UNIVERSITY OF SOUTHAMPTON

X-RAY CRYSTALLOGRAPHIC STUDIES OF ENZYMES BY MOLECULAR REPLACEMENT

by
FIYAZ MOHAMMED

A thesis submitted for the degree of
DOCTOR OF PHILOSOPHY

Department of Biochemistry

February 2001

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
DIVISION OF BIOCHEMISTRY

Doctor of Philosophy

X-RAY CRYSTALLOGRAPHIC STUDIES OF ENZYMES BY MOLECULAR REPLACEMENT

by Fiyaz Mohammed

The structures of four enzymes, namely human ubiquitous R167Q porphobilinogen deaminase (uPBGD), bovine inositol monophosphatase (IMPase), D303E methanol dehydrogenase (MDH) from *Methylobacterium extorquens* and 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid 5,6-hydrolase (MhpC) from *E. coli* have been solved by X-ray crystallography using molecular replacement.

PBGD is the third enzyme in the haem biosynthetic pathway and assembles four molecules of porphobilinogen to form the linear tertrapyrrole preuroporphyrinogen III. The R167Q mutation in human PBGD retains 6-7 % activity of the wild type enzyme and is one of those responsible for the hereditary disorder, acute intermittent porphyria. Initial phases for human R167Q-uPBGD were obtained using the *E. coli* PBGD as the search model and the structure has been refined to 2.65 Å. Preliminary structural analysis demonstrates that the dipyrromethane cofactor is in the reduced (active) form covalently linked to Cys-261 in domain 3. A sequence insertion in domain 3 is located after strand $\beta 3_3$. It stabilises the domain 1 interface and the hydrophobic core in domain 3. The additional electron density present at the active site cleft may arise from a bound substrate molecule.

IMPase is a key enzyme in the phosphatidylinositol signalling pathway and catalyses the hydrolysis of inositol monophosphate to liberate inositol. The enzyme is the putative target for the mood stabilising drug, lithium. A bovine IMPase crystal grown under inhibitory conditions (20 mM $MgCl_2$) with a dimer in the asymmetric unit diffracted X-rays to 1.65 Å resolution. Initial phases were obtained by molecular replacement with the human IMPase structure as the search model. Although the refined structure shows that each subunit possesses 3 bound Mg^{2+} ions, only two are implicated in the mechanism. The third Mg^{2+} may diminish the ability of Glu-70 to activate a water molecule.

MDH is a periplasmic quinoprotein in methylotrophic bacteria that utilises pyrrolo-quinoline quinone (PQQ) to oxidize methanol to formaldehyde. The reaction is presumably initiated by the abstraction of a proton from methanol by Asp-303. Site directed mutagenesis of the putative nucleophile (D303E) yields an active mutant although the apparent K_m has increased 80-000-fold. The D303E-MDH crystal with two $\alpha_2\beta_2$ tetramers in the asymmetric unit diffracted X-rays to 3.0 Å resolution. Initial phases for D303E-MDH were obtained using the wild type structure as the search model. The refined structure of D303E-MDH reveals that the extended Glu-303 side-chain forms new interactions with the C-5 carbonyl group of PQQ and the Ca^{2+} ion, thus reducing its likelihood of abstracting a proton from methanol. The reduced affinity for methanol can be attributed to either steric hindrance by Glu-303 or to the change in the co-ordination sphere of the Ca^{2+} ion.

MhpC, a bacterial C-C hydrolase involved in the degradation of 3-hydroxy phenylpropionic acid is of great interest to the bioremediation industry. MhpC catalyses the hydrolytic cleavage of a C-C bond of 2-hydroxy-6-keto-nona-2,4-diene-1,9 dioic acid to yield 2-hydroxypenta-2,4-dienoic acid and succinic acid. Attempts to obtain phases by molecular replacement using a number of weak structural homologues proved unsuccessful. A retrospective analysis of MhpC using the co-ordinates from an independent multiple anomalous dispersion analysis suggests that a combination of factors colluded to prevent a successful molecular replacement solution, the most significant being a large structural difference after strand 6 (a series of helices and large connecting loops) between MhpC and the search models employed.

Contents

List of Figures

List of Tables

Chapter 4 Structural Studies of D303E Methanol Dehydrogenase from *Methylobacterium extorquens*

Chapter 5 Structural Studies of MhpC from *Escherichia coli*

Abbreviations

Å	Ångström (10^{-10} m)
a, b, c	unit cell dimensions (real space)
α, β, γ	depending on the context, unit cell angles or CCP4 Eulerian angles
ADH	alcohol dehydrogenase
AIP	acute intermittent porphyria
ALA	5-aminolaevulinic acid
ALAD	5-aminolaevulinic acid dehydratase
ALAS	5-aminolaevulinic acid synthase
BphD	biphenyl hydrolase
CCD	charge-coupled device
CCP4	Collaborative Computing Project Number 4
CD	circular dichroism
CMP-PA	cytidine monophosphorylphosphatidate
D303E-MDH	D303E methanol dehydrogenase
DAG	1,2-diacylglycerol
DDT	dithiothreitol
DEPC	diethylpyrocarbonate
DHP	2,3-dihydroxyphenylpropionic acid
D-Ins(1)P	D-inositol 1-phosphate
DPM	dipyrromethane cofactor
ESRF	European Synchrotron Radiation Facility
F1,6-BP	fructose 1,6-bisphosphatase
FFT	fast Fourier transform
Glu-6-P	glucose-6-phosphate
GDH	glucose dehydrogenase
HPD	2-hydroxypenta-2,4-dienoic acid
HPK	4-hydroxy-2-ketopentanoic acid
HOPD	2-hydroxy-6-oxo-(phenyl/chlorophenyl)hexa-2,4-dienoic acid
IMPase	myo-inosoitol monophosphatase
IP_3	inositol 1,4,5-triphosphate

IPP	inositol polyphosphate 1-phosphatase
k_{cat}	turnover number
K_d	dissociation constant
K_m	dissociation constant for substrate
KNDA	4-keto-nona-1,9-dioic acid
λ	wavelength
L-Ins(1)P	L-inositol 1-phosphate
MAD	multiwavelength anomalous dispersion
MDH	methanol dehydrogenase
MhpC	2-hydroxy-6-keto-nona-2,4 diene-1,9-dioic acid-5,6-hydrolase
MIR	multiple isomorphous replacement
MPD	2-methyl-2,4-pentanediol
NCS	non-crystallographic symmetry
PBG	porphobilinogen
PBGD	porphobilinogen deaminase
PDB	protein data bank
PEG	polyethylene glycol
PC	Patterson correlation
pH	$-\log_{10}$ hydrogen ion concentration
PHMB	*para*-hydroxymercuribenzoate
PI	phosphatidylinositol
pKa	acid dissociation constant
PMSF	phenylmethylsulfonyl fluoride
PPO	protoporphyrin IX oxidase
PQQ	pyrrolo-quinoline quinone
PIP_2	phosphatidylinositol 4,5-bisphosphate
θ_1, θ_2, θ_3	X-PLOR Eulerian angles
R167Q-uPBGD	R167Q ubiquitous porphobilinogen deaminase
RFP	2-hydroxy-6-keto-nona-2,4 diene-1,9-dicarboxylate
rms	root mean square
W3A1-MDH	methanol dehydrogenase from *Methylophilus* W3A1
WT-MDH	wild-type methanol dehydrogenase (*Methylobacterium extorquens*)
x, y, z	unit cell co-ordinates (real space)

Acknowledgements

I would like to thank my supervisors Dr Jon Cooper and Professor Steve Wood for their support and guidance throughout my research and for providing me with the opportunity to learn X-ray crystallography.

I am especially grateful to two people whose help and encouragement has made this thesis possible. Firstly, Dr Darren Thompson for teaching me both the theory and practical side of crystallography and secondly, Dr Raj Gill for allowing me to work on his crystallographic projects and for his assistance in the laboratory. I would also like to thank both for reading this thesis and providing useful comments.

I would like to express my thanks to Dr Maria Hidalgo Lara for her help in the laboratory.

I would like to thank my crystallography colleagues (past and present) for their help at various times and for providing a pleasant atmosphere at the workplace. This includes Gareth, Peter, Alun, Doug, Graham, Nicky, Alan, Leighton, Mark and Terry.

I wish to extend my gratitude to EPSRC for funding me for the duration of this PhD.

I am also very grateful to the following people;

Professors Peter Shoolingin Jordan, Chris Anthony and Muhammed Akhtar for useful discussions.

Tom, Paul, Rajji, Abeer and Julie for providing me with either purified protein or crystals.

Friends made in Southampton including Neil, Jo, Suzy, Kris, Jeff, Aisling, Mark, Andrew, Andy, Sanjay and Cheryl.

I am also grateful to my family for their support.

This thesis is dedicated to my wife Azra for her constant love, support and encouragement in the last three years.

General Introduction

General Introduction

The complete sequencing of several genomes (including *Escherichia coli*, *Haemophilus influenzae*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Homo sapiens*), have provided a wealth of information on the genes which control the function of a wide variety of organisms (Fleischmann *et al.*, 1995; Goffeau *et al.*, 1996). In order to understand the complex workings of biological systems, detailed knowledge of the correlation between the amino acid sequence and the structure and function of each protein is essential. This will aid in our comprehension of the molecular basis of metabolism and reproduction and shed light on how life has evolved. Accurate data will only be available when an adequate number of high resolution three-dimensional structures of proteins have been determined to establish the principles of protein folding. It is only at this stage that the correlation between amino acid sequence and the unique properties of each protein is feasible.

As our understanding of protein structure becomes more complete so will our ability to understand and predict protein structure. This knowledge will eventually lead to a greater insight into the functions of many broadly conserved yet presently uncharacterised genes. Moreover, powerful computer programs will be capable of predicting the structure and function of any protein from its primary sequence. Biochemical information at the atomic level will also likely provide drug designers with new targets for drugs and also help in eliminating those that have no pharmaceutical use.

Today it is well established that proteins can have similar folds even though their sequences may be considerably different (Pastore and Lesk, 1990). The number of protein sequences deposited in GenBank (> 600 thousand) far exceeds the number of distinct folds that have been identified to date. One of the major objectives of structural genomics research is to develop rapid techniques to provide structures and theoretical models for all proteins encoded in the genome. An intriguing approach to solving this problem is to use computer parsing algorithms to efficiently examine the complete distribution of protein structures in an analogous manner to the geneticists, who used DNA sequencing machines to decode entire genomes. To optimise the use of genomic data, a number of structural genomic centers have been established; the US National Institute of General Medical

Science has initiated projects on protein structure, whereby 10,000 three-dimensional protein structures, each representing a protein family will be solved in the next decade by seven regional groups.

The success of these projects will largely depend upon the development of new computational approaches. Currently, the rate determining step in structure determination by X-ray crystallography is in the initial protein crystallisation, since there is no simple correlation between properties of a protein and the large number of crystallisation parameters that have to be sampled. Companies such as the Protein Structure Factory in Berlin have been established to develop technology for accelerating virtually every aspect of the structural determination process, creating ever more efficient systems for cloning genes, expressing proteins, growing crystals and collecting X-ray data. Furthermore, the quality and speed of data collection has been improved by the exclusive use of synchrotron radiation and charged-coupled devices (CCD's), while the interpretation of data has been made easier and faster by better computers and software.

A major hurdle in the analysis of protein structures is in the determination of the experimental phases following data collection. In the past the phase problem was overcome either by tedious trial and error searches for isomorphous derivatives or by molecular replacement phasing if a suitable homologous protein was available. Recent advances in synchrotron radiation research have lead to an increasing number of structures being solved routinely by multi-wavelength anomalous dispersion (MAD) methods (Hendrickson and Ogata, 1997). The recombinant protein expressed in a methionine auxotroph of *E. coli* can be labelled with selenomethionine and subjected to MAD phasing. The latter stages of protein structure determination (ie model building and refinement) are also becoming increasingly automated (Lamzin and Wilson, 1997). As a consequence, the number of structures deposited in the Protein Data Bank (PDB) has risen exponentially each year (currently there are 14 thousand entries) and this growth will continue in the near future as initiatives set by the structural genomic centers are implemented.

In order to understand and make efficient use of the huge amount of data available, several classification schemes have been introduced including SCOP (Structural Comparison of

Proteins; Murzin *et al.*, 1995), CATH (C)lass, (A)rchitecture, (T)opology and (H)omologous superfamily; Orengo *et al.*, 1997) and FSSP (Fold classification based on Structure-Structure alignment of Proteins; Holm and Sander, 1996). The SCOP database is based on human expertise, providing a detailed description of the structural and evolutionary relationships between all proteins whose structure is known. In the latest SCOP release 1.53 (July 2000), there are 564 folds and 859 superfamilies that have been identified from the deposited entries. One of the earliest discoveries from SCOP was that the distributions among all levels of hierarchy are distorted; there are a number of frequently occurring folds that have been identified (ie ferredoxin-like fold, α/β-barrel and immunoglobulin fold are adopted by 19, 13 and 9 superfamilies, respectively) (Brenner *et al.*, 1997).

The existence of a finite number of protein superfamilies was initially proposed independently by Dayhoff (1974) and Zuckerkandl (1975), who both estimated a value around 1000. This was later confirmed by Cyrus Chothia (1992) based on results obtained from the yeast (Oliver *et al.*, 1992), nematode (Sulston *et al.*, 1992) and human genome projects (Adams *et al.*, 1992), and two analyses of known protein structures (Sander and Schneider, 1991; Pascarella and Argos, 1992). Proteins are grouped into families whose members have diverged from a common ancestor and so have similar folds. However, a more recent analysis suggests that the number of protein superfamilies ranges from 1000 to 6000 (Swindells *et al.*, 1998). As we achieve a more complete coverage of the expected repertoire of protein folds found in nature, it is inevitable that the method of molecular replacement (Rossmann and Blow, 1962) will become an ever more influential tool in three-dimensional structure analysis.

The molecular replacement method makes use of a structural homologue (search model) to derive initial phases for a new structure (target). The success of this method is heavily dependent on the extent of structural similarity between the search model and the target protein (higher the better). The maps generated by the Fourier transformation of the diffracted intensities of the target molecule and the putative structural homologue (using the Patterson function) for which three-dimensional co-ordinates are already available in the PDB, are compared. These maps of interatomic vectors (the Patterson map) enable the relative orientation and position of the target structure with respect to the search model to

be determined in the target asymmetric unit. The unknown structure can then be solved following alternating cycles of refinement and model building.

In this thesis I describe the use of the molecular replacement method to determine the three-dimensional structures of four enzymes, namely, human ubiquitous R167Q porphobilinogen deaminase, bovine inositol monophosphatase, D303E methanol dehydrogenase from *Methylobacterium extorquens* and 2-hydroxy-6-keto-nona-2,4 diene-1,9-dioic acid-5,6-hydrolase from *Escherichia coli*. The problems encountered for each of the enzymes vary in terms of the size of the protein (from 29 to 149 kDa), the number of molecules present within the asymmetric unit (2 or 4 molecules), the quality of the X-ray data (from 3.5 to 1.65 Å) and the quality of the search model available (sequence identity to search model ranges from 19 to 100 %). Each of the proteins is of specific interest in its own field, but as a group, their successful structure analysis demonstrates the power of the molecular replacement method.

References

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) *Nature*. **355** 632-634

Brenner, S.E., Chothia, C. and Hubbard, T.J.P. (1997) *Curr. Opin. Struc. Biol.* 7 369-376

Chothia, C. (1992) *Nature*. **357** 543-544

Dayhoff, M.O. (1974) *Feder Proc* **33** 2314-2316

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A. and Merrick, J.M. (1995) *Science* **269** 496-512

Goffeau, A., Barrell, B.G., Bussey, H., Davies, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Adn, E.J., Louis, M.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) *Science* **274** 563-567

Hendrickson, W.A. and Ogata, C.M. (1997) *Methods Enzymol.* **277** 131-157

Holm, L. and Sander, C. (1996) *Science* **273** 595-602

Lamzin, V.S. and Wilson, K.S. (1997) *Methods Enzymol.* **277** 269-305

Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) *J. Mol. Biol.* **247** 536-540

Oliver, S.G., Vanderaart, Q.J.M., Agostonicarbone, M.L., *et al.* (1992) *Nature* **357** 38-46

Orengo, C., Michie, A., Jones, S., Hutchinson, E. and Thornton, J. (1997) *Structure* **5** 1093-1108

Pascarella, S. and Argos, P. (1992) *Prot. Eng.* **5** 121-137

Pastore, A. and Lesk, A.M. (1990) *Proteins Struct. Funct. Genet.* **8** 133-155

Rossmann, M.G. and Blow, D.M. (1962) *Acta Crystallogr.* **15** 24-31

Sander, C. and Schneider, R. *Proteins.* **9** 56-68

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. and Waterston, R. (1992) *Nature* **356** 37-41

Swindells, M., Orengo., C., Jones, D., Hutchinson, E. and Thornton, J. (1998) *Bioessays* **20** 884-891

Zuckerkandl, E. (1975) *J. Mol. Evol.* 7 1-57

Chapter 1

X-ray Crystallographic Methods

1.1 An Overview of X-ray Crystallography

X-ray crystallography is an experimental technique used to determine the three-dimensional arrangement of atoms in a protein molecule. An overall outline of structure determination, from protein expression to the final structure is illustrated in Figure 1.1. The technique depends on overcoming three significant problems. Firstly, sufficient amounts of the protein of interest must be purified to facilitate the initial crystallisation trials. Secondly, the protein must be crystallised and diffract to a resolution that will allow atomic detail to be revealed. Finally, the phases must be determined for each reflection in order to calculate an electron density map. The techniques used in protein expression and purification are beyond the scope of this thesis and thus will not be mentioned in great detail.

1.2 Crystallisation of Proteins

1.2.1 Introduction

Currently, the rate limiting step in protein crystallography is in obtaining protein crystals that are well ordered and diffract X-rays to enable the structure to be solved to a reasonable resolution. Crystals are grown by bringing the protein solution to a supersaturated state which may develop into a crystalline or amorphous phase. The formation of well ordered aggregates is the first phase of crystallisation and is called the nucleation phase. The nucleation phase is succeeded by the growth phase, in which an increasing number of protein molecules are recruited to the growing faces of the crystal, resulting in the formation of a large well ordered crystal. Finally, the last phase of crystallisation is the cessation of crystal growth resulting from the propagation of growth defects, the presence of contaminants and by depletion of protein molecules (Ducruix and Giegé, 1992).

