.4) indicated no

improvement upon the catalytic turnover (k_{cat}) and a mild improvement on the substrate affinity (K_M).

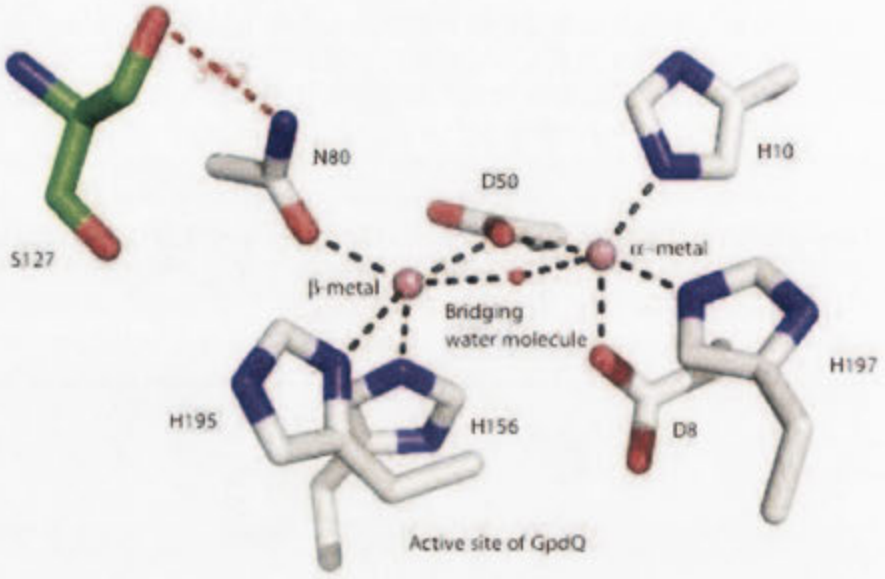


Figure 6.5: Location of S127 at the periphery of the GpdQ active site.

6.4 CONCLUSIONS ON DIRECTED EVOLUTION OF GpdQ (PART TWO)

With improved epPCR, StEP protocols and screening strategies, the second part of GpdQ evolution, which was constituted by the fifth to eighth rounds, successfully produced three GpdQ enzymes whose overall k_{cat}/K_M values had reached the realm of 10^6 $M^{-1}s^{-1}$, which was an improvement of three orders of magnitude compared to the wild type. The high activity of these three mutants, namely 8-3, 8-5 and 8-7, was associated with the dimeric form.

Mutants 8-1 and 8-2 dissociated into an equilibrium of three oligomeric forms: hexamer, dimer and monomer. The monomeric forms of mutants 8-1 and 8-2 were more active than their dimeric counterparts, but their activity was still significantly lower than the activity of 8-3, 8-5 and 8-7 dimeric

enzymes (Table 6.3). Therefore, it appeared that the GpdQ dimer was the oligomeric form that was best optimised for b*p*NPP hydrolytic activity.

At the end of each round of screening, the DNA sequences of the best 15 mutants were analysed. Once again, as stated in Chapter 4, mutations of any active residue or other conserved residues were never observed. Low mutation rates of epPCR and StEP reactions were unlikely to be responsible for this observation. A more probable explanation would be that any mutation of these conserved residues would result in drastic deleterious effect on the enzyme activity and so these mutants were eliminated during screening, through a process called 'purifying selection' by Hughes [7, 8]. However, mutations of residues that have interactions with active site residues were observed. Examples of such mutations are N53D, S127A and H217R. N53D was a rationally designed GpdQ mutant that was introduced into the StEP reaction at the sixth round. The mutation was retained after the sixth round screening but not later.

Much of the success of the second part of GpdQ directed evolution was also attributed to the inclusion of several rationally designed mutants as templates for gene recombination during the process of generating libraries for screening. The most notable among these mutations were F21K and C269A as they were retained at the end of eighth round screening among the mutants that displayed the highest activity improvement (Table 6.2). The availability of GpdQ crystal structures facilitated the possibility of rational approaches to engineer the enzyme activity. A good understanding of the protein structure and function greatly expanded the capabilities of GpdQ enzyme engineering.

CHAPTER SIX REFERENCES

1. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology 16*, 258-261.
2. Huynen, M. A., Stadler, P. F., and Fontana, W. (1996) Smoothness within ruggedness: The role of neutrality in adaptation. *Proceedings of the National Academy of Sciences of the United States of America 93*, 397-401.
3. Bloom, J. D., and Arnold, F. H. (2009) In the light of directed evolution: Pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America 106*, 9995-10000.
4. Bershtein, S., Goldin, K., and Tawfik, D. S. (2008) Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of Molecular Biology 379*, 1029-1044.
5. Gupta, R. D., and Tawfik, D. S. (2008) Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods 5*, 939-942.
6. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.
7. Hughes, A. L. (1999) Adaptive evolution of genes and genomes. *Oxford University Press.*
8. Hughes, A. L. (2007) Looking for Darwin in all the wrong places: The misguided quest for positive selection at the nucleotide sequence level. *Heredity 99*, 364-373.

7

FINAL CONCLUSIONS AND FUTURE DIRECTIONS

CHAPTER SEVEN – FINAL CONCLUSIONS AND FUTURE DIRECTIONS

7.1 MUTATIONAL ANALYSIS OF ACTIVE SITE RESIDUES AND MECHANISTIC STUDIES OF GpdQ

Perhaps the most significant outcome of the mutagenesis studies of GpdQ active site residues was the confirmation that the α- and β-metal ions bind with varying affinities. Jackson *et al.* first proposed the difference in the metal-binding affinities when the Bijvoet-difference Fourier map of the active site of the Zn^{2+}-GpdQ holoenzyme exhibited nearly full and partial occupancies of the α- and β-sites with Zn^{2+} ions [1].

In this study, the hypothesis was investigated further with the D8N/A and N80D/A mutants. The residues D8 and N80 were believed to be responsible for the non-identical binding affinities at the metal sites. The N80D mutation that was expected to increase the binding affinity at the β-site produced an inactive GpdQ enzyme; while N80A was still catalytically active. These kinetic findings indicate that flexibility in metal binding at the β-site was essential for GpdQ's diesterase activity. Meanwhile, both the D8N and D8A mutants were completely inactive, thus highlighting the importance of having a strong metal-binding affinity at the α-site.

Electron paramagnetic resonance (EPR) was employed by fellow collaborators on GpdQ at the University of Queensland to monitor the binding of Mn^{2+} to the GpdQ apoenzyme [2]. They also performed magnetic circular dichroism (MCD) studies on the wild-type GpdQ, N80D and N80A enzymes [3]. All these experiments confirmed the non-identical binding affinities at the α- and β-sites. This provided an insight into the possible catalytic mechanism of GpdQ: The α-

metal is responsible for the activation of the nucleophile (a catalytic water molecule, OH^-) that catalyses the cleavage of the phosphodiester bond while the β-metal is associated with substrate binding.

7.2 ENGINEERING GpdQ's BIS (*P*-NITROPHENYL) PHOSPHATE ACTIVITY WITH DIRECTED EVOLUTION AND RATIONAL DESIGN – A SEMI-RATIONAL APPROACH

Combining both directed evolution and rational strategies in what is sometimes referred to as a 'semi-rational' approach in protein engineering, this study successfully developed GpdQ mutants that were ~500 times (k_{cat}/K_M) more active than the wild type towards b*p*NPP. The k_{cat}/K_m ratios of some of the eighth round mutants had even entered the realm of 10^6 $M^{-1}s^{-1}$ which are two orders of magnitude below diffusion-limited catalytic rates (Table 6.3, Chapter 6).

Rational design played a major role in the success of the engineering of GpdQ's bis (p-nitrophenyl) phosphate (b*p*NPP) activity. At the end of the first part of directed evolution studies, the take-home message was that apart from the improvement upon the catalytic activity, the selection pressure was also driving GpdQ towards oligomeric dissociation. Throughout the many kinetic assays conducted in this project, it had been shown that the high activity was associated with simpler oligomeric forms (dimer and monomer). This prompted the structural analysis of GpdQ to identify mutations that could be made to the protein to design a stable dimer or monomer. These mutations were introduced into the second part of directed evolution studies. Two of these rationally-designed mutations, F21K and C269A, were retained in many of the most catalytically active eighth round variants (Table 6.2, Chapter 6).