Crystal growth is often illustrated by way of a phase diagram, which is divided into saturated and unsaturated regions (Figure 1.2). Within the unsaturated region a protein will never crystallise. In contrast, the saturated region supports crystallisation and is subdivided into three distinct zones (precipitation, nucleation and metastable). In the precipitation zone, the protein partitions from the solution into an amorphous state. Within

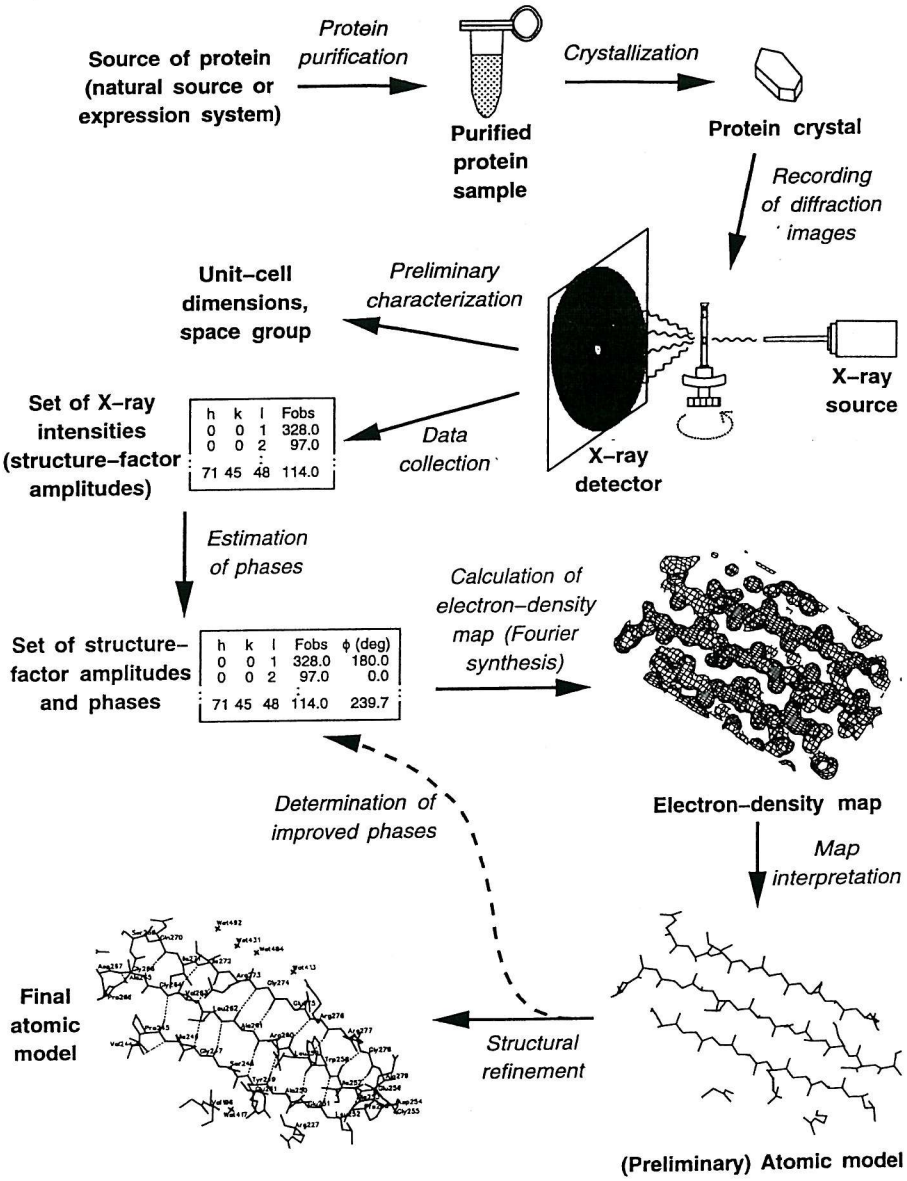


Figure 1.1: An overview of protein structure determination by X-ray crystallography (Louie, 1996). The protein, either isolated from its natural source or obtained by expression in a bacterial system or, is purified to allow the initial crystallisation trials. The crystals that are grown are exposed to an X-ray beam. The resulting diffraction patterns are processed (determine the intensity of each diffraction spot) to initially provide information about the unit cell dimensions (size of the repeating unit that forms the crystal) and space group (crystal packing symmetry). For each reflection (hkl), structure factor amplitudes are combined with estimated phases, to calculate an initial electron density map. The electron density map can then be interpreted to allow the molecular structure to be modelled. The resulting structure is subsequently refined to fit the map more accurately and to adopt a thermodynamically favoured conformation. Alternative cycles of model building and refinement are performed until atoms in the molecular structure best fit the data (final model).

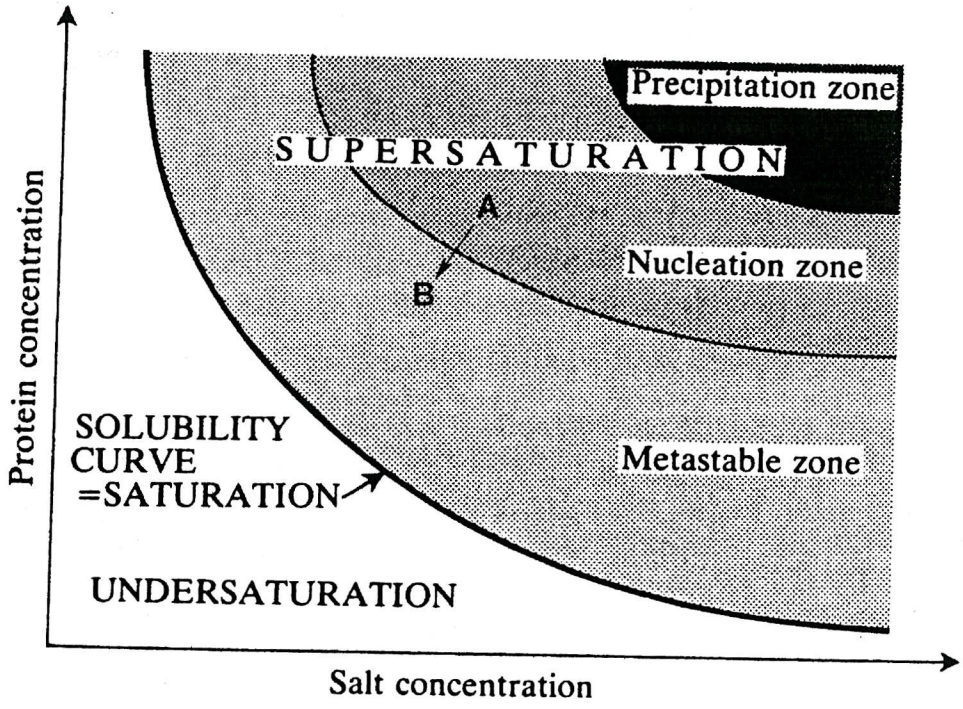


Figure 1.2: A two-dimensional solubility diagram illustrating the affect of varying the protein and salt concentration on crystal growth. Proteins fail to crystallise while in the unsaturated region. The saturated region supports crystal growth and is subdivided into three distinct zones (precipitation, nucleation and metastable region). The growth of stable nuclei and hence partitioning of protein from solution into a crystalline form occurs in the metastable and nucleation zones. Spontaneous formation of stable nuclei can only be supported in the nucleation zone and the rate at which this occurs depends on how far into the zone the system is coerced. The optimum strategy in growing large crystals therefore involves coercing the system beyond the metastable zone (A). As crystal growth occurs within this region, the protein solution is depleted and returns to the metastable state (B), where stable nuclei formation is diminished and the few nuclei previously initiated, grow to a large size.

the nucleation zone, spontaneous formation and growth of stable nuclei results in the partioning of protein from solution into the crystalline state. Finally, within the metastable zone, stable nuclei can grow but cannot be initiated. As a result, the protein solution remains supersaturated for long periods without nucleation unless it is mechanically shocked or a seed crystal is introduced (McPherson, 1990).

The optimal strategy for obtaining crystals can therefore be achieved by coercing the system as far into the nucleation zone as possible without causing precipitation. Here, nucleus formation and speed of growth is greatest, resulting in the formation of an extensive and uncontrolled shower of crystals. Crystals grown in such a manner are often small and contain defects due to the rapid speed of crystal growth. Ideal crystal growth requires nucleation to occur just beyond the metastable zone in which growth takes place and the protein solution, as a result of depletion, returns to the metastable state where stable nucleus formation is diminished (Figure 1.2). Hence the few nuclei already initiated, grow to a large size and at a pace free from defect formation.

1.2.2 Factors Influencing Protein Crystallisation

Several parameters influence nucleation and crystal growth by reducing the solubility of a protein and bringing the solution to a supersaturated state. An important factor affecting protein solubility is pH since the solubility of a protein is at its lowest at the isoelectric point. Other factors that affect the solubility of a protein include temperature, ionic strength and the presence of organic solvents; these parameters are often varied to induce nucleation or optimise crystal growth.

Solutions with a low ionic strength induce protein molecules to interact among themselves in order to balance their electrostatic charges. By adding salt ions to a solution of low ionic strength, the solubility of the protein increases and this is commonly known as the 'salting in' phenomenon. By increasing the ionic strength of the solution further, the added salt ions start to compete with each other and the protein for water molecules which gradually reduces the solubility of the protein. This phenomenon is referred to as the 'salting out' effect and ammonium sulphate is frequently used in this capacity in protein crystallisation.

Organic solvents (ethanol or 2-methyl-2,4-pentanediol (MPD)) and polymers such as polyethylene glycol (PEG) have a similar dehydrating affect on protein solutions. In addition, they can reduce the dielectric constant of solutions, and consequently increase the strength of electrostatic forces and promote self association. Moreover, PEG has the ability to restructure the natural solvent, which reduces the solubility of the protein further by volume exclusion (Blundell and Johnson, 1976). The addition of substrates, coenzymes and inhibitors may aid in protein crystallisation by allowing the protein to take a more stable form. The inclusion of metal ions essential for enzyme activity can promote protein crystallisation by stabilising structural features of the protein (McPherson, 1990). Protein purity can also affect crystallisation as poor purity is often associated with unsuccessful crystallisation or yields crystals that are small, disordered and diffract X-rays badly (Blundell and Johnson, 1976).

1.2.3 Techniques Employed to Crystallise Proteins

There are a number of methods available for crystallising proteins (vapour diffusion, batch, equilibrium dialysis), all of which aim to bring the protein solution to a supersaturated state. With many of these methods, the parameters affecting protein crystallisation can be varied with ease. These techniques have been reviewed elsewhere (Blundell and Johnson, 1976; McPherson, 1982) and will not be dealt with here. The vapour diffusion method is the most common technique used for crystallising proteins and therefore will be described in detail.

In the vapour diffusion method, crystals are grown either in a sitting or a hanging drop. In the latter case, equal volumes of protein solution (typically 2-5 μl) and reservoir solution containing the crystallisation agents (precipitant, salt, buffer), are mixed on a siliconised coverslip. The coverslip is inverted and placed on a greased well over the reservoir solution, thus producing a closed system. The precipitant concentration in the drop is less than in the reservoir solution, therefore water diffuses from the drop (via vapour diffusion) until equilibrium is achieved between the drop and reservoir. The protein concentration in the drop slowly increases, eventually leading to a supersaturated state from which crystal growth occurs. The technique is simple, inexpensive and effective in growing large crystals (Rhodes, 1993).

1.3 Data Collection and Processing

1.3.1 Introduction

Once well ordered crystals of a suitable size have been grown, they can be mounted using an appropriate method, depending on the temperature at which data is to be collected. Data collection involves exposing crystals to an X-ray beam and recording the intensities of the resulting Bragg reflections. The diffracted reflections have to be measured as rapidly and efficiently as possible, since protein crystals deteriorate in the X-ray beam. The instruments used in data collection include X-ray sources, detectors and cameras. An X-ray source produces a narrow beam of radiation which passes from the collimator through the crystal mounted on a goniometer head. This allows the crystal to be positioned at different orientations in the beam and is monitored by a camera. The diffracted X-rays can then be recorded using image plates, multiwire detectors or CCD cameras. Once a full data set is collected, it is processed, which involves calculating structure factor amplitudes from the measured intensities after applying a number of corrections.

1.3.2 X-ray Sources

X-ray sources used in protein crystallography include rotating anode generators and synchrotrons. X-rays from the former are generated by bombarding a metal target (usually copper) with electrons, produced by heating a filament and accelerated by an electric field. The spectrum of X-rays produced by a copper rotating anode consists of a continuous background of white radiation with two characteristic peaks (Figure 1.3). These peaks arise from high energy incoming electrons colliding with and displacing K shell electrons from the target metal atom. An electron from a higher orbital, emitting its excess energy as an X-ray photon, subsequently fills the K shell vacancy.

The two copper peaks are labelled K_α and K_β depending on the orbital from which the electrons fall into. The K_β peak is produced as a result of an M-K transition. In contrast, electrons from the L shell de-excite to the K shell vacancy and produce the K_α peak. The K_α peak (λ=1.5418 Å) is the most useful part of the spectrum and the rest of the spectrum is eliminated or minimised (Whittaker, 1981). This can be achieved by monochromating

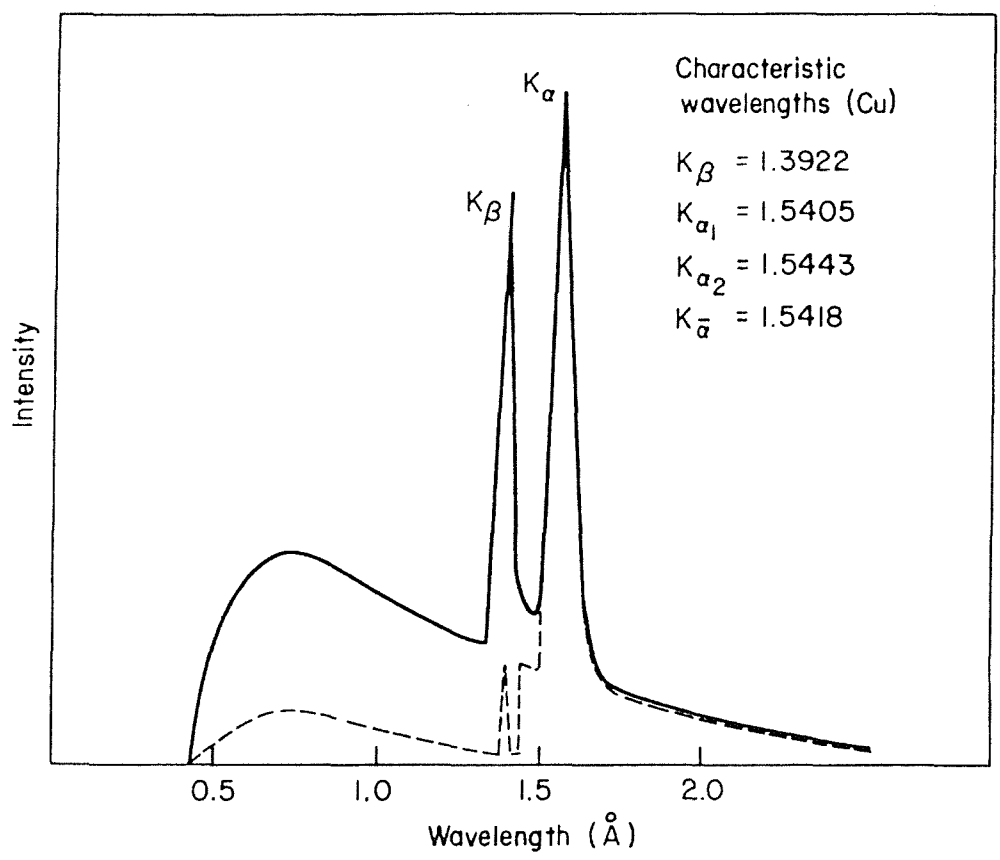


Figure 1.3: X-ray spectrum emitted by a tube with a copper target before (-) and after (---) passage through a nickel filter (Blundell and Johnson, 1976).

the primary X-ray beam using monochromaters or filters. Since the absorption edge of an element shifts to longer wavelengths as the atomic number decreases, nickel has been successfully employed in selectively reducing the K_β peak when using copper as the metal target. Cleaner radiation can also be obtained using a graphite monochromater, which allows only a very narrow band of wavelengths to be reflected thus producing a monochromatic beam (Drenth, 1994).

Synchrotrons are the most powerful X-ray sources, consisting of electrons or positrons circulating in giant rings at velocities near the speed of light. This circular motion is maintained by the presence of powerful magnets. These electrons emit energy (synchrotron radiation) in the form of X-rays when forced into a curved motion. Accessory devices such as wigglers and undulators give additional bending of the beam, producing higher intensity radiation at shorter wavelengths (Rhodes, 1993). This increased intensity allows the collection of high resolution data using short exposures. Furthermore, with synchrotron sources, specific wavelengths can be selected which can aid in determining the structure of a protein.

1.3.3 Image Plates

Image plate scanners are routinely used to collect X-ray diffraction data. The MAR scanner has a plate radius of 90 or 150 mm. The surface of the image plate consists of barium fluoride doped with europium ions (Eu^{2+}), which are excited into a metastable state by X-rays (Eu^{2+}-Eu^{3+}). The scanner then uses a red He-Ne laser to induce light emission from the europium atoms. The intensity of the emitted light is proportional to the number of absorbed X-rays and can be amplified by a photomultiplier system and measured. The image plate is subsequently erased by a flash of visible light, ready for the next exposure (Dauter, 1993).

1.3.4 Rotation Method

The rotation method (Wonacott, 1977), involves rotating the crystal through a small angle about an axis perpendicular to the X-ray beam and the diffracted X-rays are recorded on an image plate placed perpendicular to the incident beam. As the crystal is rotated, a

limited number of reciprocal lattice planes pass through the Ewald sphere and their intensities are recorded. As the reciprocal lattice plane passes through the sphere of reflection, the crystal is oscillated back and forth in order to average out any beam fluctuations accurately. Data collection consists of rotating the crystal through a small angle, recording all reflections, then rotating the crystal through another small angular step to measure additional reflections. This continues until a full dataset is collected. Partially recorded reflections (ie those recorded on two adjacent images) can be added together to give accurate intensity values (Giacovazzo, 1992).

1.3.5 Data Collection at Cryogenic Temperatures

Crystals exposed to X-ray radiation at room temperature (especially from very intense synchrotron sources) often suffer from radiation damage. Cryocooling is often used to increase the lifetime of the crystal since free radical damage is reduced at lower temperatures. The process involves soaking the crystal in mother liquor containing a cryoprotectant, and then flash freezing to cryogenic temperatures (100 K). At this temperature the liquid surrounding the crystal is transformed into glass, a process known as vitrification. The most commonly used cryoprotectant includes glycerol, ethylene glycol, PEG 400 and MPD. The advantages for collecting data at 100 K include an increased signal-to-noise ratio, fewer systematic errors and gentler crystal mounting techniques. Moreover, the prospect of storing and transporting crystals allows optimum utilisation of synchrotron beam time. The disadvantages associated with the method include increased mosaicity and the time spent in finding the appropriate cryoprotectant conditions (Garman and Schneider, 1997).

1.3.6 Data Processing

Data processing can be divided into several stages:

1 Detector calibration

2 Determination of crystal orientation, unit cell dimensions and Bravais lattice

3 Prediction of expected reflection positions

4 Intensity integration by profile fitting

5 Application of Lorentz, polarisation and absorption corrections

6 Sorting of reflections on the basis of their indices

7 Scaling of images to a common scale

8 Merging of symmetry related reflections

9 Conversion of intensities to structure factor amplitudes

Programs for processing raw image plate data include MOSFLM (Leslie, 1992, 1994), DENZO (Otwinowski, 1993) and XDS (Kabsch, 1988 a,b). During data processing the relative intensities of symmetry related reflections are used as a measure of data quality. This is known as the merging R-factor and can be expressed as

$$R_{merge} = \frac{\sum_{hkl} | I - I_{mean} |}{\sum_{hkl} | I_{mean} |} \qquad \text{Equation 1.1}$$

where I is the intensity of the reflection and I_{mean} is the mean intensity. For crystals yielding high quality data, the merging R factor is typically between 3 and 10 %.

1.4 Diffraction Theory

When a X-ray beam impinges on a crystal, it is the electrons surrounding the nuclei of atoms that scatter X-rays. The X-rays scattered in certain directions interfere constructively to give reflections which can be recorded on film. Mathematically, the description of X-ray scattering is best described as a Fourier summation. The amplitude of each reflection is described by a structure factor F_{hkl}; each term within F_{hkl} describes the contribution of one atom to the reflection hkl and can be expressed as

$$F_{hkl} = \sum f_j \, e^{2\pi i(hxj + kyj + lzj)} \qquad \text{Equation 1.2}$$

The contribution of atom j to F_{hkl} depends on the element (ie its atomic structure factor (f_j)) and on the position of atom j within the unit cell (x_j, y_j, z_j). Thus the atomic scattering factor and the position of atom j contribute to the net amplitude and phase information.

Because electron density is a complicated periodic function, it can also be described as a Fourier series and expressed as

$$\rho_{(x,\,y,\,z)} = 1/V \sum_{h}\sum_{k}\sum_{l} F_{hkl}\, e^{-2\pi i(hx+ky+lz)} \qquad \text{Equation 1.3}$$

Therefore electron density can be calculated by constructing a Fourier series using structure factors F_{hkl}. Because F_{hkl} is a periodic function describing a complicated wave it consists of a frequency, amplitude and phase. The frequency is that of the X-ray source since it is describing a diffracted X-ray. The amplitude is proportional to $(I_{hkl})^{1/2}$, the square root of the measured intensity I_{hkl} for reflection hkl. The phase is unknown and is required in order to calculate an electron density map and obtain the position of atoms in the protein molecule. The so called 'phase problem' can be overcome by various methods which include molecular replacement, multiple isomorphous replacement (MIR), MAD and direct methods.

1.5 Molecular Replacement

1.5.1 Introduction

Phases calculated from structure factors of a known homologous protein (search model) can be used as initial estimates for the unknown protein (target), providing that the approximate orientation and position of the target molecule is known. This method is known as molecular replacement and involves placing the search model in the target cell. The phases from atomic structure factors used to calculate F_{hkl} are highly sensitive to the location of atoms in the unit cell. By superimposing the search model with the target protein in its unit cell, the phases from the correctly orientated and positioned model can be calculated, and used as initial estimates for the target protein. The method involves determining the position and orientation of the search model in the unit cell with respect to the target protein (Figure 1.4).

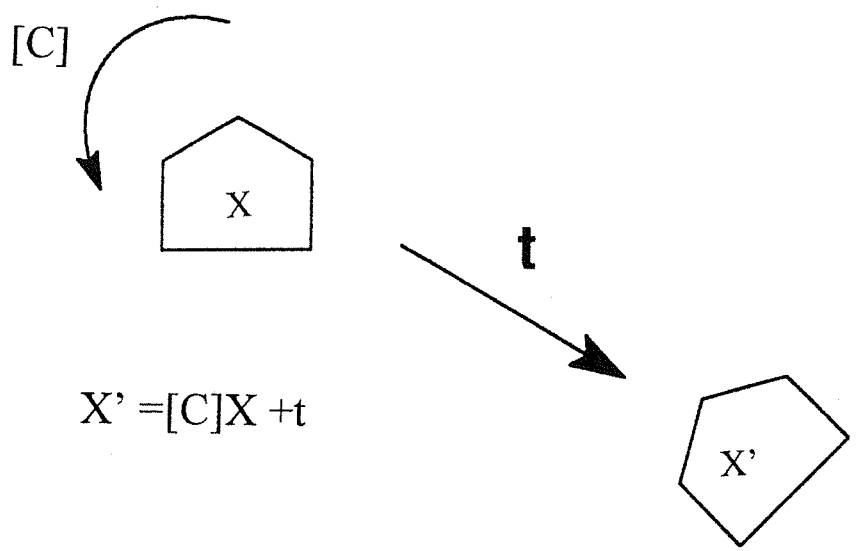


Figure 1.4: Diagram illustrating the molecular replacement method by superimposing the known structure (X) on the target structure (X') in its unit cell. The transformation of $\underline{X}$ to $\underline{X}$' is described by $\underline{X}$'=[C]$\underline{X}$ +$\underline{t}$ where [C] is rotation matrix that rotates co-ordinates X into a new orientation and $\underline{t}$ is the translation vector defining the position. The success of the procedure depends on determining 3 orientational parameters (rotation function) and 3 translational parameters (translation function).

If X represents the matrix of position vectors of the search model and X' the matrix of position vectors of the target, the transformation between them can be described by

$$\underline{X'} = [C]\underline{X} + \underline{t} \qquad \text{Equation 1.4}$$

where [C] is a matrix that rotates co-ordinates X into the new orientation and t is a vector defining the translation. The rotation matrix consists of three parameters usually defined as Eulerian angles α, β and γ (or θ_1, θ_2, and θ_3), while the translation vector (t) is defined by 3 translational parameters (t_x, t_y, t_z). In principle it is possible to search for all six parameters at the same time but in practice this is time consuming, even for the fastest computer. Rossmann and Blow (1962) pioneered the solution to this problem, by demonstrating that the search for the orientation of similar subunits in a crystal cell could be achieved by using the Patterson function (see section 1.5.2). Consequently, the six parameter search is split into two parts, a rotation function and a translation function. The success of molecular replacement depends on several factors which are discussed in greater detail in section 1.5.6.

1.5.2 Patterson Function

The Patterson function (first reported by A.L. Patterson in 1934) is the Fourier transform of the experimental intensities (ie it requires no phase information) and can be expressed as

$$P_{(u,v,w)} = 1/V \sum_h \sum_k \sum_l |F_{hkl}|^2 \; e^{-2\pi i(hu+kv+lw)} \qquad \text{Equation 1.5}$$

The Patterson map that is generated consists of peaks corresponding to all interatomic vectors. Each pair of atoms in the real unit cell give rise to two peaks in the Patterson cell; for N atoms in the unit cell there would be N^2 peaks in the Patterson cell. The Patterson cell also contains an origin peak which consists of vectors of zero length, corresponding to vectors between each atom and itself. Therefore, the Patterson unit cell consists of N^2-N non-origin peaks (Figure 1.5). By increasing the number of atoms in the real cell, the

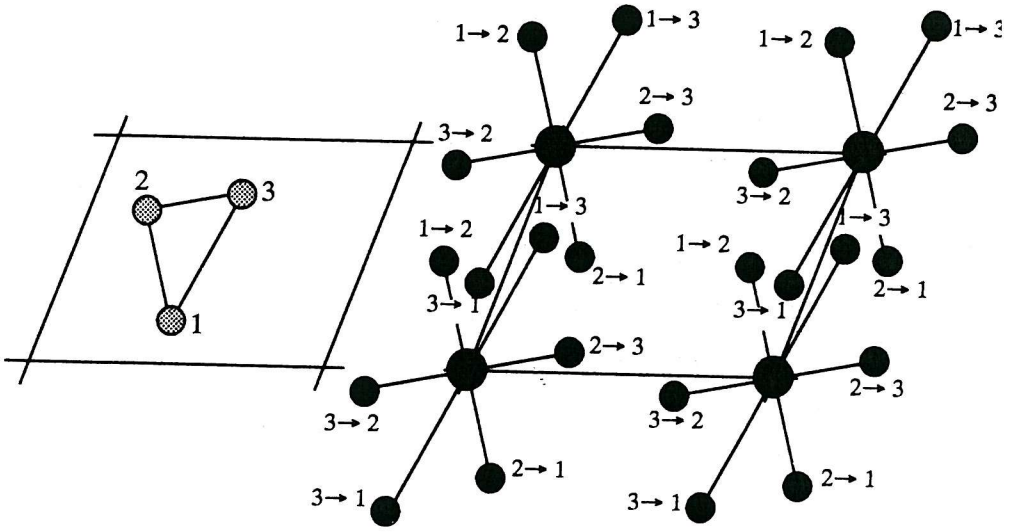


Figure 1.5: An illustration of the Patterson function (Drenth, 1994). The two-dimensional real cell containing three atoms (a) and its corresponding Patterson cell (b) is shown. The Patterson cell consists of peaks corresponding to all interatomic vectors. Each pair of atoms in the real unit cell give rise to two peaks in the Patterson cell (N^2 where N= number of atoms in the real cell); for 3 atoms in the unit cell there are 9 peaks in the Patterson cell. The Patterson cell consists of N vectors of zero length at the origin (corresponding to vectors between each atom and itself) and therefore contains N^2-N non-origin peaks (6 for this example).

corresponding Patterson cell becomes more densely packed and hence more difficult to interpret.

1.5.3 Rotation Function

1.5.3.1 Introduction

The rotation function may be evaluated in real (Patterson) or reciprocal space with differences existing in the numerical approximations that are employed. During the rotation searches the Patterson function or its reciprocal space analogue for the observed intensities and that of the search model are compared. Rotation functions are normally evaluated on a search grid in Eulerian or polar angles covering the entire region of rotational space or the region of interest. Grid points that coincide with the correct angular parameters are generally expected to give high rotation function values. At other grid points the calculation of the rotation function values provide estimates for the background noise.

1.5.3.2 Real Space Rotation function

The Patterson function in real space is computed for the target (P_x) and search model (P_m) by the fast Fourier transform (FFT) at a specific resolution. In general, the search for maximum overlap is achieved by rotating one Paterson map with respect to the other, with the product of the two Patterson functions being significant when the two Patterson maps superimpose (Equation 1.6).

$$R(\Omega) = \int_u P_x(\underline{r}).P_m(\Omega r)\, dV \qquad \text{Equation 1.6}$$

where (Ω) = rotation matrix described by 3 Eulerian angles

$R(\Omega)$ = product function

r = integration variable

U = volume of integration

The computational time is vastly reduced by selecting the strongest grid points of the search model Patterson (P_m) in the integration, and calculating the value of the target Patterson (P_x) at these grid points by interpolation (Huber, 1965; Steigman, 1974). The integration is normally performed over a spherical shell that removes the origin peak and Patterson vectors of a specific length. By judiciously choosing the integration variable (r), the spherical shell region should contain a maximum number of intramolecular vectors and the fewest possible intermolecular vectors. Because the numerical integration of equation 1.6 can be imprecise, Patterson maps are prone to Fourier series truncation errors when computed at 3 to 4 Å resolution. As these errors are not restricted to the origin, the elimination of short Patterson vectors at the origin fails to completely remove noise ripples caused by the truncation, although the error contribution is reduced (Brünger, 1997).

1.5.3.3 Reciprocal Space Rotation Function

If the Patterson functions of the target and search model are expressed as the Fourier transforms of their individual diffraction intensities, the reciprocal space expression for the rotation function is obtained (equation 1.7). The evaluation of equation 1.7 is time consuming and therefore it is often referred to as the slow rotation function.

$$R([C]) = \sum_h \sum_p F(h)^2 F_m(p)^2 G_{h,p} \qquad \text{Equation 1.7}$$

where F(h) = the Fourier coefficients of the target crystal

$F_m(p)$ = the coefficients of the Fourier transform of the search model, rotated by C

p = represents a point in reciprocal space of a continuous transform

Finally $G_{h,p}$ is the interference function whose magnitude depends on h, p and the volume used in the integration (equation 1.8).

$$G_{h,p} = \frac{3(\sin 2\pi H R - 2\pi H R \cos 2\pi H R}{(2\pi H R)^3} \qquad \text{Equation 1.8}$$

The interference function represents the Fourier transform of a sphere of radius R. The length of the reciprocal space vector h+p[C] is given by H. The interference function

assumes the maximum value of 1 when h+p[C] = 0, hence only those points that are close together in the two reciprocal lattices being compared are significant in the double summation. Since the rotation function is proportional to the intensity of each reflection, the rotation function values are most likely dominated by the strongest reflections (also referred to as large terms). Consequently, by using only the large terms calculations for the slow rotation function are accelerated. When using a relatively fine search grid (intervals of 3-4°), it is unlikely that neighbours of a grid point will have a high rotation function value if the grid point itself has a low value. Hence, disregarding such neighbouring grid points can also speed up the slow rotation function calculations (Tong and Rossmann, 1997).

The fast rotation function developed by Crowther (1972) is an expansion of the slow rotation function in terms of spherical harmonic and Bessel functions. By utilising the special properties of the spherical harmonics, the rotation function is reduced to a series of two-dimensional Fourier summation. The evaluation of the two-dimensional summation can be achieved with the FFT, thus greatly accelerating the rotation function calculation. The large term approach can also be used for the fast rotation function, although more terms will be required in comparison to the slow rotation function due to the poor convergence properties of spherical Bessel functions.

1.5.3.4 Direct Rotation Function

The direct rotation function (DeLano and Brünger, 1995) differs from other rotation functions because the search model is placed in a P1 box with cell dimensions and angles equal to those of the target crystal structure. The search model is rotated rather than the corresponding Patterson map, and the criterion for best fit is defined by the correlation coefficient between observed and calculated normalised structure factors (see section 1.5.6.6) determined at each trial orientation.

These calculations take up a considerable amount of computational time and hence the rotational space is normally sampled on a coarse grid. The precise orientation prior to the translation search can be obtained, by either repeating the rotation function using a fine grid or alternatively performing extensive Patterson correlation (PC) refinement on the highest rotation peaks (see section 1.5.5.2). The major advantage of this method is that all

intramolecular vectors of the search model are considered during the rotation function (rather than a subset), thus increasing the signal-to-noise ratio. This is particularly applicable for crystals with low solvent content and/or with a large number of molecules per asymmetric unit (Brünger, 1997).

1.5.4 Translation function

1.5.4.1 Introduction

Once the orientation of the search model has been determined, the translation function is performed to determine the position of the search model with respect to the undetermined protein structure. In general, when the position of the phasing model is coincident with the target protein, the intermolecular vectors between symmetry related molecules in one Patterson would largely correspond with the other Patterson. The translation function can be evaluated by several methods that include using the Patterson Function, by using the correlation coefficient between the observed and calculated structure factor amplitudes or by incorporating phase information (phased translation function).

1.5.4.2 Real Space Translation Function

Determining the translation function in real space involves computing two Patterson maps, one from the observed intensities and the other from the search model using intermolecular vectors. The product of the two Patterson functions will be large when the search model has been correctly positioned with respect to the undetermined structure, giving maximum overlap. The translation function in real space can be expressed as

$$T(\underline{t}) = \int_V P_{obs}(\underline{u})\, P_{calc}(\underline{u}, \underline{t})\, d\underline{u} \qquad \text{Equation 1.9}$$

where $P_{obs}(\underline{u})$ = Patterson map of the unknown structure

$P_{calc}(\underline{u}, \underline{t})$ = calculated Patterson map of phasing model after a translation t

1.5.4.3 Reciprocal Space Translation Function

Crowther and Blow (1967) expanded equation 1.9 in reciprocal space giving

$$T(\underline{t}) = \sum_{h} I_{obs}(h)\, F_{M1}\, F^{*}_{M2}\, e^{-2\pi i h,t} \qquad \text{Equation 1.10}$$

where I_{obs} = are the observed intensities (from the target structure)

F_{M1} = is the Fourier transform of molecule 1 of search model

F^{*}_{M2} = is the complex conjugate Fourier transform of molecule 2 of the search model

h = is the reciprocal space vector hkl

t = is the translation vector between molecule M1 and molecule M2

Since the translation function is independent of the intramolecular vectors, self vectors from the observed intensities can be subtracted, thus ensuring that the new T1 function (equation 1.11) is dominated by intermolecular vectors.

$$T1(\underline{t}) = \sum_{h} I_{cross}(h)\, F_{M1}\, F^{*}_{M2}\, e^{-2\pi i h,t} \qquad \text{Equation 1.11}$$

where I_{cross} = are the observed intensities with the intramolecular vectors omitted

The T1 function is a standard Fourier summation and can be calculated by use of an FFT algorithm. The T2 function implemented in TFFC (Driessen *et al.*, 1991) improves the signal-to-noise ratio by subtracting intramolecular vectors from the search model Patterson.

1.5.4.4 The R-factor Search

This method of determining the translational parameters used to be popular because of its simplicity. It involves calculating an R-factor (or testing the agreement) between the observed data and the structure factors calculated from the search model (equation 1.12) placed at different locations in the determined unit cell. The correct position will give the

lowest R-factor; as the search model is moved throughout the asymmetric unit the R-factor varies only slightly, and therefore a sudden drop of a few percent is likely to be significant. The R-factor search requires a lot of computational time in comparison to the conventional translation function and can be sensitive to orientation errors in the search model.

$$\text{R-factor} = \frac{\sum_{hkl} \left| \left| F_{obs} \right| - k \left| F_{calc} \right| \right|}{\sum_{hkl} \left| F_{obs} \right|} \qquad \text{Equation 1.12}$$

where k is the approximate scale factor

1.5.4.5 The Correlation Coefficient

The correlation coefficient is an alternative parameter that can be used to measure the agreement between observed and calculated structure factors (equation 1.13), as the search model is moved in the undetermined cell. However unlike in the R-factor search, the correlation coefficients are independent of the scale factors between observed and calculated structure factors. In this case the correct position should give the highest correlation coeffecient.

$$CC = \frac{\langle \left| F_{obs} \right| - \left| F_{calc} \right| \rangle \langle \left| F_{obs} \right| \rangle - \langle \left| F_{calc} \right| \rangle \rangle}{(\langle \left| F_{obs} \right|^2 - \langle \left| F_{obs} \right| \rangle^2 \rangle \langle \left| F_{calc} \right|^2 - \langle \left| F_{calc} \right| \rangle^2 \rangle)^{1/2}} \qquad \text{Equation 1.13}$$

1.5.5 Molecular Replacement Packages

1.5.5.1 Introduction

Several program packages are available for calculating rotation and translation functions including X-PLOR (Brünger, 1992), AMoRe (Navaza, 1994) and MOLREP (Vagin and Teplyakov, 1997). The algorithms used in these packages will be described in greater detail in this section.

1.5.5.2 X-PLOR

Rotation searches in X-PLOR are performed in Patterson space employing the real-space Patterson search method of Huber and Steigmann (1985). First, a Patterson map from the observed intensities (P2) is computed by FFT on the specified grid. Next, the Patterson map for the search model (P1) is computed in a P1 box. During the cross rotation search, Patterson P1 is rotated with respect to Patterson P2. The vectors selected in Patterson space depend on the value of the Patterson map at the particular grid point (threshold) and the distance from the grid point to the origin (range). From this selected subset, only the strongest Patterson vectors are chosen for the cross rotation search. The selected Patterson vectors are rotated using pseudo-orthogonal Eulerian angles as defined by Lattman (1972).

Once the rotation function is completed, PC refinement is used to optimise the search-model orientation prior to the translation search (Brünger, 1990). The PC refinement procedure is useful because it can identify putatively correct solutions from the rotation function, improve the search model and reduce the large number of orientations to be checked by the subsequent translation function. The procedure involves selecting a large number of rotation function peaks (it is assumed that the correct orientation is present in the selected subset of peaks). In the next and most important step, a number of parameters (p) best describing the expected structural dissimilarities between the search model and the crystal structure (ie large rigid groups, domains, secondary structural elements), are introduced. These parameters (p) are refined against a negative correlation coefficient (-PC) for each selected orientation. Only the correct orientations are expected to produce a large correlation coefficient after PC refinement.

The PC refined orientation can then be used as input in the translation function. However unlike AMoRe, the translation function in X-PLOR can only perform the calculations one orientation at a time. The translation search can be evaluated with correlation coefficients between Fs, Es, F^2s, and E^2s with the standard residual, with the vector residual or with the packing function. The search is performed over a three-dimensional grid that can be specified in fractional or orthogonal co-ordinates. The search routine computes the structure factors of the translated primary molecule and the symmetry related molecules by

applying appropriate phase shift operators in reciprocal space to the calculated structure factors of the original (un-translated) molecule and its symmetry mates.

1.5.5.3 AMoRe

The AMoRe package allows the calculation of a large number of rotation and translation function solutions in a fast automatic manner. However, one of the biggest drawbacks of AMoRe is that it requires a lot of memory which may cause problems on some machines. Rotation searches in AMoRe are performed in reciprocal space using an improved fast rotation function which reduces errors by employing numerical integration rather than Bessel expansions (Navaza, 1987). Two preliminary programs (SORTING and TABLING) are run prior to the rotation function, the first packs and sorts the data into an internal format suitable for AMoRe, whereas the second adjusts the model co-ordinates to the optimal position and prepares a table (in an internal format) of continuous Fourier coefficients from the model co-ordinates. These continuous Fourier coefficients are then used in the structure factor calculations by interpolation (Navaza, 1994).

The next step involves calculating the rotation function using the program ROTING, which performs three distinct functions. Firstly, the structure factor amplitudes for the search model are calculated in a P1 box (the size of the box depends on the radius of integration, resolution of the data and the dimensions of the search model). Secondly, the spherical harmonic coefficients for the search model and target Patterson distributions are computed. Finally the rotation function is calculated, during which a number of parameters can be modified to improve the search. For instance, sharpening of the input temperature factors (by setting the BADD value to a negative value) or using normalised structure factors (E values calculated by the CCP4 program ECALC) may increase the signal-to-noise ratio of the correct solution.

The translation search in AMoRe can be evaluated using four different fast translation functions, performed by the program TRAING. Two of the translation functions are overlapping functions (based on Crowther and Blow (1967) and a reciprocal version of the full symmetry phased translation function (Cygler and Desrochers, 1989; Bentley and Houndusse, 1992)). The other two functions compute correlation coefficients; one is an

approximation to the correlation function, in terms of the intensities (based on the method of Harada *et al.*, 1989) and the other computes the exact overlaps, in terms of intensities (Navaza and Vernoslova, 1995). The program allows automated examination of multiple rotation function peaks in the translation search, which is beneficial when the correct orientation is not obvious following the rotation search. The program judges the correctness of each translation function solution based on the R-factor and correlation coefficient calculated between the model and target structure factor amplitudes.

1.5.5.4 MOLREP

The MOLREP package can perform all stages of molecular replacement automatically and requires only the X-ray data from the target crystal and the search model as input. Alternatively, the molecular replacement calculations can be performed separately allowing manual input of a number of parameters. In contrast to other molecular replacement packages, MOLREP possesses a number of unique features in its approach to data handling and molecular replacement calculations. For each trial orientation/position of the search model, structure factors are calculated by FFT of the electron density rather than via interpolation from a table of structure factors calculated on a fine grid, as in AMoRe. Although such direct calculations for structure factors are more accurate, the computational time is vastly increased.