Many of the mutations (most notably R12L, C54G, G259R and C269A) had similar implications on the structure of GpdQ: Conformational changes to the cap domain that would lead to increased accessibility of the active site and also ultimately, oligomeric dissociation.

It is believed that oligomerisation in proteins happens as a result of the natural evolutionary process [4, 5]. The conservation of the catalytic domains of GpdQ and the other members of the α/β sandwich superfamily of dimetallophosphohydrolases lends support to the notion that these enzymes must have diverged from the same monomeric ancestor and each has evolved secondary domains to cater for the physiological needs of its host organism. In the case of GpdQ, the dimerisation and cap domains were likely to have developed as a result of divergent evolution. The formation of the GpdQ hexamer, with the cap located at the entrance of the active was probably crucial in maintaining the enzyme's selectivity for its natural substrates – the glycerophosphodiesters. A figure showing how the cap domain overhangs the active site is given in Figure 7.1.

b*p*NPP contains two large aromatic ester groups (*p*-nitrophenolate) and has a much bulkier chemical structure than glycerophosphodiesters (Figures 1.5 and 1.8, Chapter 1). It was therefore understandable that a widened entry into the active site to diminish steric entrance was required for bulky substrate molecules.

Engineering GpdQ's b*p*NPP activity may lack direct industrial and clinical applications but helps us to understand how enzymes evolve and novel functions in nature.

Here is a list of proposed follow-up experiments on the mutants with improved b*p*NPP activity in this study:

i. The activity of these mutants should be tested with glycerophosphodiesters and simple alkyl diesters like dimethyl phosphate and diethyl phosphate to examine their substrate specificity.

ii. Our ability to separate the hexamer, dimer and sometimes also monomer with a size exclusion column indicates that the rate of inter-conversion of the species is slow compared to the time required to run the experiment. The dissociate rates need to be further investigated.

iii. The thermodynamics of association/dissociation and the molecular dynamics of the GpdQ hexamer, dimer and monomer should be studied.

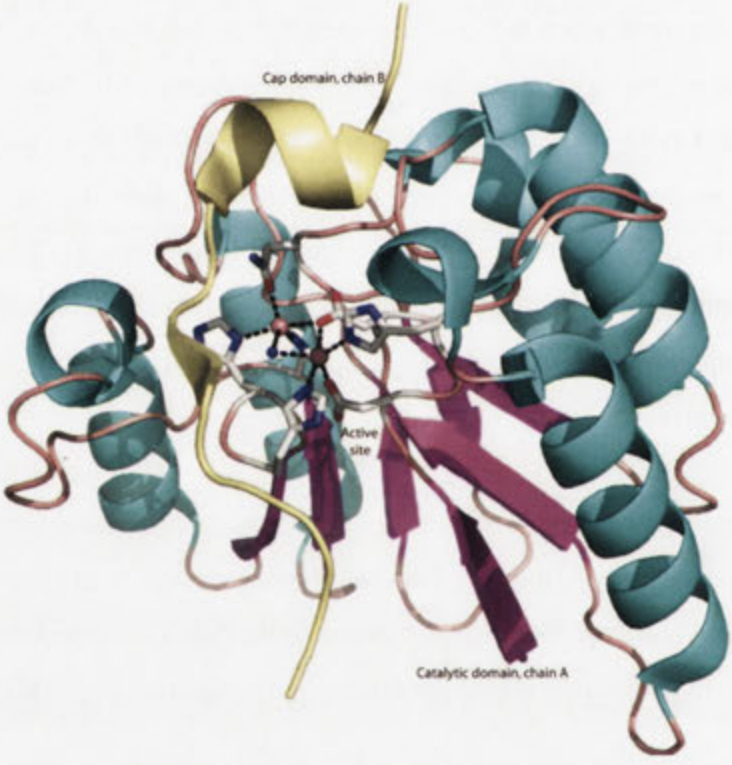


Figure 7.1: A figure showing how the cap domain of chain B overhangs the active site of chain A.

Presently in the Ollis laboratory, there is an ongoing directed evolution project, led by Miss Tracy Murray that aims to improve GpdQ's dimethyl phosphate activity. The agar plate-based growth assay method [6] is used for the preliminary screening of the GpdQ mutant libraries. The *E. coli* BL21(DE3) cells used are equipped with the recombinant putative *ugp* operon from

Enterobacter aerogenes. Hopefully the results from these studies would enable us to answer better to the question whether GpdQ has potential for bioremediation applications.

CHAPTER SEVEN REFERENCES

1. Jackson, C. J., Carr, P. D., Liu, J. W., Watt, S. J., Beck, J. L., and Ollis, D. L. (2007) The structure and function of a novel glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of Molecular Biology 367*, 1047-1062.
2. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural Flexibility Enhances the Reactivity of the Bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.
3. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from Enterobacter aerogenes. *Journal of the American Chemical Society 130*, 14129-14138.
4. Ali, M. H., and Imperiali, B. (2005) Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry 13*, 5013-5020.
5. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure 29*, 105-153.
6. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of *Escherichia coli* coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Applied and Environmental Microbiology 70*, 404-412.

APPENDIX A – MANUFACTURERS AND SUPPLIERS OF EQUIPMENT, KITS, CHEMICALS AND REAGENTS

EQUIPMENT

30°C and 37°C temperature control units	Qualtex Solidstat
AKTA FPLC system	GE Healthcare
All glassware	Duran, E-Mil England, Hirschmann Germany, LMS Germany, Pyrex
All plasticware	Greiner Bio-one, Sarstedt
Autoclave model ASB270BT	Astell Scientific
Auto vortex mixer	Chiltern
Cell density meter	WPA Biowave
Centrifuge 5415, 5804	Eppendorf
Centrifuge 3K18	Sigma Laboratory Supplies
Centrifuge RC5C, RC6+	Thermo Electron Corporation
Centrifuge tubes	Nalgene
Centrifugal filter devices	Amicon, Millipore
Class II Biological safety cabinet	Flow Laboratories
Combination pH electrode	Orion Pacific Pty Ltd
DEAE Sepharose™ fast flow column	GE Healthcare
Dri-Bath Type 17600	Thermolyne
Electricity power supply	Bio-Rad, Pharmacia LKB
French® pressure cell press	SLM Aminco
Fume cabinet	Dynaflow
GeneAmp PCR system 9700	P. E. Applied Biosystems
Hot plate	IKA Laboratory Equipment
iCycler®	Bio-Rad
Labsystem Multiskan Ascent plate reader	Pathtech Pty Ltd
Ligh scattering detector	Wyatt
Micropipettes	Gilson
MicroPulser™	Bio-Rad
Microscope	Olympus
Mighty small gel tank	Hoefer
MF™ 0.45 µm membrane filter	Millipore
Millex®-GV 0.22 µm filter unit	Millipore
Milli-Q reagent water system	Millipore
Minisart® 0.20 µm filter unit	Sartorius
NanoDrop® ND-1000 spectrophotometer	NanoDrop
Orbital mixer and incubator	New Brunswick Scientific, Ratek Instruments
Orbital shaker	Bio-Line
Orbital shaking water bath	Paton Scientific
Peristaltic pump	Gilson
Phenyl Sepharose™ fast flow column	GE Healthcare
Q Sepharose™ fast flow column	GE Healthcare
Qik Spin microcentrifuge	United Biosciences
Robotic dispensing system	Cartesian

Rotors 11133	Sigma Laboratory Supplies
Rotors SA-600, SLA-3000, SS-34	Sorvall
Sephadex™ G-75 column	GE Healthcare
Sonicator	Soniclean Pty Ltd
SpectraMax® M2/M2^e multi-detection microplate reader	Molecular Devices
Superdex™ 200 column	GE Healthcare
Syringe	Terumo
UV transilluminator 312 nm with attached camera for photo acquisition	UVItec
UV transilluminator 365 nm	Hanimax Statesman, UVP Inc.
Varian Cary UV/Vis spectrophotometer	Varian
Water bath with temperature control	Polystat, Thermomix
Weighing scales	Sartorius

KITS

QIAprep® Spin Miniprep Kit	Qiagen
QIAquick® PCR Purification Kit	Qiagen
QIAquick® Gel Extraction Kit	Qiagen
QuikChange™ Site Directed Mutagenesis Kit	Stratagene
Wizard® SV Gel and PCR Clean-Up System	Promega