The rotation and translation functions performed in MOLREP are essentially overlapping functions between the target and calculated model Pattersons. The algorithm used in the rotation search is the Crowther fast rotation function (1972), which is evaluated using FFT. The noise level in the rotation function is considerably reduced because the program computes the rotation function for 3 different orientations of the search model and averages them. The translation search algorithm (Vagin, 1989) is carried out in reciprocal space using FFT. The correctness of each translation function solution is evaluated by monitoring the R-factor and correlation coefficient calculated between the model and target structure factor amplitudes. The original full-symmetry translation function differs from other translation functions in two significant ways; firstly, the translation function for the symmetry operators are multiplied rather than summed therefore forming a map with

higher contrast, and secondly, by multiplying the translation function by the packing function, peaks corresponding to incorrect solutions with bad packing can be eliminated.

1.5.6 Factors Affecting Molecular Replacement

1.5.6.1 Introduction

Increased automation of the software means that molecular replacement can be a trivial exercise in cases where the search model has a high sequence homology with the target structure. However, as the sequence homology gets lower, solving a structure by molecular replacement becomes less straightforward. Depending on how complex the problem is, a number of factors have to be considered to ensure accurate structure determination by molecular replacement. The factors that have a pronounced affect on a molecular replacement solution are examined in greater detail below.

1.5.6.2 The Target Crystal

It is vital that the experimental data collected from the target crystal is complete (ideally 100 %). The presence of multiple copies of the protein molecule in the asymmetric unit is a major obstacle to successful molecular replacement solution. In such cases the traditional approach is to locate the first molecule, fix it and resume the search for the second molecule etc. However in such circumstances, the location of the first molecule is often difficult since the signal-to-noise ratio is reduced for the correct peaks owing to an increased number of vectors being present. In special cases, interpretation of the self rotation function by rotating P_{obs} against itself, provides non-crystallographic symmetry (NCS) information which can be used in a locked rotation/translation function search (Tong and Rossmann, 1990) to improve the signal-to-noise ratio.

1.5.6.3 The Search Model

For a successful molecular replacement solution the best available search model should be selected. Whenever possible, refined co-ordinates should always be used as this can often make the difference between obtaining a clear solution and no solution at all. Furthermore,

at least 30-40 % sequence similarity must exist between the search model and the target structure, although structural similarity (within 1.0-1.5 Å rms deviation) is more important. Since the inclusion of many wrong atoms in the model increases the overall noise level, the search model can be improved by removing parts that are known not to exist in the undetermined structure (eg loop regions and insertions). Similarly, side chains that are known to differ from the sequence alignment can be trimmed to alanine. However, a balance has to be maintained when trimming the search model, since removal of too many correct atoms will also reduce the signal one is searching for.

1.5.6.4 Radius of Integration and Size of the P1 Box

The calculated Patterson map contains peaks that correspond to both intramolecular and intermolecular vectors. The search for the best orientation is dependent on the intramolecular vectors, which tend to be shorter and closer to the Patterson origin. Therefore during the cross rotation function, the aim is to choose a sphere of radius (radius of integration) such that a sufficient number of intramolecular peaks are included, whilst minimising the number of intermolecular vector peaks. The choice of integration is a matter of trial and error but suggestions include using 75-80 % of the diameter of protein for a spherical molecule (Blow, 1985) or 75 % of the smallest orthogonal dimension of the molecule (Tickle, 1994).

The size of the search model unit cell has to be carefully selected to minimise the number of intermolecular vectors contributed by the search model. The search model is usually placed in a P1 unit cell, with dimensions chosen equal to the size of the model plus the radius of integration in each direction (Lifchitz, 1983). However, it is important to bear in mind that if the cell is too big then too many structure factors will be calculated.

1.5.6.5 Resolution Limits

The resolution limits determine the number of reflections that are included in the molecular replacement calculations and are selected on a trial and error basis. High resolution reflections (> 3.5 Å) are markedly different even between homologous structures since they reflect the precise conformation of residues, whereas low resolution reflections are

similar since they reflect the grosser features of the protein (ie relationship between secondary structural elements). In contrast, very low resolution reflections are heavily dominated by the arrangement of solvent and protein and hence reflections within the resolution range of 25-10 Å are normally omitted from the calculation. The highest resolution data should only be used when the structural differences between the target structure and the search model are small (Aguilar *et al.*, 1993). The suggested ideal high resolution cut off is approximately two or three times the expected mean co-ordinate difference (Blow, 1985).

1.5.6.6 Sharpening of Patterson Peaks

There are two methods available for sharpening Patterson peaks and hence increasing the contrast of the Patterson map, since Patterson peaks in general are much broader than electron density peaks. The first method involves temperature factor sharpening where high resolution intensities are scaled up by applying a negative temperature factor. In the second method, normalised structure factors are used by dividing data into resolution bins and scaling the data so that the mean intensity does not decrease with resolution. For both methods the contrast of the Patterson map is increased allowing overlapping peaks to be better resolved.

1.6 Refinement and Model Building

1.6.1 Introduction

Refinement is the process of improving the agreement between the calculated and observed structure factor amplitudes. This method is often referred to as reciprocal space refinement and involves adjusting atomic positions (x,y,z) to values that best fit the observed structure factor amplitudes. With least squares refinement, atom positions are selected that minimise the squares of differences between corresponding calculated and observed structure factor amplitudes. This is expressed by

$$\phi = \sum w_{hkl} \left(|F_o| - |F_c| \right)^2_{hkl} \qquad \text{Equation 1.14}$$

where ϕ = sum of squared differences between observed and calculated structure factor amplitudes

W_{hkl} = weighting of the difference depending on the reliability of the measured intensity.

The parameters to be refined include atomic positions, temperature factors (B-factors) and occupancy. A major problem associated with refining protein molecules is the large number of atoms present in the structure resulting in a poor observations-to-parameter ratio. Therefore, constraints or restraints are used in refinement to overcome this problem. With constrained refinement a fixed value is used for a certain parameter e.g. to fix bond lengths and angles, allowing only dihedral angles to vary thus, decreasing the number of parameters to be refined. In comparison, with restrained refinement elastic conditions are imposed on parameters e.g. all bond lengths and bond angles to be within a specific range of values. Therefore, inclusion of stereochemical information and non-bonded information such as Van der Waals interactions effectively increases the number of observations and helps to maintain the correct stereochemistry of the model.

The function ϕ is complicated because it exhibits many local minima corresponding to multiple model conformations. With least squares refinement, the first minimum close to the starting point is located, and therefore it is important that the starting model structure is close to the global thermodynamic minimum, otherwise refinement will converge to an incorrect local minimum. Alternatively, the simulated annealing protocol can be used which explores phase space, crossing barriers between minima by taking uphill steps by subjecting the model to high temperatures (2000 K-3000 K) using molecular dynamics (Brünger, 1987). Consequently, the radius of convergence is increased when compared with restrained least squares refinement and barriers separating large regions of phase space can be crossed, thus lifting the model out of local energy minima. The system is then cooled to restrict access to successively smaller regions of phase space until a configuration closer to global minimum is obtained.

There are a number of packages available for refining protein structures and these include RESTRAIN (Driessen *et al.*, 1989), REFMAC (Murshudov *et al.*, 1997), SHELXL (Sheldrick and Schneider, 1997) and CNS (Brünger, *et al.*, 1998). The CNS suite offers a

variety of refinement protocols (ie rigid body, positional, simulated annealing and temperature factor (overall, group or individual refinement).

1.6.2 Electron Density Maps and Model Building

During refinement, manual model building will be required, especially if the model has been refined to a local minimum. This involves calculating electron density maps so that the model can be rebuilt into density wherever necessary, thus allowing phase barriers to be crossed. Two types of electron density maps are normally calculated during model building. Maps calculated using *Fo-Fc* as the coefficient provides information about the new structure, whilst with a *2Fo-Fc* map, an approximation of the electron density of the model and a difference electron density map is obtained.

The Fourier transform allows the crystallographer to move back and forth between real and reciprocal space to nurture the model into agreement with the data (Figure 1.6). The initial phases can either be filtered in real space by density modification or filtered in reciprocal space by weighting the reflections according to the estimated accuracy of their phases. In real space, errors in the trial model can be removed by improving the fit between the model and map.

Electron density maps are phased either using MIR, MAD or from a model structure placed by molecular replacement. The initial experimental phases may be imperfect although the observed intensities will have little errors. The initial phases may not even be good enough to give an interpretable map and very often as a result will have to be refined. The phase angles can be improved by modifying electron density and carrying out it's Fourier transform to obtain better phases, hence finally producing a better map. The density modification process can be expressed as

$$\rho_{MOD}(\mathbf{r}) = M(\rho)\,\rho(\mathbf{r}) \qquad \text{Equation 1.15}$$

$\rho(\mathbf{r})$ = electron density,

$M(\rho)$ = modification function or physical constraint

$\rho_{MOD}(\mathbf{r})$ = modified electron density distribution

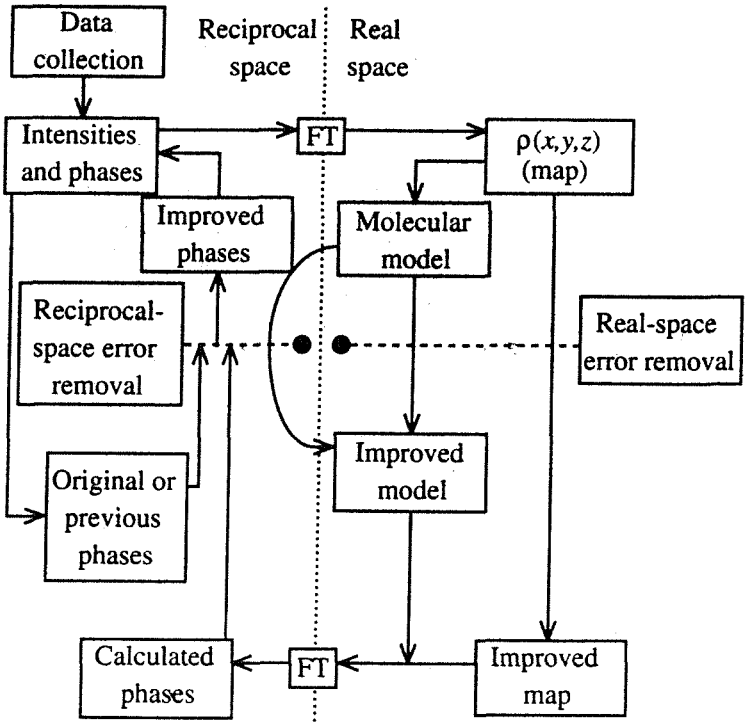


Figure 1.6: Diagram illustrating the use of the Fourier transform in refinement (Rhodes, 1993). The crystallographer can travel back and forth between real and reciprocal space during refinement and model building to nurture the model into agreement with the data. Errors in the initial phases can either be filtered in real space by density modification or filtered in reciprocal space by weighting the reflections according to the estimated accuracy of their phases. Errors in the molecular model can be removed in real space by improving the fit between the model and electron density map (model building).

Generally, all modification functions will attempt to enhance the correct features of an electron density map. The phases of the new electron density can then be combined with the observed intensities to produce another electron density distribution, which in turn can be modified, initiating a cyclic process. There are several methods available to remove the noise from the maps to make them easier to interpret and include NCS averaging (Bricogne, 1974), solvent flattening (Wang, 1985) and histogram matching (Zhang, 1990).

1.6.3 Convergence and R-factor

The improvement in the quality of the model during refinement can be followed by monitoring changes in the crystallographic R-factor (equation 1.12). As refinement converges to the correct structure, the agreement between observed (F_{obs}) and calculated (F_{calc}) structure factor amplitudes also improves. Values for the R-factor range from 0-0.6; an R-factor of 0.6 is obtained when a set of F_{obs} is compared with a random set of amplitudes. An early model with a R-factor near 0.4 is likely to improve when applying the various refinement protocols. For a protein model refined with data to 2.5 Å, the final R-factor should be around 0.2.

The R-factor is related to the residual being minimised and thus can be reduced by refining against an increased number of parameters, without improving the quality of the model. This problem is overcome by monitoring the free R-factor (R_{free}) in conjunction with the crystallographic R-factor (Brünger, 1992). The R_{free} gives an idea of the extent of any model bias and is calculated using 5 or 10 % of the total reflections, of which are not included in the refinement.

1.7 References

Aguilar, C.F., Newman, M.P., Aparicio Sanz, J., Cooper, J.B., Tickel, I.J. and Blundell, T.L. (1993) *Acta Crystallogr.* **A49** 306-315

Bentley, G.A. and Houdusse, A. (1992) *Acta Crystallogr.* **A48** 322-328

Blow, D.M. (1985) in *Introduction to rotation and translation functions* 8-13, Molecular Replacement, (Machin, P.A. comp), Proceedings of the Daresbury Study Weekend, S.E.R.C, Daresbury Laboratory, Daresbury, U.K

Blundell, T.L. and Johnson, L.N. (1976) in *Protein Crystallography*, Academic Press, New York

Bricogne, G. (1974) *Acta Crystallogr.* **A30** 395-405

Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) *Science.* **235** 458-460

Brünger, A.T. (1990) *Acta Crystallogr.* **A46** 46-57

Brünger, A.T. (1992) *Nature.* **355** 472-475

Brünger, A.T. (1997) *Methods Enzymol.* **276** 558-580

Brünger, A.T., *et al.*, and Warren, G.L. (1998) *Acta Crystallogr.* **D54** 905-921

Crowther, R.A. and Blow, D.M. (1967) *Acta Crystallogr.* **23** 544-548

Crowther, R.A (1972) in *The Molecular Replacement Method* 173-178, Ed. Rossmann, M.G., New York, Gordon and Breach

Cygler, M. and Desrochers, M. (1989) *Acta Crystallogr.* A45 563-572

Dauter, (1993) in *Data collection and processing* 119-122, Eds. Sawyer, L., Isaacs, N. and Bailey, S., S.E.R.C. Daresbury Laboratory, Daresbury, U.K

DeLano, W.L. and Brünger, A.T. (1995) *Acta Crystallogr.* **D51** 740-748

Drenth, J. (1994) in *Principles of protein X-ray crystallography*, Springer

Driessen, H., Haneef, M.I.J., Harris G.W., Howlin, B., Khan, G. and Moss, D. (1989) *J. Appl. Crystallogr.* **22** 510-516.

Driessen, H., Bax, B., Slingsby, C., Lindley, P.F., Mahadevan, D., Moss, D.S. and Tickle, I.J. (1991) *Appl. Crystallogr.* **B47** 987-997

Ducruix, A. and Giegé, R. (1992) in *Crystallization of nucleic acids and proteins. A practical approach,* Oxford University Press

Garman, E.F. and Schneider, T.R. (1997) *J. Appl. Crystallogr.* **30** 211-217

Giacovazzo, C., Monaco, H.L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G. and Catti, M. (1992) in *Fundamentals of Crystallography*, Ed Giacovazzo, C., Oxford Univeristy Press

Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981) *Acta Crystallogr.* **A37** 398-406

Huber, R. (1985) in *Experiences with the application of Patterson search techniques* 58-61, Molecular Replacement, (Machin, P.A. comp), Proceedings of the Daresbury Study Weekend, S.E.R.C, Daresbury Laboratory, Daresbury, U.K

Kabsch, W. (1988a) *J. Appl. Crystallogr.* **21** 67-71

Kabsch, W. (1988b) *J. Appl. Crystallogr.* **21** 916-924

Kleywegt, G.J. and Jones, T.A. (1994) in *From first map to final model* 59-66, Eds. Sawyer, L., Isaacs, N. and Bailey, S., S.E.R.C. Daresbury Laboratory, Daresbury, U.K

Lattman, E.E. (1972) *Acta Crystallogr.* **B28** 1065-1068

Leslie, A.G.W. (1992) Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography No 26, Daresbury Laboratory, Warrington, UK

Leslie, A.G.W. (1994) Mosflm User guide

Lifchitz, A. (1983) *Acta Crystallogr.* **A39** 130-139

Louie, G. (1996) in *Encyclopaedia of Molecular Biology and Molecular Medicine* Volume 5 110-115, Ed Meyers, R.A.,VCH, Weinheim

McPherson, A. (1990) *Eur. J. Biochem.* **189** 1-23

McPherson, A. (1982) in *The Preparation and Analysis of Protein Crystals*, John Wiley and Sons, New York

Mursudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) *Acta Crystallogr.* **D53** 240-255

Navaza, J. (1987) *Acta Crystallogr.* **A43** 645-653

Navaza, J. (1994) *Acta Crystallogr.* **A50** 157-163

Navaza, J. and Vernoslava, E. (1995) *Acta Crystallogr.* **A51** 445-449

Otwinowski, Z. (1993) in *Data Collection and Processing* 56-62, Eds. Sawyer, L., Isaacs, N. and Bailey, S., S.E.R.C. Daresbury Laboratory, Daresbury, U.K

Rhodes, G. (1993) in *Crystallography Made Crystal Clear*, Academic Press

Rossmann, M.G. and Blow, D.M. (1962) *Acta Crystallogr.* **15** 24-31

Sayre, D. (1974) *Acta Crystallogr.* **A30** 180-184

Sheldrick, G.M. and Schneider, T.R. (1997) *Methods Enzymol.* **277** 319-343

Steigemann, W. (1974) Ph.D. Thesis, Technische Universitaet Muenchen

Tickle, I.J. (1994) BCA Summer School, Bath

Tong, L. and Rossmann, M.G. (1990) *Acta Crystallogr.* **A46** 783-792

Tong, L. and Rossmann, M.G. (1997) *Methods Enzymol.* **276** 594-611

Tollin, P., Main, P. and Rossmann, M.G. (1966) *Acta Crystallogr.* **20** 404-407

Vagin, A. (1989) *CCP4 Newslett. Protein Crystallogr.* **29** 117-121

Vagin, A. and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30** 1022-1025

Wang, B.C. (1985) *Methods Enzymol.* **115** 90-112

Whittaker, E.J.W (1981) in *Crystallography : An introduction for Earth Science students*, Pergamon Press, Oxford

Zhang, K.Y.J. and Main, P. (1990) *Acta Crystallogr.* **A46** 377-381

Chapter 2

Structural Studies of Human Ubiquitous R167Q Porphobilinogen Deaminase

2.1 Introduction to Porphobilinogen Deaminase

2.1.1 Tetrapyrrole Biosynthetic Pathway

Most living organisms make use of tetrapyrroles present as prosthetic groups in proteins which are responsible for a diverse range of functions including transporting and storing oxygen, light gathering in photosynthesis, electron transport in respiration and detoxification of foreign compounds. By altering substituents on the periphery of the macrocycle and inserting a different metal ion into the centre of the macrocycle ring (eg iron in haem, magnesium in chlorophyll, cobalt in vitamin B_{12} and nickel in factor F_{430}), these functions can be achieved using a basic interchangeable structural template.

The pathway for the biosynthesis of naturally occurring tetrapyrroles from 5-aminolaevulinic acid (ALA) is the same in all organisms, up to and including uroporphyrinogen III. Subsequent steps involve methylation at positions 2 and 7 of the tetrapyrrole to form precorrin-2, a common precursor of factor F_{430} and vitamin B_{12}. Alternatively, decarboxylation of uroporphyrinogen III followed by further modifications of the macrocycle substituents yields protoporphyrin IX (Figure 2.1). This key porphyrin lies at the root of two branches; the iron pathway (haem biosynthesis) and the magnesium pathway (chlorophyll and bacteriochlorophyll biosynthesis) (Sano and Granick, 1961).