CHEMICALS AND REAGENTS

1 kb DNA marker	New England Biolabs
10X BugBuster	Merck
10X PCR buffer	Roche
10X PCR buffer without $MgCl_2$	Roche
2% (w/v) N,N'-methylene-bisacrylamide	Merck
40% (w/v) acrylamide solution	Bio-Rad
Acetic acid, anhydrous	Ajax Finechem
Adenosine 5'-monophosphate sodium	Sigma-Aldrich
Adenosine 5'-triphosphate	Sigma-Aldrich
Agar technical	Bacto Laboratories
Agarose, for routine use	Sigma-Aldrich
Alkaline phosphatase, calf intestinal	New England Biolabs
Ammonium chloride	BDH Chemicals
Ammonium molybdate	Chem Supply
Ammonium persulfate	Bio-Rad
Ammonium sulfate, for molecular biology	Sigma-Aldrich
Ampicillin	Amresco
*BamH*I	New England Biolabs
β-mercaptoethanol	Sigma-Aldrich
BigDye terminator v3.1	BRF, JCSMR, ANU
bis (*p*-nitrophenyl) phosphate sodium	Sigma-Aldrich
Bis-Tris	Sigma-Aldrich
Boric acid	BDH Chemicals
Bovine serum albumin	Sigma-Aldrich

Bromophenol blue	Chem Supply
Buffer 3	New England Biolabs
Buffer 4	New England Biolabs
Buffer H	Roche
Calcium chloride	Ajax Finechem
Casamino acids	Difco Laboratories
CHES	Sigma-Aldrich
Cobalt (II) chloride hexahydrate	Sigma-Aldrich
Crystallisation screens	Hampton Research, JENA Bioscience
Cupric sulfate	May and Baker
D-(+)- Glucose, anhydrous	Sigma-Aldrich
Diethyl phosphate	Chem Service
Dimethyl phosphate	Acros
Dimethyl sulfoxide	Sigma-Aldrich
Dipotassium hydrogen orthophosphate	Ajax Finechem
dNTP mix	Roche
*Dpn*I	New England Biolabs
*EcoR*I	New England Biolabs
Ethanol, absolute	Ajax Finechem
Ethidium bromide	Sigma-Aldrich
Ethylenediaminetetraacetic acid disodium	Ajax Finechem
Gelcode® blue stain reagent	Pierce
Glycerol	Ajax Finechem
Glycine, for electrophoresis	Sigma-Aldrich
Guanidine hydrochloride	Sigma-Aldrich
HEPES	Sigma-Aldrich
Hydrochloric acid 37%	Scharlau
High range molecular weight standard	Bio-Rad
Lithium sulfate	Sigma-Aldrich
Low range molecular weight standard	Bio-Rad
Magnesium chloride hexahydrate	BDH Chemicals
Magnesium sulfate	Mallinckrodt
Manganese (II) chloride	Sigma-Aldrich
MES	Sigma-Aldrich
MOPS	Sigma-Aldrich
N,N,N',N',-tetramethylethylenediamine	Sigma-Aldrich
Native *Pfu* DNA polymerase	Stratagene
*Nde*I	New England Biolabs
Nutrient broth	Bacto Laboratories
p-nitrophenyl phosphate disodium	Sigma-Aldrich
Paraoxon-ethyl	Riedel-de-Haen
Paraoxon-methyl	Riedel-de-Haen
Petroleum jelly	Shell
Phenylmethanesulfonylfluoride	Sigma-Aldrich
Potassium chloride	Ajax Finechem
Potassium dihydrogen orthophosphate	
Potassium sulfate	Sigma-Aldrich

Protamine sulfate	Sigma-Aldrich
*Pst*I	New England Biolabs
Sigma 7-9®	Sigma-Aldrich
Sodium acetate	ICN Biomedicals
Sodium chloride	Chem Supply
Sodium dodecyl sulfate	BDH Chemicals
Sodium phosphate dibasic	Sigma-Aldrich
T4 DNA ligase and buffer	New England Biolabs
Tacsimate 100% pH 7.0	Hampton Research
Taq DNA polymerase	Bio-Line; Roche
Tris	Amresco
Tryptone	Bacto Laboratories
*Xba*I	New England Biolabs
*Xho*I	New England Biolabs
Yeast extract	Bacto Laboratories
Zinc chloride	BDH Chemicals

APPENDIX B – RECIPES OF CULTURE MEDIA AND BUFFERS USED IN THIS STUDY

The main reference used in the preparation of the following recipes was Molecular Cloning: A Laboratory Manual by Sambrook and Russell [1].

B.1 CULTURE MEDIA

All media were sterilised by autoclaving for 20 minutes at 15 psi (1.05 kg/cm^2) before use.

Ampicillin (not autoclaved)

Ampicillin was the only selection antibiotic used throughout this study. Stock solutions of ampicillin were prepared by dissolving the ampicillin sodium salt (stored at 4°C) in 50% v/v ethanol (diluted from 100% undenatured ethanol of analytical grade) to a final concentration of 100 mg/mL, *eg.* 1 g in 10 mL. The prepared solutions were stored at -20°C until future use. Due to the presence of ethanol, the solutions did not freeze at -20°C storage and therefore need not be thawed at room temperature before use. Ampicillin was added to cooled, sterilised culture media ($\leq$ 55°C) at a final concentration of 100 µg/mL, *ie.* 1 mL of ampicillin stock solution into every 1 L of media.

Agar media

To prepare agar media, 15 g of technical agar powder was added into every 1 L of media and autoclaved together with the media. When the media cooled to about 55°C after autoclaving, 15-20 mL was poured into each agar plate in a lamina hood and left to set.

Luria-Bertani (LB) medium

To 950 mL of mQH_2O, the following were added and dissolved well:

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

The pH of the medium was adjusted with saturated NaOH solution (~10 M) to 7.5 and the final volume, 1 L.

Terrific Broth (TB)

To 900 mL of mQH_2O, the following were added and dissolved well:

Tryptone	12 g
Yeast extract	24 g
100% glycerol	4 mL

The 10X salt solution was prepared separately by dissolving the following in 100 mL of mQH_2O:

KH_2PO_4	2.31 g
K_2HPO_4	12.54 g

The solutions were autoclaved separately and allowed to cool before mixing them together.

M9 minimal medium

To 750 mL of sterile mQH_2O (cooled to 50°C), the following were added:

5X M9 salts	200 mL
1 M $MgSO_4$	2 mL
20% glucose	20 mL
1 M $CaCl_2$	0.1 mL
Sterile mQH_2O	to 980 mL final volume

5X M9 salts were prepared by dissolving the following in mQH_2O to a final volume of 1 L:

$Na_2HPO_4.7H_2O$	64 g
KH_2PO_4	15 g
NaCl	2.5 g
NH_4Cl	5.0 g

The 5X M9 salt solution was divded into 5 200 mL aliquots and autoclaved. The $MgSO_4$ and $CaCl_2$ solutions were also prepared and autoclaved separately.

Modified MOPS phosphate-free minimal medium

The following recipe was modified from the original one by Neidhardt *et al.* [2]

MOPS minimal medium

Per liter:

10X MOPS mixture	100 mL
mQH_2O	900 mL

The pH was adjusted with saturated NaOH solution (~10 M) to 7.2 and filter sterilised.