2.1.2 Haem Biosynthetic Pathway

Proteins such as haemoglobin, cytochromes, catalases and peroxidases utilise a tetrapyrrole pigment called haem (Sano and Granick, 1961). The pyrrole rings of these tetrapyrrole pigments are synthesised from ALA, a non-protein amino acid that is formed via two distinct pathways. In plants, algae and strict anaerobes, glutamate is converted to ALA via a multi-enzyme pathway (C5 pathway) (Kannagara *et al.*, 1994). In contrast, ALA synthesis in animals, fungi and facultative anaerobes involves the Shemin pathway (C4 pathway), in which glycine and succinyl CoA are condensed in a single enzyme catalysed step to form ALA (Figure 2.2). This step is catalysed by 5-aminolaevulinic acid synthase (ALAS), the first enzyme in the haem biosynthetic pathway. ALAS functions as a homodimer and utilises pyridoxal 5'-phosphate as a cofactor (Ferreira and

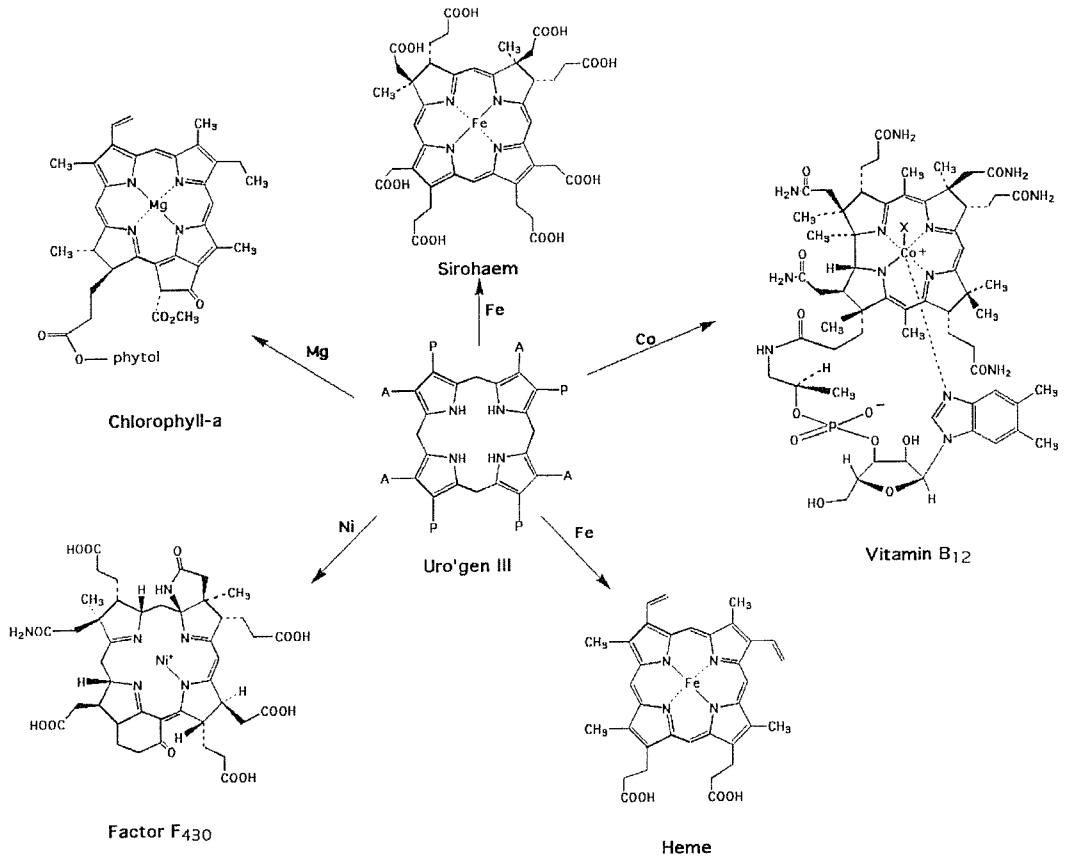


Figure 2.1: Tetrapyrrole biosynthesis from the ubiquitous uroporphyrinogen III (Warren and Scott, 1990) by altering substituents on the periphery of the macrocycle and inserting a different metal ion into the centre of the macrocycle ring (iron (Fe) in haem, magnesium (Mg) in cholorophyll, cobalt (Co) in vitamin B_{12} and nickel (Ni) in factor F_{430}).

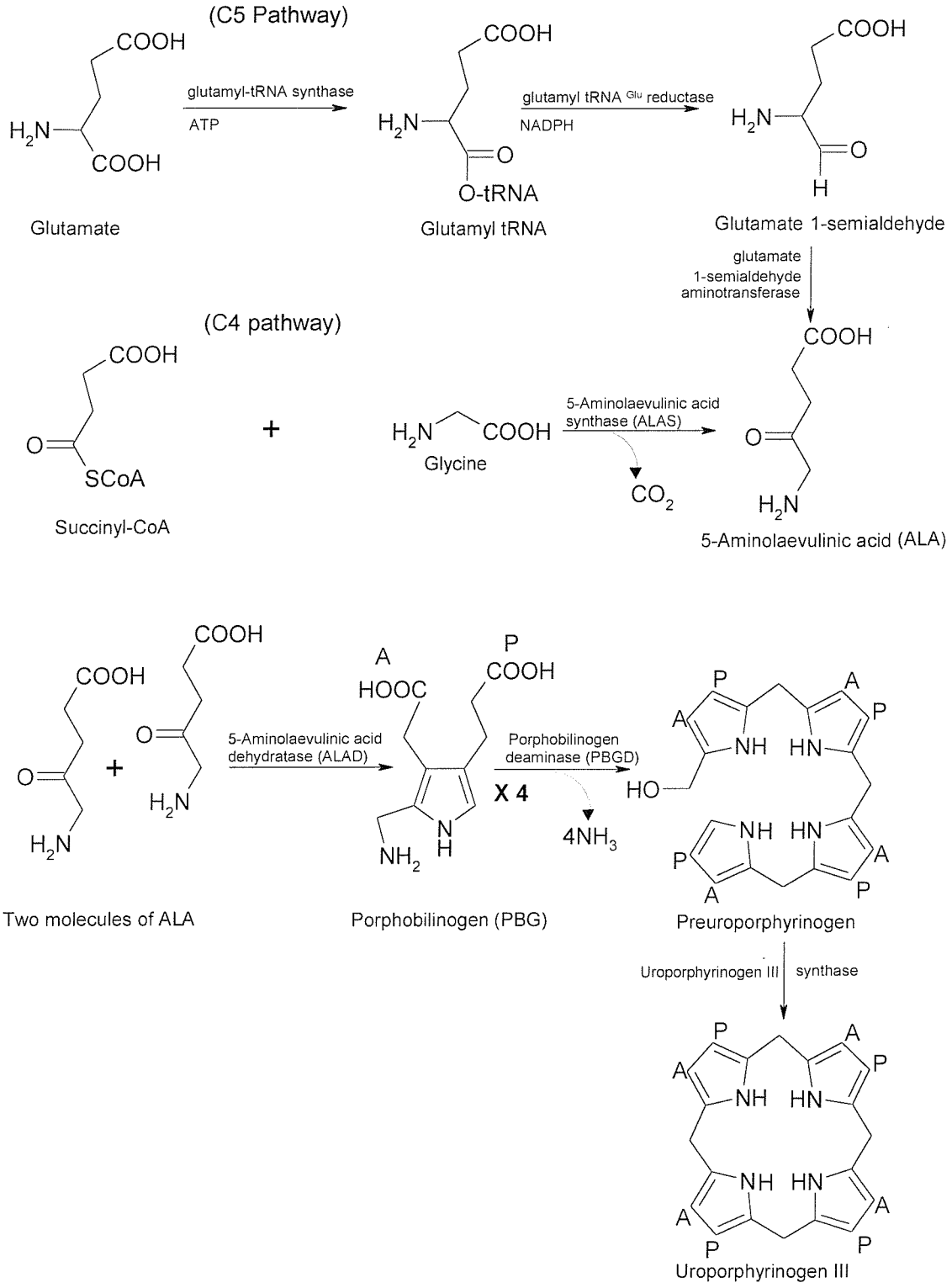


Figure 2.2: Schematic pathway for the biosynthesis of uroporphyrinogen III. 5-aminolaevulinic acid (ALA) is formed via two distinct pathways, the C5 pathway (in plants) and the C4 or Shemin pathway (in animals). In the presence of ALAD two ALA molecules are condensed to form porphobilinogen (PBG) (A and P refers to the acetic and propionic acid side chains). PBGD catalyses the condensation of 4 PBG molecules to form preuroporphyrinogen. Uroporphyrinogen III synthase cyclises the tetrapyrrole ring and inverts the final pyrrole to form the ubiquitous uroporphyrinogen III.

Gong, 1995). ALA dehydratase (ALAD), the second enzyme in the Shemin pathway, is a metalloprotein that catalyses the condensation of two ALA molecules (Figure 2.2) to form porphobilinogen (PBG) (Shemin and Russell, 1953). ALAD possesses two substrate binding sites termed the A and P site, as well as two metal binding sites. The latter provide a catalytic and structural role, and are classified as the α and β sites respectively. While the mammalian, yeast and *E. coli* ALAD require Zn^{2+} for catalytic activity, the plant enzymes are Mg^{2+} dependent. In the third step of the haem biosynthetic pathway, four molecules of PBG are condensed to form preuroporphyrinogen III (Figure 2.2), in the presence of porphobilinogen deaminase (PBGD, EC 4.3.1.8). PBGD utilises a novel dipyrromethane cofactor, which serves as a primer for the sequential addition of four PBG molecules. The initial cyclic intermediate of the tetrapyrrole biosynthetic pathway is produced by the conversion of preuroporphyrinogen III to uroporphyrinogen III (Figure 2.2). This reaction, catalysed by uroporphyrinogen III synthase, involves inversion of ring D and closure of the macrocyclic ring (Leeper, 1994).

Uroporphyrinogen III decarboxylase catalyses the fifth step in the haem biosynthetic pathway (Figure 2.3), which entails removing carboxylate groups from the acetate side chains at positions 2, 7, 12 and 18 of uroporphyrinogen III to form coproporphyrinogen III (Akhtar, 1994). Coproporphyrinogen III is subsequently converted to protoporphyrinogen IX by the action of coproporphyrinogen III oxidase (CPO), the sixth enzyme in the haem biosynthetic pathway (Figure 2.3). CPO catalyses the oxidative decarboxylation of the propionate side chains of rings A and B of coprogen at positions 3 and 8, to form vinyl groups (Kappas *et al.*, 1989). In the penultimate step of the Shemin pathway, protoporphyrinogen IX is converted to protoporphyrin IX by the action of protoporphyrin IX oxidase (PPO). PPO utilises tightly bound flavin adenine nucleotide and the obligatory electron acceptor is O_2. Finally, the terminal step of the haem biosynthetic pathway (Figure 2.3) is the chelation of ferrous iron into the centre of protoporphyrin IX to form haem, a reaction catalysed by ferrochelatase (Ferreira *et al.*, 1995).

2.1.3 Acute Intermittent Porphyria

Insufficient activity of enzymes in the haem biosynthetic pathway results in haem deficiency and the accumulation of tetrapyrrole intermediates. A group of several related

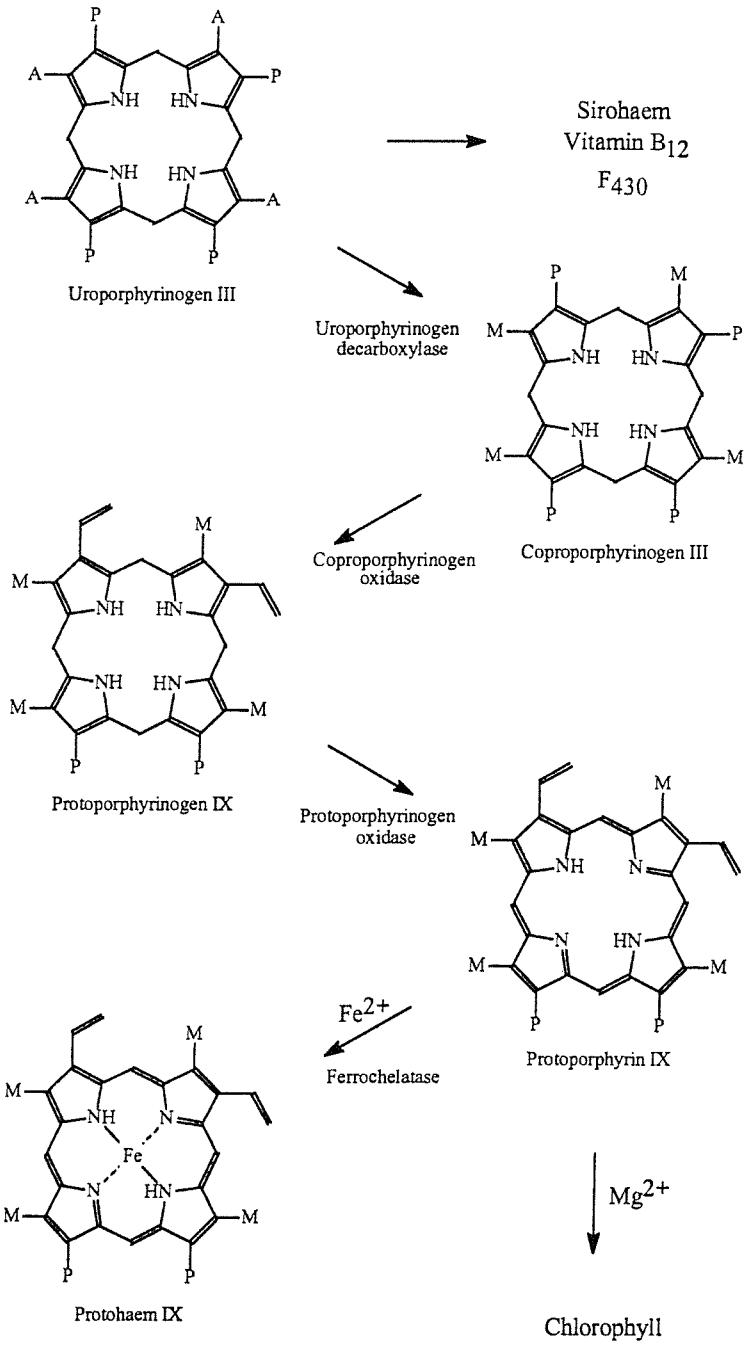


Figure 2.3: The haem biosynthetic pathway from ubiquitous uroporphyrinogen III. Uroporphyrinogen III decarboxylase catalyses the fifth step in the pathway, which entails removing carboxylate groups from the acetate side chains at positions 2, 7, 12 and 18 of uroporphyrinogen III to form coproporphyrinogen III. In the next step, coproporphyrinogen III oxidase (CPO) catalyses the oxidative decarboxylation of propionate side chains of rings A and B at positions 3 and 8 of coproporphyrinogen III to form protoporphyrinogen IX. In the penultimate step, protoporphyrin IX oxidase (PPO) catalyses the conversion of protoporphyrinogen III to protoporphyrin IX. Finally, the terminal step of the pathway involves chelation of ferrous iron (Fe^{2+}) into the centre of protoporphyrin IX to form haem, a reaction catalysed by ferrochelatase.

diseases linked to this pathway are referred to as porphyrias (Table 2.1). People with porphyria may develop skin photosensitivity or a condition known as the 'acute attack'. Acute intermittent porphyria (AIP) is an autosomal dominant disorder caused by deficiencies in PBGD, the third enzyme in the haem biosynthetic pathway. Typically patients exhibit a 50 % reduction in PBGD activity although approximately 90 % of gene carriers lead a normal life without developing any symptoms (Kappas *et al.*, 1989).

The gene coding for PBGD is located on chromosome 11 (Wang *et al.*, 1981). The enzyme is coded for by two distinct mRNA species, expressed in a tissue specific manner. The housekeeping (ubiquitous) isozyme is expressed in all cells and contains 361 residues (44 kDa) (Raich *et al.*, 1986), while the erythroid form is present only in red blood cells and is 17 residues shorter with 344 residues (42 kDa) (Grandchamp *et al.*, 1987). To date, more than one hundred different mutations have been identified in patients suffering from AIP in the form of single base substitutions and frame-shift mutations arising from insertions, deletions and splicing defects.

Patients susceptible to AIP are at risk of developing an acute attack, precipitated by alcohol consumption, drugs, infection, or low calorie intakes due to dieting or fasting (Grandchamp, 1998). The characteristic symptoms accompanying an acute attack include abdominal pain, cramps, constipation, nausea, anxiety and vomiting. Neurological symptoms associated with AIP may be due to raised levels of ALA, which is structurally similar to the inhibitory neurotransmitter 4-aminobutyric acid (Bagust *et al.*, 1985). However, more recently it is considered that haem deficiency may be the main cause of the neurological symptoms. The accumulation of haem precursors, PBG and ALA, causes a 'port wine' discolouration of the urine, which is the classic signature of an acute attack (Moore, 1993).

Successful treatment of an acute attack depends on early diagnosis, removal of precipitating factors and provision of intensive support therapy consisting in the administration of a high carbohydrate intake and haem arginate transfusion. The latter involves intravenous administration of haem arginate, which raises intracellular levels of haem in the liver, and suppresses the activity of ALAS, the initial and rate determining enzyme of the haem biosynthetic pathway, resulting in decreased levels of porphyrin

precursors in the plasma and urine. In mild attacks, poor carbohydrate intakes aggravate the condition in patients and therefore an adequate oral intake of glucose polymer drinks is essential. Various drugs are also used to overcome the symptomatic effects of an acute attack.

Haem Biosynthetic Enzyme	Associated disorder
5-Aminolaevulinic acid synthase	X-linked sideroblastic anaemia
5-Aminolaevulinic acid dehydratase	Doss porphyria and acute lead poisoning
Porphobilinogen deaminase	Acute intermittent porphria
Uroporphyrinogen synthase	Congenital erythropoietic porphyria
Uroporphyrinogen decarboxylase	Porphyria cutanea tarda
Coproporphyrinogen oxidase	Hereditary coproporphyria
Protoporphyrinogen oxidase	Variegate porphyria
Ferrochelatase	Erythropoietic protoporphyria

Table 2.1: Disorders linked with defects in enzymes of the haem biosynthetic pathway

2.1.4 Functional Studies of PBGD

PBGD is a cytoplasmic enzyme that assembles four PBG molecules to form the highly unstable linear tetrapyrrole preuroporphyrinogen. The enzyme was first described by Bogorad in 1958 and has since been isolated from a variety of sources including human erythrocytes (Raich *et al*, 1986), mouse (Beaumont *et al.*, 1989), yeast (Keng *et al.*, 1992) and has been expressed in *E.coli* (Thomas and Jordan, 1986). The high degree of sequence homology between PBGD enzymes from different species (45 %) suggests they all have a similar mechanism of action. The enzymes also share similar optimal pH ranging from 8.0-8.5, turnover numbers of about 0.5 s^{-1} and M_r's of 34,000-44,000.

The assembly process is initiated with ring A binding first followed by the sequential addition of rings B, C and D (Jordan and Seehra, 1979; Battersby *et al.*, 1979). The intermediates, each consisting of a different number of PBG units, are termed ES, ES_2, ES_3 and ES_4 (Figure 2.4). The existence of these intermediary forms were confirmed by treating PBGD with tritiated PBG and analysing the enzyme species that formed (Anderson and Desnick, 1980). Following this discovery, a search for the active site residue responsible

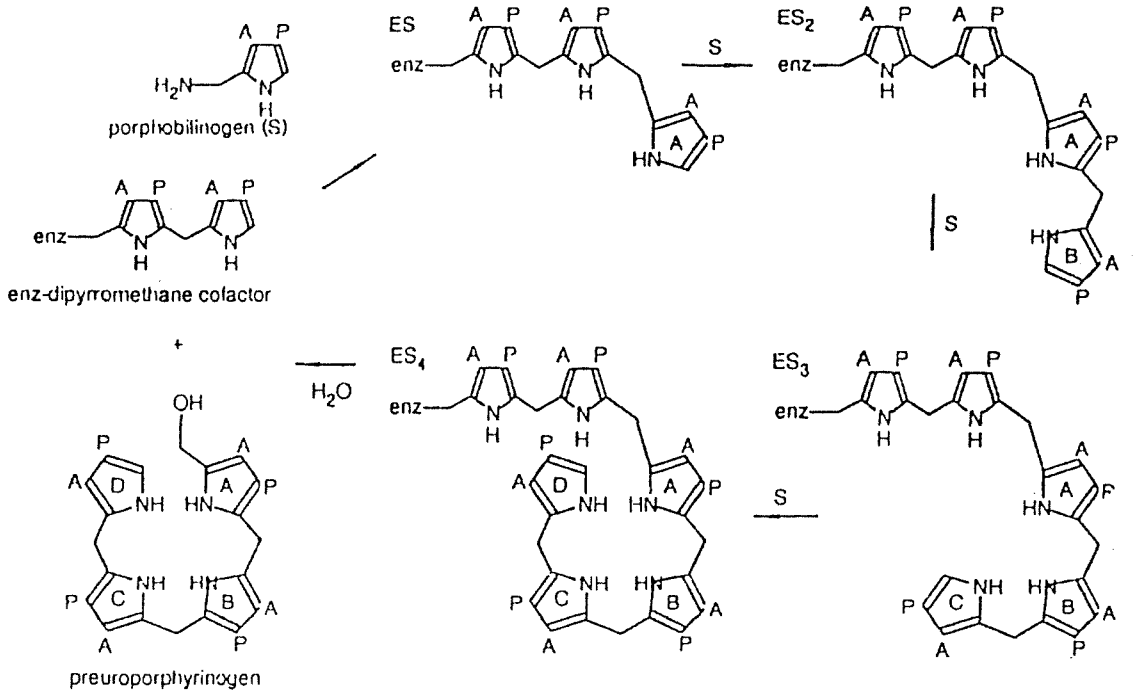


Figure 2.4: The sequential assembly of four porphobilinogen (PBG) molecules by porphobilinogen deaminase to form a highly unstable product, preuroporphyrinogen III. The dipyrromethane cofactor (DPM) serves as the primer during the assembly process; ring A binds to DPM first followed by the sequential additions of rings B, C and D. The intermediate steps, each with a different number of PBG units, are termed ES, ES_2, ES_3 and ES_4 and are isolable. The final hexapyrrole intermediate (ES_4) is subsequently cleaved with water to yield the highly unstable preuroporphyrinogen III, leaving the cofactor intact for another round of catalysis (Jordan, 1995).

for the covalent linkage to the substrate revealed that a lysine residue might be involved (Battersby *et al.*, 1983; Hart *et al.*, 1984). However, subsequent experiments with ^{14}C-PBG revealed a pink unlabelled dipyrrole chromophore suggesting that the pyrrole units were already present in PBGD before any substrate had been added (Jordan and Berry, 1983 unpublished). This pink chromophore was termed the dipyrromethane (DPM) cofactor (Jordan and Warren, 1987), and its structure was confirmed following ^{14}C labelling (Warren and Jordan, 1988) and ^{13}C NMR (Hart *et al.*, 1987) studies.