10X MOPS mixture

To a total volume of 440 mL, the following were added:

MOPS	83.72 g
Tricine	7.17 g

The pH of the MOPS/Tricine solution was adjusted with NaOH to 7.4. Then 0.01 M of $FeSO_4$ solution was prepared separately, of which 10 mL was added into the MOPS/Tricine solution. Then the following were added to the MOPS/Tricine/ $FeSO_4$ solution according to the order shown:

1.9 M NH_4Cl	50 mL
0.276 M K_2SO_4	10 mL
0.02 M $CaCl_2.2H_2O$	0.25 mL
2.5 M $MgCl_2$	2.1 mL
5 M NaCl	100 mL
Micronutrient stock	0.2 mL
Autoclaved mQH_2O	to 1000 mL final volume

The 10X MOPS mixture was sterilised using vacuum filtration units, dispensed into 100 mL aliquots and stored at -20°C. All the solutions above were prepared individually. To make the micronutrient stock, the following salts were dissolved in mQH_2O, adjusted to a final volume of 50 mL, autoclaved and stored at room temperature:

Ammonium molybdate $[(NH_4)_6Mo_7O_{24}.4H_2O]$	0.009 g
Boric acid (H_3BO_3)	0.062 g
$CoCl_2$	0.018 g
$CuSO_4$	0.006 g
$MnCl_2$	0.040 g
$ZnSO_4$	0.007 g

B.2 DNA AGAROSE GEL ELECTROPHORESIS

1.0 % agarose gel with ethidium bromide

1.0 % w/v agarose gel was prepared by dissolving 1 g of agarose into 100 mL SB or TBE buffer. The solution was brought to a brief boil in the microwave, shaken to dissolve all agarose and cooled at room temperature to about 55°C. 5 µl of ethidium bromide solution (10 mg/mL) to a final concentration of 0.5 µg/mL. For gel extraction in library making purposes, the amount of ethidium bromide was increased to 1.0 µg/mL.

20X SB (sodium hydroxide-boric acid) buffer

Per liter as dissolved in mQH_2O:

0.2 M NaOH adjusted to pH 8.0 with boric acid (~48 g).

10X TBE (Tris-boric acid-EDTA) buffer

Per liter:

Tris	108 g
Boric acid	55 g
0.5 M EDTA	20 mL

6X Loading dye

20 mM Tris pH 8.0, 10% w/v of bromophenol blue (BPB) and 40% v/v glycerol.

B.3 SDS-PAGE

Both the running and stacking gel stock solutions were stored at 4°C. 10% w/v APS solution was stored at -20°C. TEMED was stored at room temperature.

17.5% running gel

17.5% running gel stock solution	3.75 mL
10% w/v APS	25 µl
TEMED	3 µl

17.5% running gel stock solution

40% acrylamide	21.9 mL
2% bis-acrylamide	1.8 mL
2 M Tris pH 8.8	9.4 mL
10% w/v SDS	0.5 mL
mQH_2O	to 50.0 mL final volume

Stacking gel

Stacking gel stock solution	1.7 mL
10% w/v APS	25 µl
TEMED	3 µl

Stacking gel stock solution

40% acrylamide	6.3 mL
2% bis-acrylamide	3.3 mL
2 M Tris pH 6.8	3.1 mL
10% w/v SDS	0.5 mL
mQH_2O	to 50.0 mL final volume

10X running buffer

Per liter in mQH_2O:

Glycine	144 g
Tris	30 g
SDS	10 g

Cracking buffer

200 mM Tris pH 6.8, 2% w/v SDS, 0.1% w/v bromophenol blue (BPB) and 5% v/v β-mercaptoethanol.

B.4 PROTEIN PURIFICATION

All buffers in protein purification were filtered with syringe driven 0.20 µm filters before use.

DEAE anion exchange chromatography

Starting buffer (buffer A) – 50 mM HEPES (11.92 g in 1 L) pH 8.0 at 4°C (adjusted with NaOH).

Eluting buffer (buffer B) – 50 mM HEPES pH 8.0 at 4°C, 1 M NaCl (58.44 g in 1 L).

Phenyl Sepharose hydrophobic interaction chromatography

Buffer A – same as for the DEAE column above.

Buffer B – 50 mM HEPES pH 8.0 at 4°C, 1 M ammonium sulfate (132.1 g in 1 L).

Q Sepharose anion exchange chromatography

Buffer A – 50 mM Tris (6.07 g in 1 L) pH 8.0 at 4°C (adjusted with hydrochloric acid HCl).

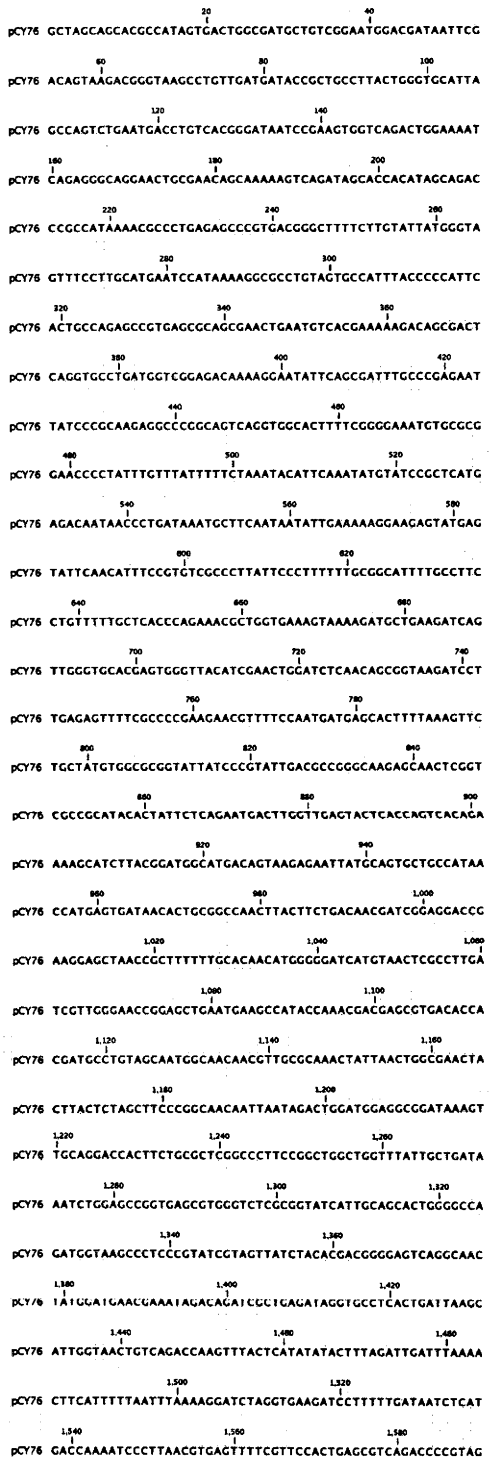
Buffer B – 50 mM Tris pH 8.0 at 4°C, 1 M NaCl.

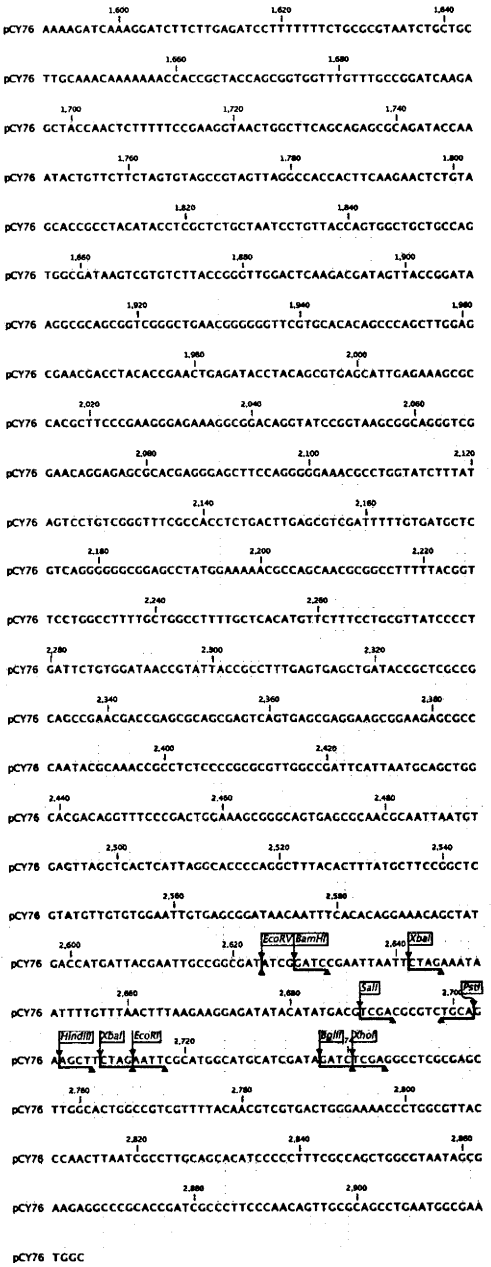
Superdex 200 size exclusion chromatography

One buffer, with the same formula as buffer A for the Q Sepharose column above.

The recipes of all other buffers and reagents have been described in good detail in Chapter 2.

APPENDIX C – NUCLEOTIDE SEQUENCE OF PLASMID pCY76




pCY76 AAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGC
pCY76 TTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGA
pCY76 GCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAA
pCY76 ATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
pCY76 GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAG
pCY76 TGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
pCY76 AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG
pCY76 CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGC
pCY76 CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCG
pCY76 GAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTAT
pCY76 AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC
pCY76 GTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
pCY76 TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCT
pCY76 GATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCG
pCY76 CAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
pCY76 CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGG
pCY76 CACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGT
pCY76 GAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTC
pCY76 GTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTAT
pCY76 GACCATGATTACGAATTGCCGGCGATATCGCATCCGAATTAATTCTAGAAATA
pCY76 ATTTTGTTTAACTTTAAGAAGGAGATATACATATGACGTCGACGCGTCTGCAG
pCY76 AAGCTTCTAGAATTCGCATGGCATGCATCGATATCGATCTCGAGGCCTCGCGAGC
pCY76 TTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTAC
pCY76 CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG
pCY76 AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAA
pCY76 TGGC
EcoRV
BamHI
XbaI
SalI
PstI
HindIII
XbaI
EcoRI
BglII
XhoI

APPENDIX D – SEQUENCES OF PRIMERS USED IN THIS RESEARCH PROJECT

Primer	Forward (5' → 3')	Reverse (5' → 3')
Sequencing	GCC GGC GAT ATC GGA TCC G	GTA AAA CGA CGG CCA GT
Chapter 3		
D8N	TGT TAG CGC ACA TTT CC**A** ATA CCC ATT TCC GCA GC	GCT GCG GAA ATG GGT AT**T** GGA AAT GTG CGC TAA CA
D8A	TTA GCG CAC ATT TCC **GC**T ACC CAT TTC CGC AGC	GCT GCG GAA ATG GGT **AGC** GGA AAT GTG CGC TAA
H10A	CAC ATT TCC GAT ACC **GC**T TTC CGC AGC CGC GG	CCG CCG CTG CGG AAA **GC**G GTA TCG GAA ATG TG
D50N	GTG GTG GTG AGC GGC **AA**T ATC GTC AAC TGC G	CGC AGT TGA CGA TAT **T**GC CGC TCA CCA CCA C
D50A	GGT GGT GAG CGG C**GC** TAT CGT CAA CTG CG	CGC AGT TGA CGA TA**G** **C**GC CGC TCA CCA CC
N80D	CTC ATC CCC GGC GA**C** CAC GAT GAT AAA GCG	CGC TTT ATC ATC GTG **G**TC GCC GGG GAT GAG
N80A	CTC ATC CCC GGC G**CC** CAC GAT GAT AAA GCG	CGC TTT ATC ATC GTG **GG**C GCC GGG GAT GAG
H81Q	ATC CCC GGC AAC CA**G** GAT GAT AAA GCG	CGC TTT ATC ATC **C**TG GTT GCC GGG GAT
H81A	ATC CCC GGC AAC **GC**C GAT GAT AAA GCG	CGC TTT ATC ATC G**GC** GTT GCC GGG GAT
H156Q	GCA GCG GCG GGT GCT GCA **T**AA AGA TCG TTG C	GCA ACG ATC TTT **A**TG CAG CAC CCG CCG CTG C
H156A	CGG CAA CGA TCT TTA TGG CC**C** ACC CGC CGC T	AGC GGC GGG TGG G**C**C ATA AAG ATC GTT GCC G
H195Q	CGC ATC TTT TGC GGT CA**G** AAC CAT AGC CTG ACC AT	ATG GTC AGG CTA TGG TT**C** TGA CCG CAA AAG ATG CG
H195N	CGC GCA TCT TTT GCG GT**A** ATA ACC ATA GCC TGA CC	GGT CAG GCT ATG GTT AT**T** ACC GCA AAA GAT GCG CG
H195A	CGC GCA TCT TTT GCG GT**G** **C**TA ACC ATA GCC TGA CCA	TGG TCA GGC TAT GGT TA**G** **C**AC CGC AAA AGA TGC GCG
Chapter 4		
epPCR, StEP	GGA GAT ATA CAT ATG CTG TTA GCG	GCC ATG CGA ATT CCT TAT TAG CGC
C54G	GCG ATA TCG TCA AC**G** GCG GCC GTC CGG A	TCC GGA CGG CCG C**C**G TTG ACG ATA TCG C
S122C	GTT TAT CGA CTC C**T**G CCG CGC CGG CAC	GTG CCG GCG CGG C**A**G GAG TCG ATA AAC
N164D	CTG CCG CTG GGC GA**T** GCG CAG ATG G	CCA TCT GCG C**A**T CGC CCA GCG GCA G
H217R	CCC GGC ACC GTC C**GT** CAG GTG CCT TAC	GTA AGG CAC CTG A**C**G GAC GGT GCC GGG
Y221N	ACC GTC CAT CAG GTG CCT **AA**C TGC CAC GA	TCG TGG CAG T**T**A GGC ACC TGA TGG ACG GT
Y221H	ACC GTC CAT CAG GTG CCT **CA**C TGC CAC GA	TCG TGG CAG T**G**A GGC ACC TGA TGG ACG GT
S233A	CGT ATT ACG ATC TC**G** CGC CGG CTT CGT GC	GCA CGA AGC CGG CG**C** GAG ATC GTA ATA CG
S253L	GGT GAG CTA CCA GCA C**TT** GCT GGC CCA	TGG GCC AGC **AA**G TGC TGG TAG CTC ACC
G259R	CCC ACT ACG CC**A** GGC CGT GGC TG	CAG CCA CGG C**C**T GGC GTA GTG GG
Y263C	CGC CGG GCC GTG GCT GT**G** CGA CGA AAA	TTT TCG TCG **C**AC AGC CAC GGC CCG GCG
I267T	GCT GTA CGA CGA AAA CA**C** CAG TTG TCC AAC GGA AG	CTT CCG TTG GAC AAC TG**G** TGT TTT CGT CGT ACA GC
Chapter 5		
F21K	GCG AGA AGC TGT ACG **GCA AG**A TCG ACG TCA ACG CCG C	GCG GCG TTG ACG TCG AT**C** **TT**G CCG TAC AGC TTC TCG C
N53D	CGG CGA TAT CGT CGA C**T**G CGG CCG TCC	GGA CGG CCG CA**G** TCG ACG ATA TCG CCG
N53A	CGG CGA TAT CGT C**GC** CTG CGG CCG TCC G	CGG ACG GCC GCA G**GC** GAC GAT ATC GCC G
Y204stop	TGA CCA TGA CCC AGT A**G**C GCC AGG CGC	GCG CCT GGC G**C**T ACT GGG TCA TGG TCA

Y257stop	CTG GCC CAC TA**A** GCC GGG CCG TGG CTG TAC G	CGT ACA GCC ACG GCC CGG C**T**T AGT GGG CCA G
W261R	GGG CCG **A**GG CTG TAC GAC GAA AAC ATC AGT TG	GTC GTA CAG CC**T** CGG CCC GGC GTA GTG
C269S	CGA CGA AAA CAT CAG **TA**G TCC AAC GGA AGA GCG	CGC TCT TCC GTT GGA C**T**A CTG ATG TTT TCG TCG
C269A	GTG GCT GTA CGA CGA AAA CAT CAG **TGC** TCC AAC GGA AGA G	CTC TTC CGT TGG **AGC** ACT GAT GTT TTC GTC GTA CAG CCA C
Chapter 6 (Primers for epPCR and StEP as given under Chapter 4)		
R12L	GAT ACC CAT TTC C**T**C AGC CGC GGC GAG	CTC GCC GCG GCT **GA**G GAA ATG GGT ATC
N103D	CGG TAG CGA TGC C**GA** TAA TAT GCG C	GCG CAT ATT AT**C** GGC ATC GCT ACC G
S127A	CGC CGG CAC **TGC** AAA AGG CTG GCT G	CAG CCA GCC TTT TG**C** AGT GCC GGC G
Site saturation mutagenesis		
C54SSM	GGC GAT ATC GTC AAC **NNN** GGC CGT CCG GAG GAG TAT C	GAT ACT CCT CCG GAC GGC **CNN** NGT TGA CGA TAT CGC C
Y221SSM	CCG TCC ATC AGG TGC C**T**N **NN**T GCC ACG AAG ACA CAC TCGC	GCG AGT GTC TTC GTG GCA **NNN** AGG CAC CTG ATG GAC GG
G259SSM	CTG GCC CAC TAC GCC **NNN** CCG TGG CTG TAC GAC GAA AAC	GTT TTC GTC GTA CAG CCA CGG **NNN** GGC GTA GTG GGC CAG
Appendix E		
D168A	AAT GCG CAG ATG G**C**C CCG ATT GCC	GGC AAT CGG G**GC** CAT CTG CGC ATT

Notes

- The pair of primers used for error-prone PCR (epPCR) and staggered extension process (StEP) is referred to as SY-F (forward primer) and SY-R (reverse primer) in Chapter 2. The *Nde*I and *EcoR*I cloning sites in SY-F and SY-R respectively, are underlined in this table.
- Primers used in PCR cycle sequencing are referred as 4pCY76 (forward primer) and M13 (reverse primer) in Chapter 2.
- The rest of the primers listed in this table are mutagenic, with the nucleotide changes highlighted in bold.