Further studies revealed that the cofactor is comprised of two PBG molecules covalently linked together. Since the cofactor is not turned over during the assembly process, the final intermediate consists of six pyrrole units, four arising from PBG and two contributed by the cofactor. Because each pyrrole ring contains negatively charged acetate and propionate side chains, PBGD must have a special protein organisation in stabilising the increasing negative charge as each PBG unit is incorporated into the growing chain. Sequence alignment comparisons of PBGD enzymes from various sources (Figure 2.5) have identified highly conserved arginine residues that may interact with the carboxylate groups of the pyrrole rings during the assembly process. Site directed mutagenesis of these invariant arginine residues disrupt various stages in the PBGD catalytic cycle (Table 2.2) (Lander *et al.*, 1991; Woodcock and Jordan, 1991).

Mutation	Specific activity (μmoles/h/mg	K_m (μm)	Presence of cofactor	Stage affected with 10 μM PBG
Wild-type	43	17	+	None
R11H	0.1	not determinable	+	E→ES
R131H	not determinable	not determinable	-	Cofactor assembly
R132H	not determinable	not determinable	-	Cofactor assembly
R149H	11.1	200	+	ES→ES_2
R155H	0.5	not determinable	+	E→ES
R176H	6.0	30	+	ES→ES_2

Table 2.2: The effect of mutating highly conserved arginine residues on the assembly process catalysed by PBGD (Woodcock and Jordan, 1991).

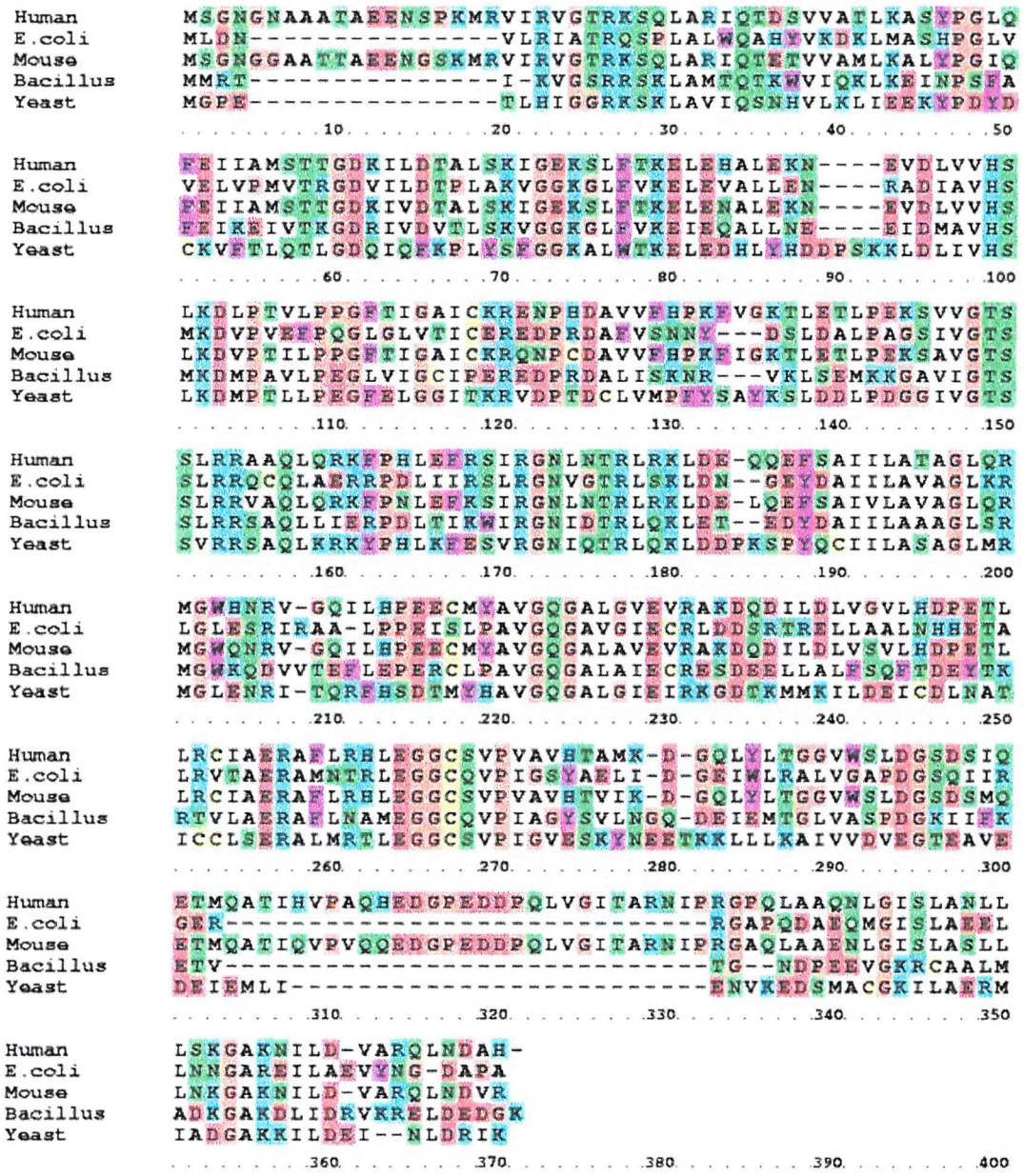


Figure 2.5: Primary sequence alignments of porphobilinogen deaminase (PBGD) from various sources (CINEMA 2.1; Attwood *et al.,* 1997). The alignment was performed using MALIGN (Johnson, 1990). The sequence identity between human PBGD and *E.coli* PBGD is 47 %. Note that there are several highly conserved arginine residues that may form interactions with the carboxylate groups of the pyrrole units during the assembly process. The colour key for residue types is the following: polar positive (blue), polar negative (red), polar neutral (green), non-polar aliphatic (white), non-polar aromatic (purple), glycine and proline (brown) and cysteine (yellow). The sequences were obtained from the SWISS-PROT/TrEMBL protein sequence database (Bairoch and Apweiler, 2000).

2.1.5 X-Ray Structure of *E .coli* PBGD

The structure of oxidised PBGD from *E. coli* has been determined at 1.76 Å resolution (Louie *et al.*, 1992; 1996) using MIR, while the reduced form of the enzyme has been solved at 1.9 Å resolution using MAD (Hädener *et al.*, 1992; 1999). *E. coli* PBGD is made up of three equally sized domains and has overall dimensions of 57 Å x 43 Å x 32 Å. Domains 1 and 2 possess a similar topology consisting of a doubly wound, parallel β sheet of five strands, whereas domain 3 is an open faced, three stranded anti-parallel β sheet (Figure 2.6). The structure is also stabilised by hydrophobic interactions within the hydrophobic core present in each domain.

The large catalytic cleft (approximately 15 Å x 13 Å x 12 Å) is located at the interface between domains 1 and 2 and is large enough to accommodate the cofactor and the growing polypyrrole chain. Nearby the cleft there is a disordered loop region (residues 48-58) which may function as a flexible lid as well as forming interactions with the substrate. Further roles for this mobile loop region may include transiently shielding the growing polypyrrole chain from solvent exposure or controlling the environment of the active site (Louie *et al.*, 1996). The cofactor is linked covalently to Cys-242, located on a loop in domain 3, and projects into the catalytic cleft between domains 1 and 2. Structural studies of *E. coli* PBGD have identified 3 pyrrole ring binding sites (Lambert *et al.*, 1994), two of which are occupied by the C1 and C2 rings of the DPM cofactor. The third site adjacent to the C2 ring of DPM, has been identified as the putative substrate binding site (S).

The DPM cofactor can adopt two conformations that differ at the bridging carbon depending on its oxidation state (Figure 2.7). In the reduced form, the dihedral angles of the cofactor are less obtuse than those of the oxidised form, creating a distinct elbow at the methylene group. As a result, there is little change in the conformation of ring C1 between the two states. However, ring C2 in the oxidised state occupies a position at the front of the cleft close to the S site, interacting with Arg-11, Ser-13, Phe-62, Arg-149 and Arg-155 (Louie *et al.*, 1992; Lambert *et al.*, 1994). In addition, X-ray structural studies of *E. coli* PBGD reveal the importance of Phe-62, since it is able to stack against the PBG pyrrole ring at the S site (Louie *et al.*, 1992).

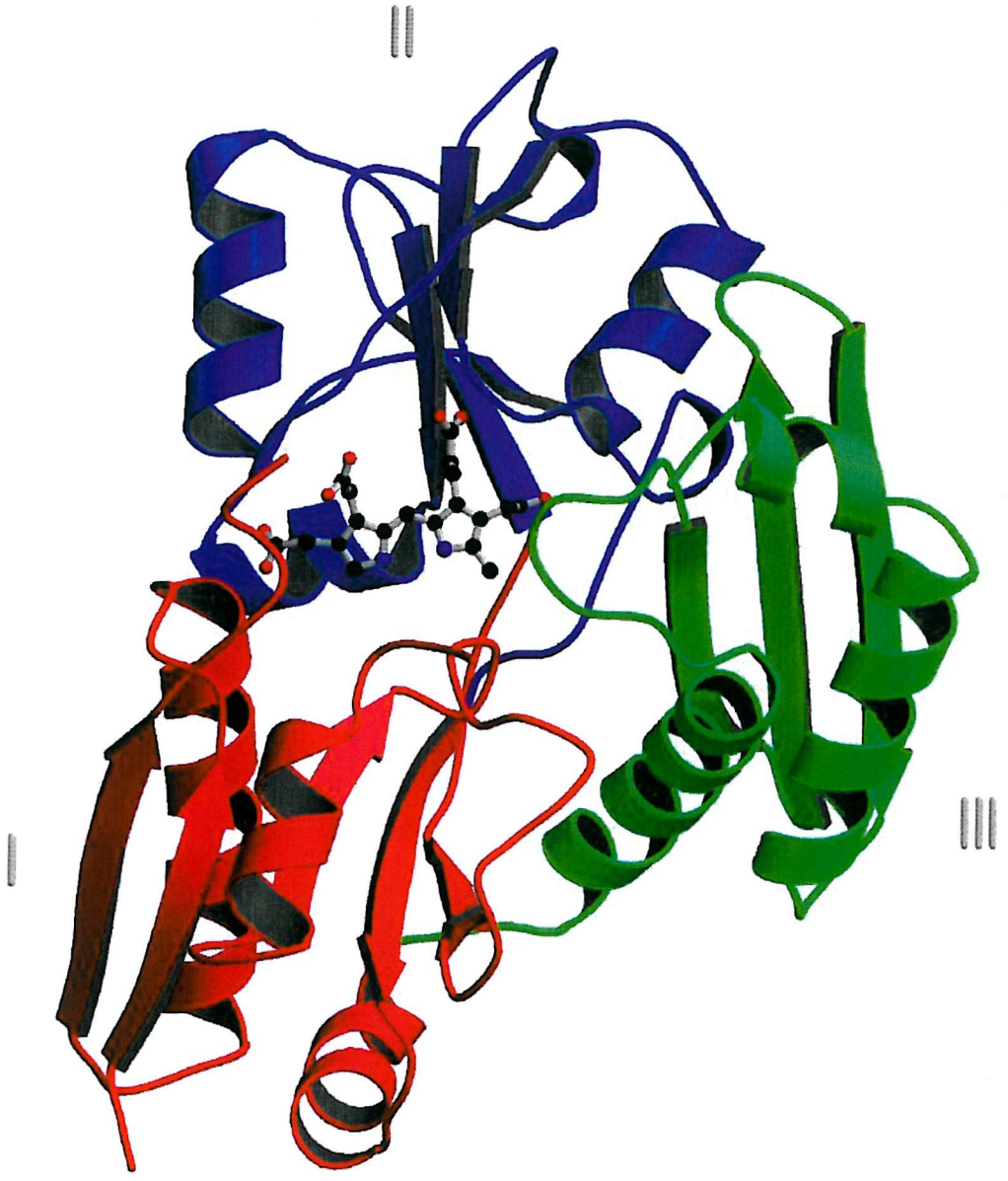


Figure 2.6: The three-dimensional structure of *E. coli* porphobilinogen deaminase (PBGD) (drawn with MOLSCRIPT (Kraulis, 1991) and rendered using Raster3D (Merritt and Bacon, 1997)). PBGD is made up of three equally sized domains (domain 1 in red, domain 2 in blue and domain 3 in green). Domains 1 and 2 have a similar overall topology, each comprising of a doubly wound, mainly parallel β sheet of five strands with α helical segments packing against each face of the sheet. Domain 3 is an open faced, three-stranded anti-parallel β sheet with three helices covering one of the faces (Louie *et al.,* 1992). The dipyrromethane cofactor is covalently linked to a loop residue in domain 3 and displayed in CPK representation.

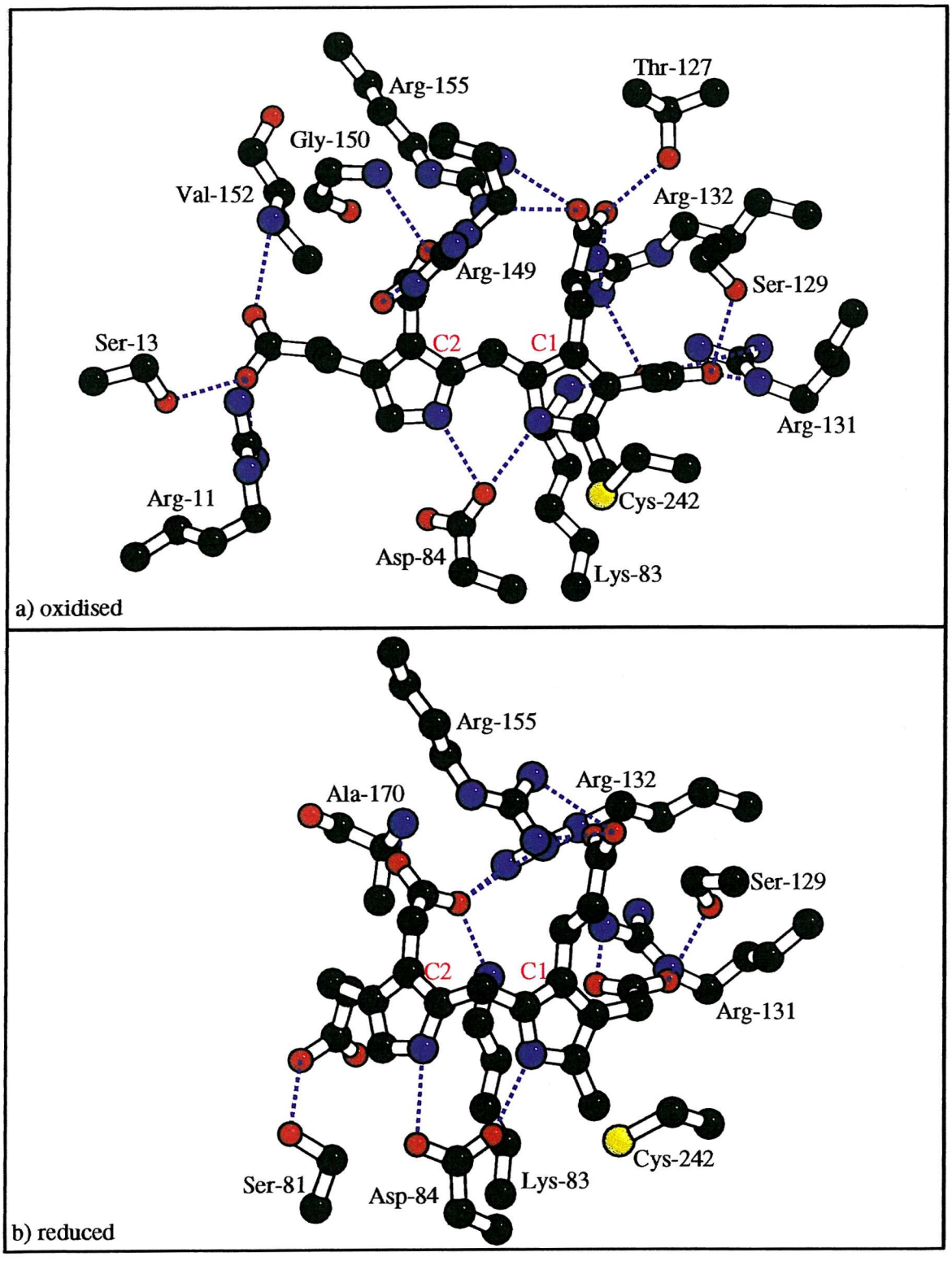


Figure 2.7: The two conformations that the dipyrromethane cofactor can adopt within the active site cleft of *E. coli* porphobilinogen deaminase depending on its oxidation state (BOBSCRIPT; Ensouf, 1997). The interactions of ring C1 remain the same in both states because there is little change in conformation, however ring C2 in the oxidised state (a) occupies a position at the front of the cleft proximal to the substrate binding site, interacting with the side-chains of several invariant residues (Arg-11, Ser-13, Asp-84, Arg-149 and Arg-155 (Louie *et al.*, 1992; Lambert *et al.*, 1994). In the reduced state (b) ring C2 is positioned towards the rear of the cleft and interacts with the side-chains of Ser-81, Lys-83, Asp-84, Arg-132 and Arg-176 (not shown).

In the reduced state, ring C2 is positioned at the back of the cleft and interacts with Ser-81, Lys-83, Arg-132 and Arg-176. The interactions between PBGD and the carboxylate groups of ring C1 are mainly salt bridges with highly conserved Arg-131, Arg-132, Arg-155 and Lys-83 residues. Furthermore, the invariant Asp-84 interacts with both pyrrole nitrogens of DPM, in both the oxidised and reduced forms. The importance of Asp-84 has been confirmed by site directed mutagenesis, since total loss of activity was observed with the D84A mutant, while only 0.5 % of wild type activity was retained with the D84E mutant (Woodcock and Jordan, 1994).

2.1.6 Mechanism of PBGD

The reaction catalysed by PBGD is initiated following the deamination of the PBG unit at the S site, resulting in the formation of an azafulvene (Pichon *et al.*, 1992). The azafulvene is an electrophilic species that reacts with the nucleophilic free α-position of the cofactor (or free α-position in the ES, ES_2 or ES_3 complexes) to form a new C-C bond (Figure 2.8). Several roles have been proposed for the highly conserved Asp-84 in the reaction including stabilisation of the positive charge on the pyrrole units after deamination, protonation of the ammonia or abstraction of the α-proton to complete the reaction (Jordan, 1994). The propionic acid side chain of substrate could also abstract the α-proton.

Following abstraction of the α-proton, it is conceivable that the newly incorporated pyrrole vacates the S site and positions itself at the C2 site. The remaining three PBG units are added sequentially in a similar stepwise fashion (rings B, C and D) resulting in the formation of ES_4. The final intermediate composed of a hexapyrrole unit is subsequently cleaved by water to yield the highly unstable preuroporphyrinogen III (Jordan, 1995). The hydrolytic reaction regenerates the cofactor ready for another round of catalysis. Chain termination may result from steric effects since the size of the active site cleft is only large enough to accommodate up to 4 pyrrole units (Louie *et al.*, 1992).