APPENDIX E – PROTEOLYTIC ACTIVATION OF GpdQ

All purified GpdQ enzymes (native and mutant), when visualised in an SDS-PAGE gel, had a thick band associated with 30.8 kDa calculated molecular weight of the monomer. Apart from this main band, there were four other smaller fragments between the estimated sizes of 7 and 23 kDa. Dr Colin Jackson suggested that GpdQ went through proteolytic activation after he showed that one batch of purified GpdQ that suffered from a higher level of proteolysis was more active than another batch with a lower level of proteolysis. He went on to predict these proteolysis sites by identifying the N-terminal sequence of the four small fragments and reconciling the identified proteolysis sites with the molecular weight of the fragments [1].

One of the proposed proteolytic cleavage sites was between D168 and P169, where the proteolysis would result in two fragments of polypeptide sized 18.6 kDa and 12.2 kDa.

Cleavage of aspartate-proline (D-P) bonds has been well documented in literature. Some cleavages are reported to be sensitive to acidic conditions [2], occur at very low efficiency [3], calcium dependent [4] and autocatalytic [5].

A mutant, D168A was made to confirm proteolysis at this site. The other suggested D-P cleavage between D227 and P228, was not explored further as it was likely to be erroneous. At position 227, the correct residue is an arginine and not aspartate.

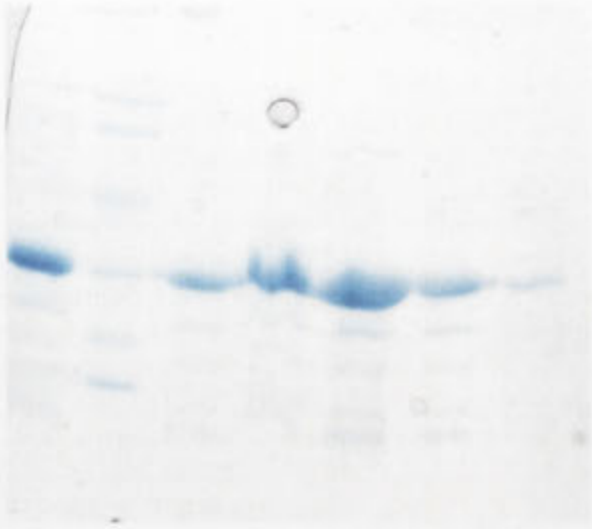


Figure E.1: SDS-PAGE visualisation of purified wild-type and D168A GpdQ enzymes.
Lane 1 shows the purified wild-type GpdQ enzyme; Lane 2 is the marker; Lanes 3-7 are fractions of D168A from the final purification step.

The SDS-PAGE gel above shows that there was no apparent difference between the level of proteolysis between D168A and the wild-type enzyme. D168A was assayed with b*p*NPP and the results are given in the following table. The mutation produced a deleterious effect on the activity of the enzyme.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_m (mM)	Relative K_m	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_m
Wild type	7.50 ± 0.1	1.00	2.10 ± 0.08	1.00	3.57×10^3	1.00
D168A	0.79 ± 0.1	0.11	3.21 ± 0.41	1.53	2.47×10^2	0.07

Table E.1: Kinetic results of D168A.

APPENDIX E REFERENCES

1. Jackson, C. J. (2006) Enzymatic catalysis of phosphate ester hydrolysis. *The Australian National University.*
2. Piszkiewicz.D, Landon, M., and Smith, E. L. (1970) Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. *Biochemical and Biophysical Research Communications 40*, 1173-1178.
3. Breddam, K., and Meldal, M. (1992) Substrate preferences of glutamic-acid-specific endopeptidases assessed by synthetic peptide substrates based on intramolecular fluorescence quenching. *European Journal of Biochemistry 206*, 103-107.
4. Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I., Havlicek, V., and Sebo, P. (2004) A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens - Covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond. *Journal of Biological Chemistry 279*, 24944-24956.
5. Lidell, M. E., Johansson, M. E. V., and Hansson, G. C. (2003) An autocatalytic cleavage in the c terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. *Journal of Biological Chemistry 278*, 13944-13951.

APPENDIX F – FACTORS INFLUENCING THE NEGATIVE COOPERATIVITY OF GpdQ MUTANT N80A

F.1 N80A PHOSPHODIESTERASE ACTIVITY WITH BIS (*P*-NITROPHENYL) PHOSPHATE

Conditions of N80A assays with bis (*p*-nitrophenyl) phosphate were varied in order to explore factors that might affect the negative cooperativity in the mutant enzyme. Table E.1 summarises the kinetic parameters (k_{cat}, K_m, k_{cat}/K_m) and values of the Hill coefficient (n) and their corresponding correlation coefficient (R^2) values.

Enzyme and assay conditions	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	n	R^2 (of n)
Wild type + 0.5 mM Mn^{2+}	7.50 ± 0.1	2.10 ± 0.08	3.57×10^{3}	1.0	1.00
N80A without added metals	0.42 ± 0.04	2.10 ± 0.03	1.99×10^{2}	0.9	0.97
N80A	2.28	38.62	5.90×10^{1}	0.6	0.99
+ 0.5 mM Mn^{2+}	0.17	0.39	4.26×10^{2}		
N80A	9.28	73.66	1.26×10^{2}	0.7	0.98
+ 1.0 mM Mn^{2+}	0.37	0.42	4.26×10^{2}		
N80A	4.13	137.05	3.01×10^{1}	0.6	0.99
+ 1.0 mM Co^{2+}	0.09	0.38	2.45×10^{2}		
N80A + 0.5 mM Mn^{2+} + 0.5 mM P_i	1.19 ± 0.01	4.85 ± 0.19	2.47×10^{2}	1.0	1.00
N80A + 0.5 mM P_i	0.73 ± 0.03	4.82 ± 0.56	1.51×10^{2}	1.0	0.99

Table F.1: A summary of the kinetic parameters and the Hill constant values of N80A assayed in various conditions.
The kinetic data for wild-type GpdQ was also included for comparison. Reactions that showed negative cooperativity are shaded in yellow while the rest are shaded in turqoise.

For all the assay reactions in Table E.1, other conditions like 50 mM CHES pH 9.0, 0.1 mg/mL BSA and temperature (20°C) were kept the same. Kinetic parameters for normal reactions with no cooperativity (shaded in turqoise in Table E.1) were calculated directly from their respective Michaelis-Menten curves (Equation 2.3, Chapter 2) using the KaleidaGraph software. Apparent values of kinetic parameters for reactions that exhibited negative cooperativity were estimated from their Eadie-Hofstee plots using

Microsoft Excel (section 3.3.4.1, Chapter 3). The Michaelis-Menten curves of the assay reactions are shown in Figure F.1.