Because X-ray structural studies have provided no evidence for a multi-catalytic site in PBGD, a single catalytic site has been proposed, implying that large conformational changes must occur during elongation of the polypyrrole chain. It is possible that the hinge

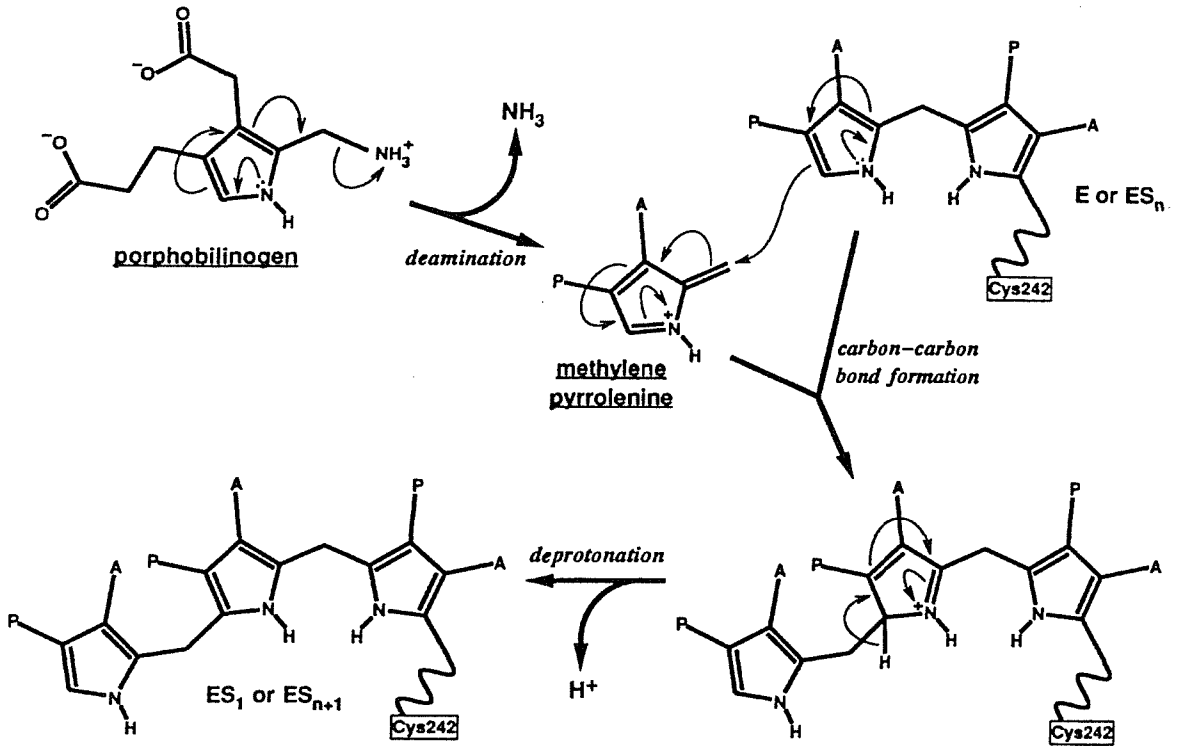


Figure 2.8: The mechanism for the porphobilinogen (PBG) ring coupling reaction catalysed by porphobilinogen demainase. Following deamination of a PBG unit at the substrate binding site, a methylene pyrrolenine (azafulvene) is generated. The azafulvene is an electrophilic species that reacts with the nucleophilic free α-position of the cofactor (or free α-position in ES, ES_2 or ES_3), forming a new C-C bond. The reaction is completed by abstraction of the α-proton, presumably by either Asp-84 or the propionic acid moiety of PBG. The remaining three PBG units are added sequentially in a similar stepwise fashion resulting in the formation of a hexapyrrole intermediate (Louie *et al.*, 1996).

regions connecting the domains allow considerable flexibility, enabling the extending polypyrrole chain to be manipulated through the substrate binding and cofactor sites. Further support for the existence of a single catalytic site is provided by the fact that the overall topology of domains 1 and 2 of PBGD are similar to that of the transferrins and periplasmic binding proteins which are known to undergo a hinge bending motion between domains 1 and 2 upon ligand binding (Louie, 1993).

Evidence for a conformational change occurring upon substrate binding has also been provided by chemical modification studies with *N*-ethylmaleimide (a thiophilic reagent), since PBGD is increasingly susceptible to alkylation, as the catalytic reaction proceeds ($ES \rightarrow ES_2 \rightarrow ES_3 \rightarrow ES_4$). The modified residue is Cys-134, which appears to be increasingly exposed during assembly of the tetrapyrrole chain due to the separation of domains 2 and 3 (Warren and Jordan, 1988; Warren *et al.*, 1995). As a corollary, the accompanying movement of domain 3 away from domains 1 and 2 would allow repositioning of the developing polypyrrole chain as the incoming PBG unit binds at the S site.

2.1.7 Research Aims

There are several conserved arginine residues which have been implicated in the reaction catalysed by *E. coli* PBGD. Among these Arg-149 when mutated disrupts the tetrapolymerisation process accumulating stable ES complexes (interfering mainly with ES→ES2). X-ray structural studies of *E. coli* PBGD suggests that Arg-149 most likely binds to the incoming substrate pyrrole molecule. The equivalent residue in human ubiquitous porphobilinogen deaminase (uPBGD) has been mutated (R167Q), displaying 90 % reduction in activity compared to the wild type. This mutant is of considerable clinical interest since it is responsible for the hereditary disease AIP. Therefore, detailed structural information on the human enzyme and its site directed mutants would provide a greater insight into the molecular basis of AIP. It is anticipated that the structure of R167Q-uPBGD will provide a stable ES complex, thus revealing how the substrate docks at the substrate binding site.

2.2 Expression, Purification and Crystallisation of Human R167Q-uPBGD

Human R167Q-uPBGD was expressed in *E. coli* and purified as described by Abeer Al-Dbass (2001). Crystallisation was performed by vapour diffusion at 4 °C in the dark in 2-4 µl hanging drops. Single crystals were grown upon equilibration against reservoirs containing 0.6 M ammonium sulphate, 1.2 M lithium sulphate, 5 % ethylene glycol, 0.05 M sodium citrate (pH 5.6) and 0.05 M dithiothreitol (DTT) (Al-Dbass, 2001).

2.3 Data Collection and Processing of R167Q-uPBGD

Synchrotron diffraction data from a single R167Q-uPBGD crystal were collected with a Mar CCD detector at the European Synchrotron Radiation Facility (ESRF; in Grenoble, France) on station BM14 (λ= 1.00). Low temperature data (100 K) was collected using an Oxford Cryosystem cooler. Prior to flash-freezing, the crystal was transferred to a reservoir solution containing 30 % (v/v) glycerol. Successful flash-freezing was obtained when the crystal was transferred from 0 % to 30 % glycerol in steps of 7.5 %. The crystal initially diffracted X-rays to a resolution of 2.65 Å (Figure 2.9) but towards the end of data collection diffraction to 3.5 Å was observed, probably due to radiation damage.

A total of 180 0.5°-oscillation frames were measured, with an exposure time of 120 s per frame and a crystal to detector distance of 145.5 mm. The raw data were processed with MOSFLM (Leslie, 1992) and scaled and merged using programs of the CCP4 (Collaborative Computing Project Number 4, 1994) suite. The merged intensities were subsequently converted to structure factor amplitudes using TRUNCATE (CCP4, 1994). The relevant data processing statistics for the R167Q-uPBGD crystal are listed in Table 2.3.

The unit cell dimensions for R167Q-uPBGD (a=81.1Å, b=105.2Å, c=109.6Å, $\alpha=\beta=\gamma=90°$) were determined using the REFIX algorithm within MOSFLM. This indicated that the crystal belonged to the orthorhombic space group. As a result, data were processed in space group P222 and pseudo precession pictures were generated using HKLVIEW. Systematic absences were observed along b (0K0 zone), consistent with the presence of a 2_1 screw axis (Figure 2.10), while reflections along H00 and 00L zones were

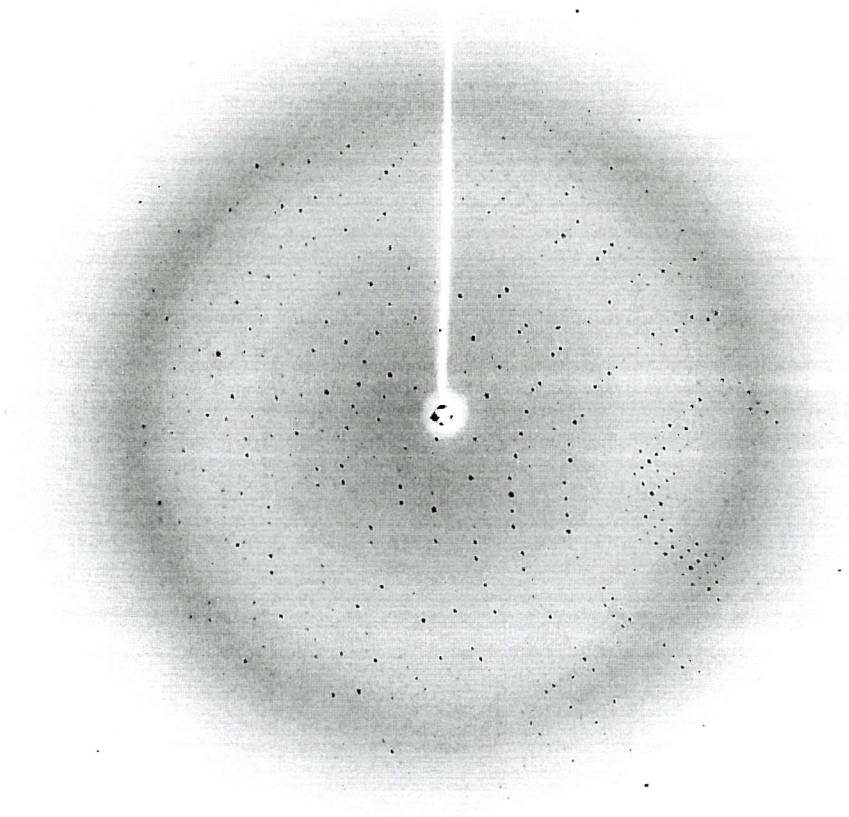


Figure 2.9: A 0.5° oscillation image from a human R167Q-uPBGD crystal collected on a MAR CCD detector at the ESRF (Grenoble, France). The crystal was flash cooled at 100 K and diffracted X-rays to a resolution of 2.65 Å. The exposure time was 120 s with a crystal-to-detector distance of 145.5 mm.

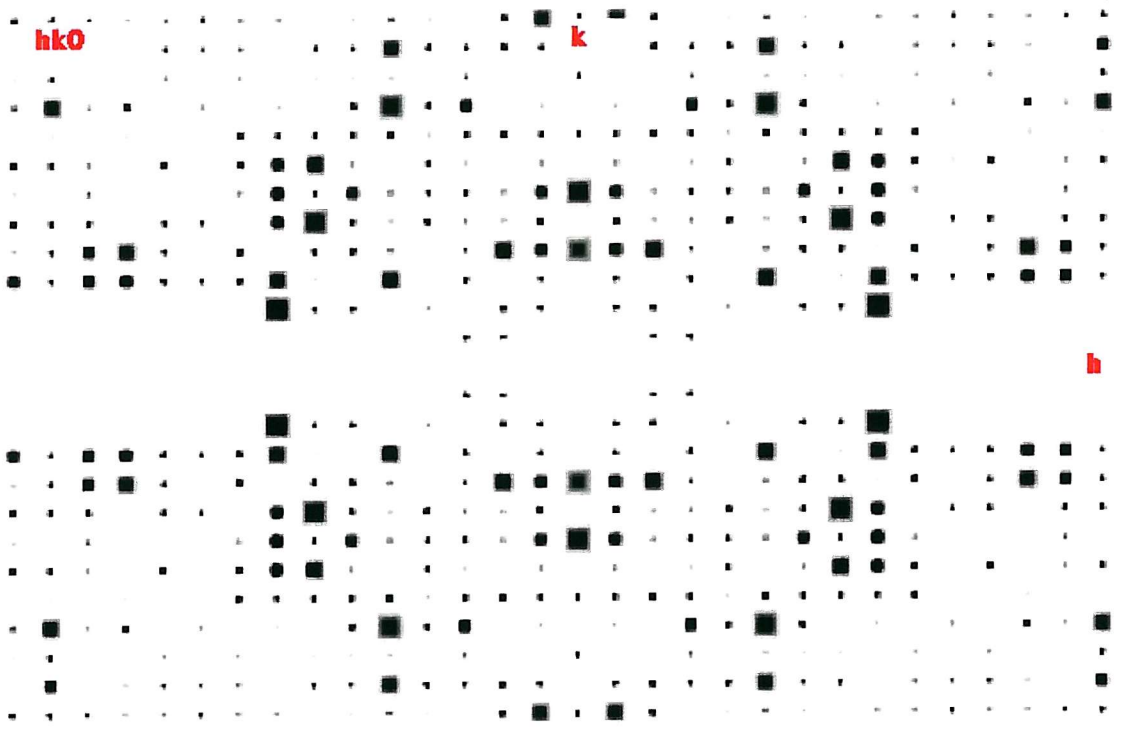


Figure 2.10: A pseudo precision picture displaying a 2_1 screw axis along zone 0K0 (k=2n) for human R167Q-uPBGD data when processed in P222. Note that data from zone H00 is missing.

missing. These observations indicated that the protein had crystallised into either space group $P222_1$, $P2_12_12_1$ or $P2_12_12$. In this case, space group ambiguity could only be overcome following a correct translation function solution during molecular replacement.

Resolution (Å)	R_{merge} (%)	N-ref_{unique}	I/σ(I)	Completeness (%)	Multiplicity
8.38	6.5	853	8.0	89.6	3.6
5.93	7.2	1530	7.5	92.9	3.7
4.84	8.4	1973	6.9	94.2	3.7
4.19	9.1	2327	5.9	94.5	3.7
3.75	10.8	2649	5.3	95.3	3.7
3.42	15.1	2931	4.2	95.9	3.7
3.17	16.3	2988	4.2	90.4	2.3
2.96	14.9	2645	4.5	75.0	1.7
2.79	21.0	2844	3.3	75.9	1.6
2.65	30.0	3069	2.3	77.6	1.6
Overall	9.6	23809	5.1	86.4	2.8

Table 2.3: Relevant data processing statistics for a crystal of human R167Q-uPBGD.

2.4 Molecular Replacement Studies of Human R167Q-uPBGD

2.4.1 Introduction

A solvent content of 53 % (V_m=2.63 $Å^3$/Da) was calculated (Matthews, 1968) assuming there are two molecules of PBGD (88,000 Da) in the asymmetric unit of the human R167Q-uPBGD crystal. The *E. coli* PBGD also crystallised in the orthorhombic space group ($P2_12_12$) with unit cell dimensions of a=88.0Å, b=75.9Å, c=50.5Å but only had one molecule in the asymmetric unit (Louie *et al.*, 1992). The calculated unit cell volume of *E. coli* PBGD is approximately three times smaller than that of the human R167Q-uPBGD, and therefore it was considered quite likely that there was a dimer present in the asymmetric unit of R167Q- uPBGD.

2.4.2 Cross Rotation Search

Initial phases for human R167Q-uPBGD were obtained by molecular replacement calculations using MOLREP. A monomer from the refined structure of *E. coli* PBGD at 2.0 Å resolution was used as the search model with the DPM cofactor omitted (accession code 1PDA; Louie *et al.*, 1992). The cross rotation search was performed using reflections within the resolution range 20 Å to 3.5 Å and yielded two significant solutions (α=144.38°, β=52.21°, γ=162.31° and α=140.09°, β=56.21°, γ=166.95°) when searching with a radius of integration of 35 Å (Table 2.4). The two solutions corresponded to the orientation of each PBGD molecule within the target asymmetric unit, as implied by the solvent content calculations.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	144.38	52.21	162.31	7.48
2	140.09	56.21	166.95	6.67
3	38.90	58.83	227.46	4.76
4	17.36	46.62	315.25	4.63
5	148.17	90.00	142.94	3.88

Table 2.4: The 5 highest peaks following cross rotation calculations performed using data between 20 Å and 3.5 Å (α, β, and γ correspond to CCP4 Eulerian angles).

2.4.3 Translation Search

The position of each PBGD molecule within the target asymmetric unit was determined following a translation search calculated in space group $P2_12_12$. The highest peak from the cross rotation function (α=144.38°, β=52.21°, and γ=162.31°) yielded a translation function peak of 13 σ (Table 2.5). The position of the first molecule corresponded to fractional co-ordinate shifts of X=0.150, Y=0.828 and Z=0.884. The translation search was subsequently repeated with the position of the first molecule fixed, since both molecules within the asymmetric unit must be positioned with respect to the same origin. Therefore by fixing the position of the first molecule, the position of the second molecule was determined with respect to the same origin (Table 2.6). The position of the second molecule with respect to the first corresponded to fractional co-ordinate shifts of X=0.258, Y=0.324 and Z=0.379 (peak number 1 from Table 2.6). The crystal packing of the

solution was visualised using MOLPACK (Wang *et al.*, 1991) and shown to possess sensible crystal contacts (Figure 2.11). For completeness, it was decided to calculate the translation function in space groups $P222_1$ and $P2_12_12_1$. However no significant translation function peaks were observed for either space group (Tables 2.7 and 2.8)..

Peak Number	X	Y	Z	Dens/σ
1	0.150	0.828	0.884	13.22
2	0.251	0.844	0.836	6.89
3	0.357	0.850	0.882	6.88
4	0.255	0.856	0.927	6.84
5	0.314	0.867	0.914	6.03

Table 2.5: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the first orientation of the cross rotation search (α=144.38°, β=52.21°, γ=162.31°).

Peak Number	X	Y	Z	Dens/σ
1	0.258	0.324	0.379	6.04
2	0.869	0.032	0.453	5.12
3	0.509	0.291	0.091	4.95
4	0.083	0.325	0.380	4.86
5	0.299	0.290	0.090	4.68

Table 2.6: The 5 highest non-crystallographic translation function peaks after fixing the position of the first molecule.

Peak Number	X	Y	Z	Dens/σ
1	0.111	0.346	0.290	5.23
2	0.139	0.346	0.301	5.10
3	0.353	0.047	0.087	5.04
4	0.112	0.090	0.286	4.94
5	0.930	0.039	0.161	4.69

Table 2.7: The 5 highest translation function peaks calculated in space group $P222_1$, using the first orientation of the cross rotation search.

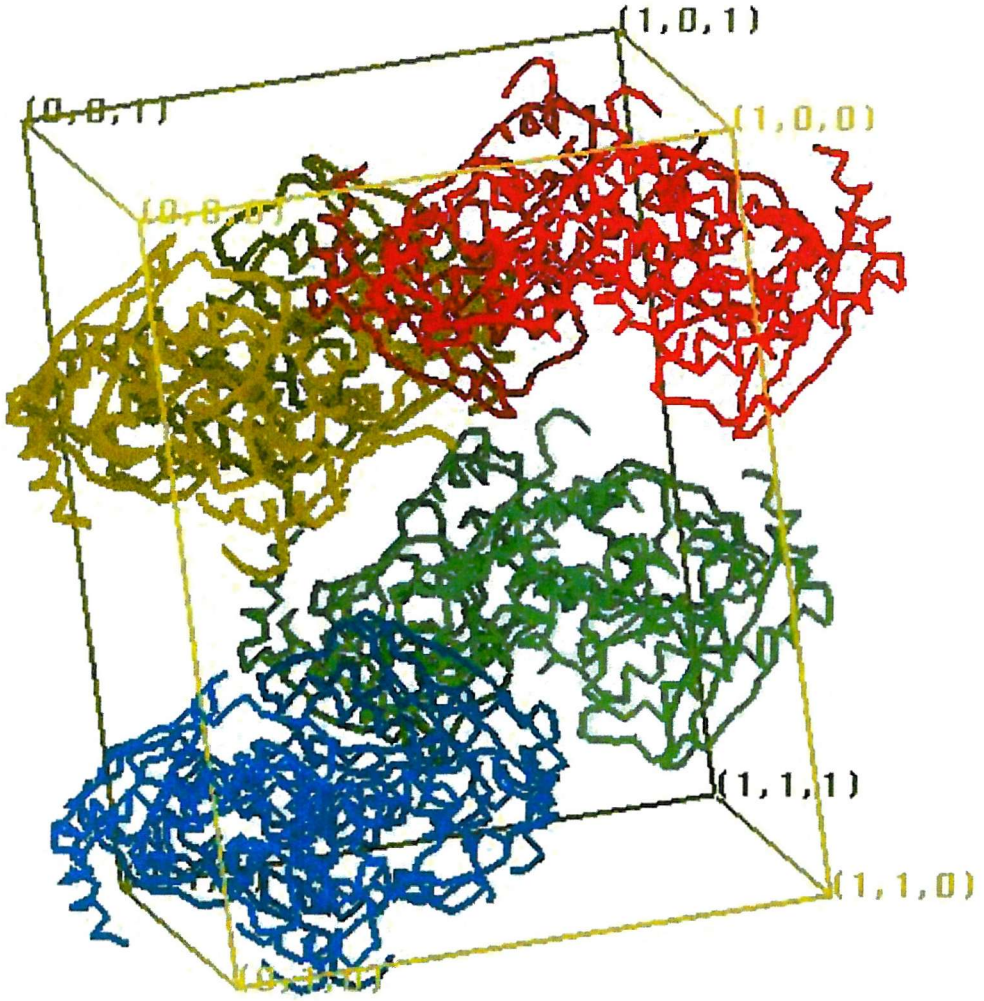


Figure 2.11: The crystal packing of human R167Q-uPBGD (MOLPACK; Wang *et al.*, 1991). The R167Q-uPBGD crystal crystallises in the orthorhombic space group $P2_12_12$ (a=81.1 Å, b=105.2 Å, c=109.6 Å, $\alpha=\beta=\gamma=90^{\circ}$) with a dimer (each depicted in a different colour) in the asymmetric unit and a solvent content of 53 %.