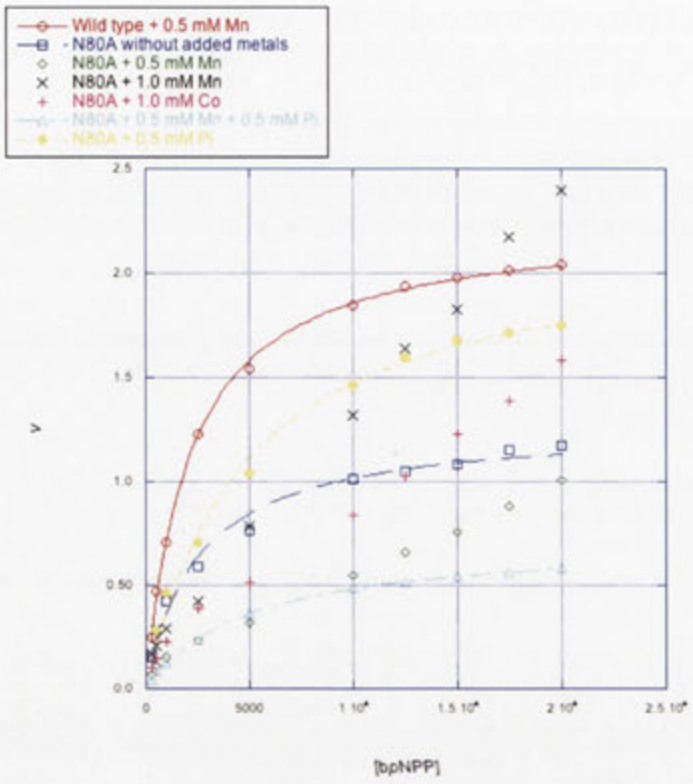


Figure F.1: The Michaelis-Menten curves of N80A assayed with b*p*NPP in different conditions.

F.2 N80A PHOSPHOMONOESTERASE ACTIVITY WITH *P*-NITROPHENYL PHOSPHATE

GpdQ is a promiscuous enzyme, showing primarily phosphodiesterase activity and low activities with many other organophosphate compounds, including phosphomonoesters and phosphotriesters [1]. GpdQ has been shown to display higher phosphomonoesterase activity at lower pH values [2, 3]. For this reason, all assays with b*p*NPP in this study were performed at pH 9.0 in order to minimise sequential phosphomonoesterase activity (b*p*NPP → *p*NPP → inorganic phosphate) that would affect the accuracy of the kinetic results.

Wild-type GpdQ, N80D and N80A were assayed with *p*-nitrophenol phosphate (*p*NPP) to characterise their phosphomonoesterase activities and to assess if the biphasic velocity curves observed of N80A could be a mixture of diesterase and monoesterase activities. Kinetic results are summarised in Table E.3. Assay conditions were 50 mM CHES pH 9.0, 0.5 mM $MnCl_2$ or $CoCl_2$, 0.1 mg/mL BSA at 20°C.

Enzyme	k_{cat} (s^{-1})	Relative k_{cat}	K_m (mM)	Relative K_m	k_{cat}/K_m ($M^{-1}s^{-1}$)	Relative k_{cat}/K_m
Wild type-Mn^{2+}	0.08 ± 0.0002	1.00	4.51 ± 0.38	1.00	1.82×10^{1}	1.00
N80D-Mn^{2+}	0.09 ± 0.005	1.13	11.74 ± 1.32	2.60	7.54×10^{0}	0.41
N80A-Mn^{2+}	0.16 ± 0.04	2.00	17.27 ± 8.32	3.83	9.57×10^{0}	0.53
Wild type-Co^{2+}	0.06 ± 0.002	1.00	3.54 ± 0.34	1.00	1.62×10^{1}	1.00
N80A-Co^{2+}	0.19 ± 0.06	3.17	25.22 ± 0.16	0.76	7.49×10^{0}	0.46

Table F.2: *p*NPP activities of wild-type GpdQ, N80D, N80A with Mn^{2+} or Co^{2+}.

Both N80D and N80A mutations resulted in a drastic drop in the affinity of GpdQ for *p*NPP. The catalytic turnover, k_{cat}, was however improved by the N80A mutation in both the assays with Mn^{2+} and Co^{2+}, although the ultimate k_{cat}/K_m ratio values were still much lower than those of the wild-type GpdQ. The kinetic progress curves, of which an example is shown below in Figure E.2, were not diagnostic of any obvious sequential hydrolysis of b*p*NPP and *p*NPP by GpdQ.

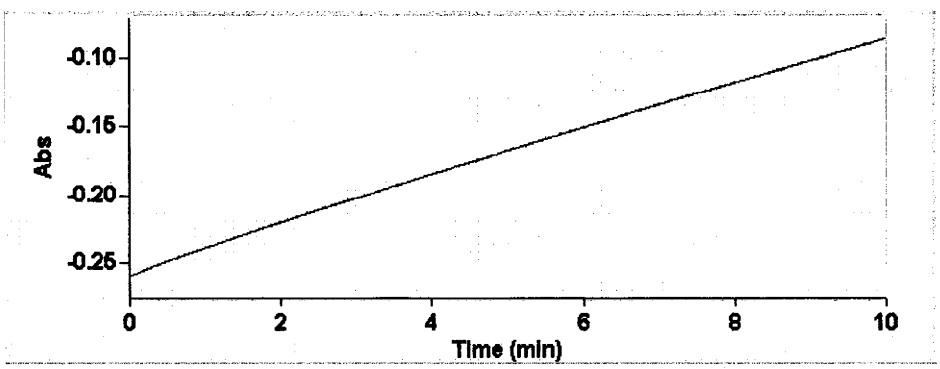


Figure F.2: A sample of a kinetic progress curve of N80A hydrolysis of 20 mM b*p*NPP.

F.3 INTRA/INTER-MOLECULAR NEGATIVE COOPERATIVITY IN N80A

Aghajanian and Engel demonstrated that negative cooperativity in the glutamate dehydrogenase from *Clostridium symbiosum* was due to interactions between NAD^+ binding sites on the six subunits of the homo-hexamer [4].

To investigate if biphasic kinetics/negative cooperativity existed between the dimeric units in the N80A hexamer, mutant 3-25-N80A was generated. 3-25 (C54G, Y221H and G259R) is a GpdQ mutant from directed evolution that dissociated into a mixture of hexamer and dimer.

3-25-N80A hexamer and 3-25-N80A dimer were purified and assayed with b*p*NPP separately. The Michaelis-Menten curves of the assays are shown in Figure E.3 while values of the Hill coefficient for both enzyme forms are given in Table E.3.

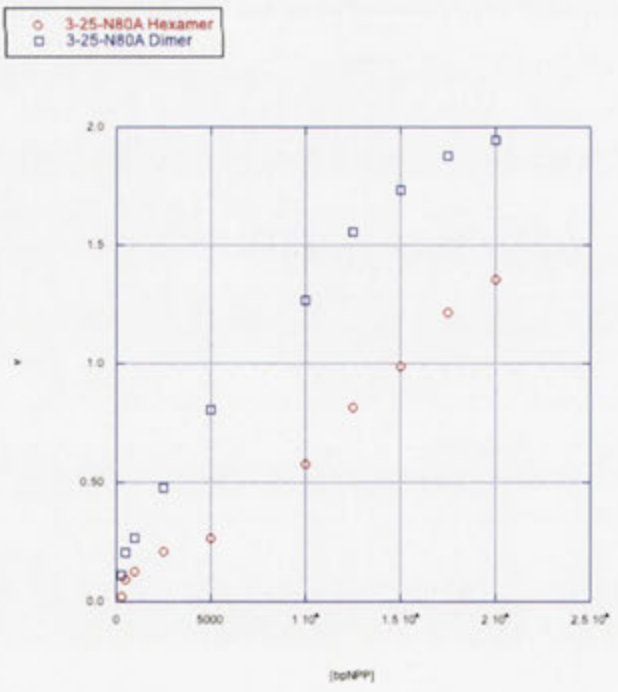


Figure F.3: Michaelis-Menten plots of 3-25-N80A hexamer and 3-25-N80A dimer assayed with b*p*NPP.

Enzyme	*n*	R^2
3-25-N80A hexamer	0.9	0.99
3-25-N80A dimer	0.9	0.96

Table F.3: Values of the Hill coefficient (*n*) and correlation coefficient (R^2) for kinetic assay reactions of 3-25-N80A hexamer and dimer with b*p*NPP.

F.4 CONCLUSIONS ON THE NEGATIVE COOPERATIVITY OF N80A

Based on the results of all the studies on the negative cooperativity in sections F.1-F.3, the following conclusions are made:

i. **Negative cooperativity was induced by the presence of metal ions**
N80A was assayed using b*p*NPP as substrate without adding any exogenous metals. The activity of the apoenzyme was extremely low but was found to obey Michaelis-Menten kinetics.

ii. **Negative cooperativity was not dependent on the concentration of metal ions**
N80A was assayed for b*p*NPP activity again with the concentration of Mn^{2+} increased to 1.0 mM. The activity of the enzyme increased but biphasic kinetics could still be observed.

iii. **Negative cooperativity was not metal specific**
The b*p*NPP activity of N80A was tested in the presence of 1.0 mM Co^{2+}. Biphasic kinetics was observed and the enzyme was less active with Co^{2+} than it was with Mn^{2+}.

iv. **The presence of P_i eliminated negative cooperativity in N80A**
It had been previously shown by Hadler *et al.* in their magnetic circular dichroism (MCD) [5] and electron paramagnetic resonance studies [6], how the addition of inorganic phosphate (P_i) promoted metal binding at the β-site in N80A. Therefore, kinetic assays of

N80A were repeated in the presence of 0.5 mM sodium monophosphate monobasic (FW = 119.98) as a supply of P_i. It was found that in the presence of P_i, N80A displayed standard Michaelis-Menten kinetics.

The addition of P_i into assay reaction mixtures resulted in an almost two-fold increase in K_m, regardless of the presence of Mn^{2+}. GpdQ has a low level of phosphomonoesterase activity, which yields P_i as a final product. It is possible that the relationship between the active site of GpdQ and P_i happened as a result of product inhibition. Product inhibition of enzymes has long been recognised as a negative feedback mechanism in the homeostatic regulation of enzyme activity [7]. In fact, the solved crystal structure of native Rv0805 3',5'-cyclic nucleotide phosphodiesterase (PDB file 2HY1) had a P_i molecule bound at its active site [8].

Interestingly, in the presence of P_i and absence of exogenous Mn^{2+}, the catalytic rate of N80A was higher than that of N80A assayed without any added $Mn^{2+}$and P_i. This implied that P_i promoted enzyme activity in the absence of excessive metal ion co-factors.

v. **Biphasic kinetics was substrate specific**

N80A did not exhibit biphasic but standard Michaelis-Menten kinetics with *p*NPP. A different catalytic mechanism involved in the hydrolysis of monoesters would be the likely reason for this.

vi. **Negative cooperativity was present between the two metal-binding sites of the GpdQ active site**

3-25-N80A hexamer and dimer, despite their different oligomeric forms, displayed slight negative cooperativity, with both enzymes having the same *n* value of 0.9. This implies that the observed negative cooperativity behaviour in N80A was most likely intra-

(within the binuclear active site of an N80A protein monomer) and not intermolecular.

Negative cooperativity in cyclic nucleotide phosphodiesterases has been well characterised [9-12]. Negative cooperativity has its advantages in that it enables the enzyme to respond to a broader range of ligand or substrate concentrations [13] and at these exceedingly high ligand concentrations, the catalytic rate accelerates greatly [14].

The negative cooperativity of N80A was most likely attributed to the mutation disabling the coordination of the β-metal ion by the N80 ligand that subsequently weakened further the metal binding affinity at the β-site. Some kinetic models conceptualise discrete and static sub-sites within a single, large active site [15]. This concept of discrete and static sub-sites is supported by mutagenesis studies done by Halpert and colleagues [16-18]. In the case of GpdQ, these two sub-sites referred to are the two α- and β-metal binding sites. Each subsite would have its characteristic affinity for each ligand. In a most extreme case with N80A, metal ions and substrate molecules bind sequentially to the high-affinity α-subsite and then β-subsite at higher substrate concentrations. If this catalytic mechanism is true, then this would mean that at higher substrate concentrations, the N80A active site would actually hydrolyse two molecules of b*p*npp at a time, one at each metal coordinating site.

APPENDIX F REFERENCES

1. Ghanem, E., Li, Y. C., Xu, C. F., and Raushel, F. M. (2007) Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry 46*, 9032-9040.
2. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of Escherichia coli coexpressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorus source. *Applied and Environmental Microbiology 70*, 404-412.
3. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-Promoted Formation of a Catalytically Competent Binuclear Center and Regulation of Reactivity in a Glycerophosphodiesterase from Enterobacter aerogenes. *Journal of the American Chemical Society 130*, 14129-14138.
4. Aghajanian, S., and Engel, P. C. (1998) Use of protein engineering to explore subunit interactions in an allosteric enzyme: construction of inter-subunit hybrids in *Clostridium symbiosum* glutamate dehydrogenase. *Protein Engineering 11*, 569-575.
5. Hadler, K. S., Tanifum, E. A., Yip, S. H. C., Mitic, N., Guddat, L. W., Jackson, C. J., Gahan, L. R., Nguyen, K., Carr, P. D., Ollis, D. L., Hengge, A. C., Larrabee, J. A., and Schenk, G. (2008) Substrate-Promoted Formation of a Catalytically Competent Binuclear Center and Regulation of Reactivity in a Glycerophosphodiesterase from *Enterobacter aerogenes*. *Journal of the American Chemical Society 130*, 14129-14138.
6. Hadler, K. S., Mitic, N., Ely, F., Hanson, G. R., Gahan, L. R., Larrabee, J. A., Ollis, D. L., and Schenk, G. (2009) Structural Flexibility Enhances the Reactivity of the Bioremediator glycerophosphodiesterase by fine-tuning its mechanism of hydrolysis. *Journal of the American Chemical Society 131*, 11900-11908.
7. Frieden, E., and Walter, C. (1963) Prevalence and significance of product inhibition of enzymes. *Nature 198*, 834-837.
8. Shenoy, A. R., Capuder, M., Draskovic, P., Lamba, D., Visweswariah, S. S., and Podobnik, M. (2007) Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from *Mycobacterium tuberculosis*. *Journal of Molecular Biology 365*, 211-225.
9. Erneux, C., Boeynaems, J. M., and Dumont, J. E. (1980) Theoretical analysis of the consequences of cyclic nucleotide phosphodiesterase negative co-operativity - amplification and positive co-operativity of cyclic AMP accumulation. *Biochemical Journal 192*, 241-246.

10. Russell, T. R., Schneide.Fw, Thompson, W. J., and Appleman, M. M. (1972) 3'-5'-Cyclic adenosine monophosphate phosphodiesterase - negative cooperativity. *Proceedings of the National Academy of Sciences of the United States of America 69*, 1791-1795.
11. Russell, T. R., Terasaki, W. L., and Appleman, M. M. (1973) Separate phosphodiesterases for hydrolysis of cyclic adenosine 3' 3',5'-monophosphate and cyclic guanosine 3' 3',5'-monophosphate in rat liver. *Journal of Biological Chemistry 248*, 1334-1340.
12. Thompson, W. J., and Appleman, M. M. (1971) Characterization of cyclic nucleotide phosphodiesterases of rat tissues. *Journal of Biological Chemistry 246*, 3145-3150.
13. Koshland, D. E. (1996) The structural basis of negative cooperativity: Receptors and enzymes. *Current Opinion in Structural Biology 6*, 757-761.
14. Sakai, S. (1994) Negative cooperativity in the prolactin-receptor interaction in the rabbit mammary gland - action of high prolactin concentration. *Journal of Dairy Science 77*, 433-438.
15. Atkins, W. M. (2005) Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annual Review of Pharmacology and Toxicology 45*, 291-310.
16. Harlow, G. R., and Halpert, J. R. (1998) Analysis of human cytochrome P450 3A4 cooperativity: Construction and characterization of a site-directed mutant that displays hyperbolic steroid hydroxylation kinetics. *Proceedings of the National Academy of Sciences of the United States of America 95*, 6636-6641.
17. Davydov, D. R., Kumar, S., and Halpert, J. R. (2002) Allosteric mechanisms in P450eryF probed with 1-pyrenebutanol, a novel fluorescent substrate. *Biochemical and Biophysical Research Communications 294*, 806-812.
18. Domanski, T. L., He, Y. A., Harlow, G. R., and Halpert, J. R. (2000) Dual role of human cytochrome P450 3A4 residue phe-304 in substrate specificity and cooperativity. *Journal of Pharmacology and Experimental Therapeutics 293*, 585-591.