Peak Number	X	Y	Z	Dens/σ
1	0.911	0.223	0.427	4.68
2	0.911	0.227	0.110	4.60
3	0.912	0.223	0.070	4.58
4	0.036	0.019	0.373	4.42
5	0.034	0.024	0.354	4.29

Table 2.8: The 5 highest translation function peaks calculated in space group $P2_12_12_1$, using the first orientation of the cross rotation search.

2.5 Refinement of the Human R167Q-uPBGD Structure

The model was refined with the program CNS using 22202 reflections in the resolution range 10-2.65 Å. A test set of reflections (5 % of data) was used for R_{free} calculations. During the early stages of refinement, NCS restraints were maintained between the two monomers. The model was subjected to an initial round of refinement consisting of rigid body refinement, simulated annealing (slow cool from 5000 to 300 K in 25 K decrements) and grouped B-factor refinement. The resulting model was used to calculate sigmaA weighted electron density maps that were subjected to two-fold NCS averaging. The maps were calculated with CNS and converted into X-PLOR format using MAPMAN (Kleywegt and Jones, 1994).

All map inspections and model manipulations were performed using QUANTA (Molecular Simulations, Inc Burlington, Massachusetts). Initial examination of these maps revealed well defined *Fo-Fc* density for the DPM cofactor and side chains of human residues (Figure 2.12). This demonstrated the validity of the molecular replacement solution and also indicated that the calculated maps were not dominated by model bias. The model was improved by several rounds of manual rebuilding, during which human PBGD residues were fitted to the *Fo-Fc* density. The model was then subjected to several alternating cycles of refinement and model building. The subsequent refinement procedures consisted of rigid body refinement, simulated annealing and grouped B-factor refinement.

During the latter stages of model building the large insert in domain 3 was built in as the electron density for this region improved (Figure 2.13). Currently, there are several regions

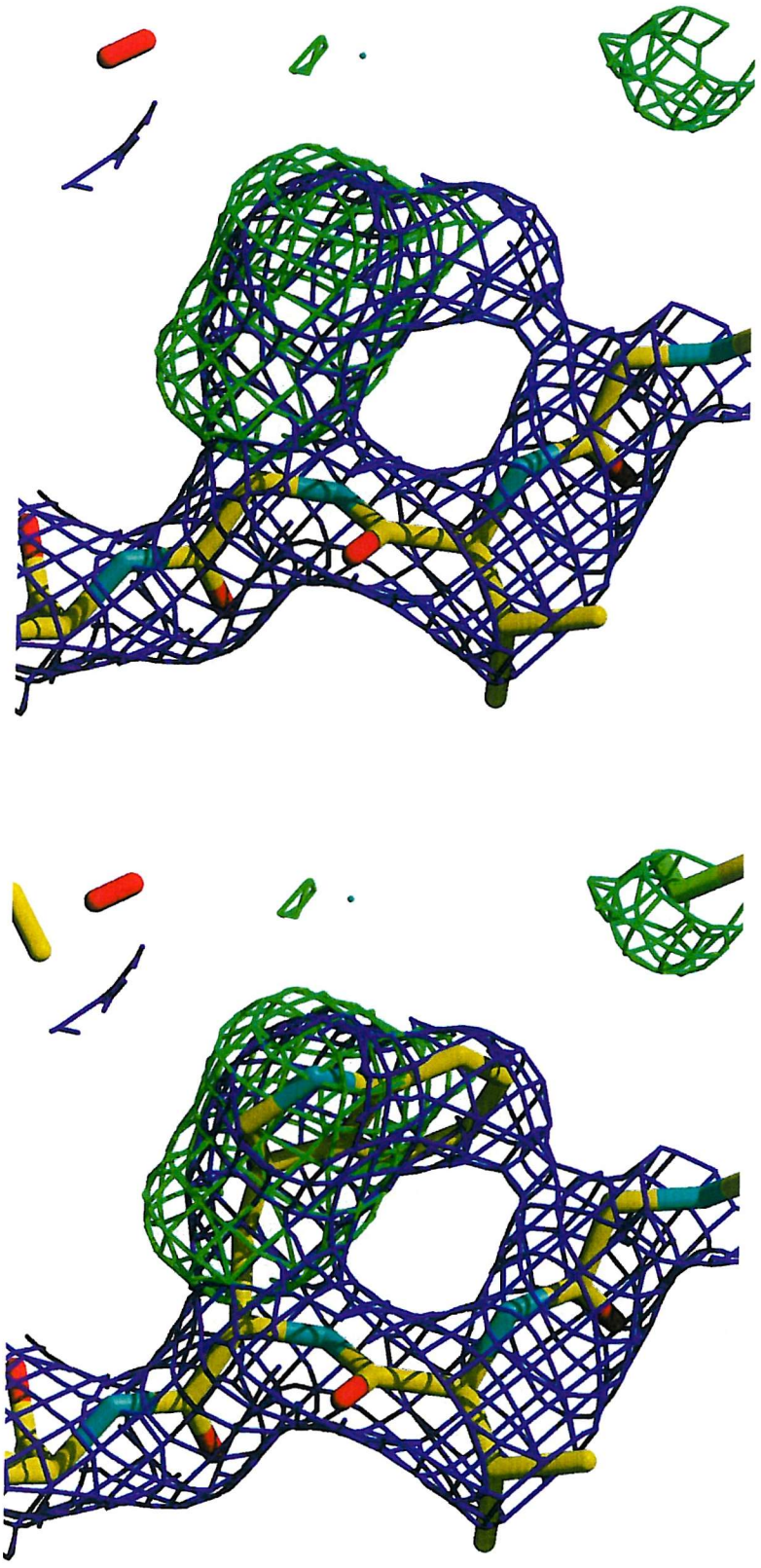


Figure 2.12: Confirmation of a correct molecular replacement solution for human R167Q-uPBGD. Features that are present in human R167Q-uPBGD crystal but absent from the *E. coli* PBGD model are displayed (XtalView (McCree, 1999) and rendered with Raster3D). *Fo-Fc* electron density (depicted in green and contoured at the 2 σ level) for Trp-283 (equivalent to Gly-264 in the *E. coli* model) can be observed (top) and is modelled in the human R167Q-uPBGD structure (bottom). The *2Fo-Fc* map is contoured at the 1 σ level (depicted in purple).

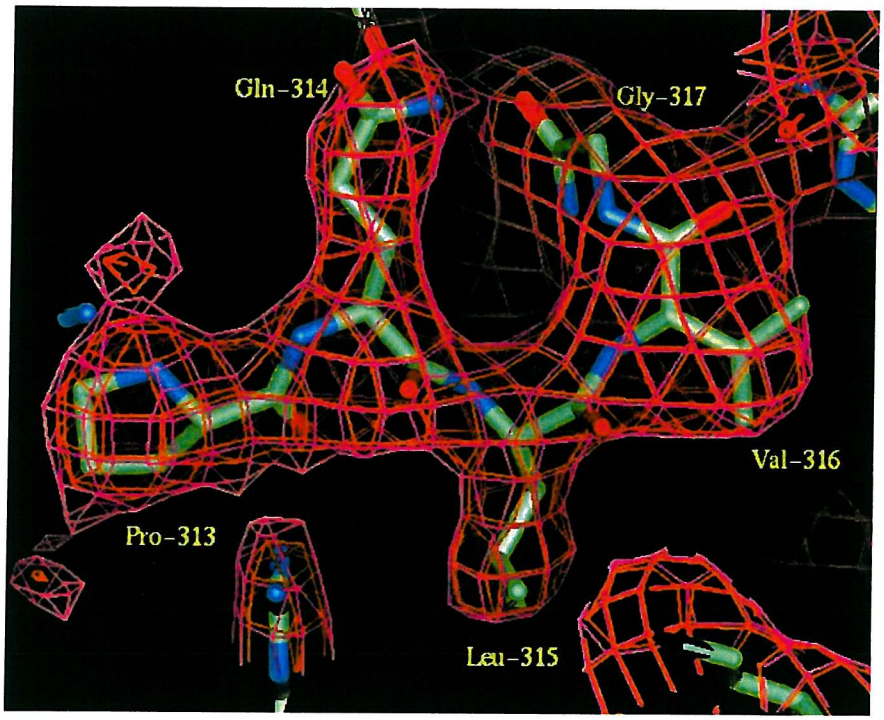


Figure 2.13: A region of the large insert (Pro-313, Gln-314, Leu-315, Val-316 and Gly-317) in domain 3 of the R167Q-uPBGD structure (QUANTA). The insert was modelled and subsequently refined to give well defined *2Fo-Fc* electron density (contoured at the 1.0 σ (pink) and 1.3 σ (orange) levels).

that exhibit high temperature factors (helices $\alpha 2_1$ and $\alpha 2_2$), and are therefore being rebuilt using omit maps. The refinement statistics after round 7 of model building and refinement are listed in Table 2.9. At present, no density can be observed for the *N*-terminus (residues 1-18), a mobile loop region in domain 1 (57-78), a part of the large insert (residues 307-312) and for the *C*-terminus (residues 359-361), thus prompting the release of the NCS restraints to observe whether differences exist between the two molecules.

Resolution range (Å)	30-2.65
R-factor (%)	27.9
R-free (%)	32.0
Number of reflections in working set	20990
Number of reflections in test set	1071
Rms bond length deviation (Å)	0.014
Rms bond angle deviation (°)	2.04

Table 2.9: Current refinement statistics for the human R167Q-uPBGD structure.

2.6 The Structure of Human R167Q-uPBGD

2.6.1 Overall Structure

The three-dimensional structure of human R167Q-uPBGD consists of three domains (Figure 2.14) with combined overall dimensions of 57 x 47 x 55 Å. The secondary structure elements of human R167Q-uPBGD from the *N* to the *C* terminus are listed in Table 2.10. The topology of domains 1 (residues 18-114 and 219-236) and 2 (residues 120-212) are similar, namely, a doubly wound, five stranded mainly parallel β-sheet (strands $\beta5_1$ and $\beta5_2$ are anti-parallel). The α-helical segments pack against each face of the β-sheet. By contrast, domain 3 (residues 241-357) is an open faced three-stranded anti-parallel β-sheet, with three α helical segments that cover one face and a large loop (residues 298-324) that connects $\beta3_3$ and $\alpha2_3$. Domains 1 and 2 are linked together by two hinge segments (residues 115-119 and 213-218), whereas the polypeptide chain connecting domains 1 and 3 involves residues 237-240.

Residue Range	Secondary Structure	Residue Range	Secondary Structure	Residue Range	Secondary Structure
1-18	disordered	141-144	strand ($\beta1_2$)	237-241	loop
19-20	loop	145-147	loop	242-255	helix ($\alpha1_3$)
21-25	strand ($\beta1_1$)	148-157	helix ($\alpha1_2$)	256-264	loop
26-28	loop	158-160	loop	265-272	strand ($\beta1_3$)
29-45	helix ($\alpha1_1$)	161-164	strand ($\beta2_2$)	273-274	loop
46-50	loop	165-170	loop	275-283	strand ($\beta2_3$)
51-56	strand ($\beta2_1$)	171-178	helix ($\alpha2_2$)	284-289	loop
57-78	disordered	179-184	loop	289-297	strand ($\beta3_3$)
79-86	helix ($\alpha2_1$)	185-189	strand ($\beta3_2$)	298-306	loop
87-91	loop	190-195	helix ($\alpha3_2$)	307-312	disordered
92-96	strand ($\beta3_1$)	196-203	loop	313-324	loop
97-107	loop	204-206	strand ($\beta4_2$)	325-343	helix ($\alpha2_3$)
108-114	strand ($\beta4_1$)	207-218	loop	344-346	loop
115-119	loop	219-225	strand ($\beta5_1$)	347-356	helix ($\alpha3_3$)
120-125	strand ($\beta5_2$)	226-228	loop	357-358	loop
126-140	loop	229-236	helix ($\alpha3_1$)		

Table 2.10: The secondary structure elements of human R167Q-uPBGD listed in order from the *N* to the *C* terminus (subscripts refer to domains).

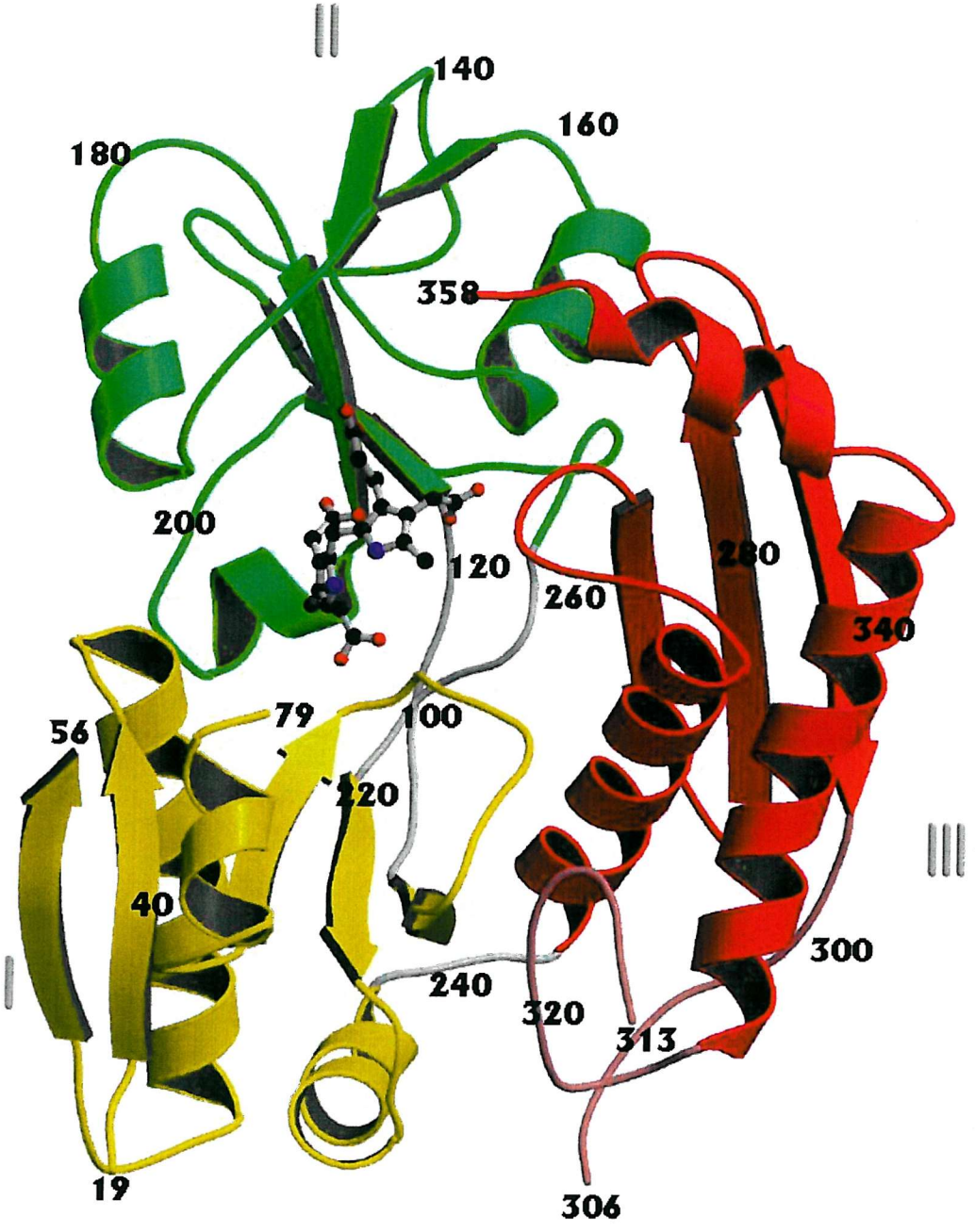


Figure 2.14: The three-dimensional structure of human R167Q-uPBGD (MOLSCRIPT and rendered with Raster3D). The R167Q-uPBGD structure is made up of three domains (domain 1 in yellow, domain 2 in green and domain 3 in red). The polypeptide chain segments connecting domains 1 and 2 and domains 1 and 3 are displayed in white. The overall topologies of domains 1 (residues 18-114 and 219-236) and 2 (residues 120-212) are similar, each comprising of a doubly wound, mainly parallel β sheet of five strands with α helical segments packing against each face of the sheet. Domain 3 (residues 241-357) is an open faced, three-stranded anti-parallel β sheet with three α helical segments that cover one of the faces and a large insertion loop (depicted in pink) that interacts with domain 1. The dipyrromethane cofactor is also displayed in CPK representation.

2.6.2 Structural Comparisons with *E. coli* PBGD

2.6.2.1 Introduction

As expected, the overall three-dimensional structure of human R167Q-uPBGD is similar to the *E. coli* PBGD structure (Figure 2.15) with the main structural differences occurring as insertions in loop regions; there is a three residue insertion occurring between strands $\beta5_2$ and $\beta1_2$ (129-131), a one residue insertion between helix $\alpha2_2$ and strand $\beta3_2$ (180) and a 29 residue insertion following strand $\beta3_3$ (298-327). The three-dimensional superposition of the R167Q-uPBGD structure on that of the oxidised and reduced form of *E. coli* PBGD yielded an rms deviations of 1.03 Å (for 283 C^{α} atoms) and 0.99 Å (for 278 C^{α} atoms), respectively.

The three-dimensional superposition of R167Q-uPBGD structure with *E. coli* PBGD in the oxidised form was repeated for the individual domains (Figure 2.16). Domain 2 of the R167Q-uPBGD superposed well with the corresponding domain of *E. coli* PBGD, yielding an rms deviation of 0.79 Å (for 89 C^{α} atoms). In contrast, domains 1 and 3 yielded rms deviations of 0.86 Å (for 93 C^{α} atoms) and 0.89 Å (for 86 C^{α} atoms), respectively. The large structural deviations (> 1.0 Å) are mainly restricted to solvent exposed loop regions and include residues 25-27 (following strand $\beta1_1$), 101-106 (following strand $\beta3_1$), 127-132 (following strand $\beta5_2$), 157-160 (following helix $\beta1_2$), 207-210 (following strand $\beta4_2$), 272-274 (following strand $\beta1_3$) and 284-289 (following strand $\beta2_3$).

2.6.2.2 Interactions Between Domains

There are few direct interactions between the domains suggesting a high degree of flexibility for the human R167Q-uPBGD structure, as in the *E. coli* structure. With the exception of polypeptide chain connections, the interactions between domains 1 and 2 include a hydrogen bond (Gln-34 OE1—Arg-195 NH1 2.67 Å) and several hydrophobic interactions involving the aliphatic moieties of Gln-29, Gln-34 and Met-196. The packing of domains 1 and 3 is primarily mediated by hydrogen-bonding interactions between $\beta3_1$, $\beta4_1$ and the large insertion loop (Pro-101 O—Arg-251 NH2 2.77 Å, Thr-102 O—Arg-251 NH2 3.02 Å, Ile-110 N—Gly-317 O 2.71 Å, Ile-110 O—Thr-314 N 2.73 Å, Thr-109

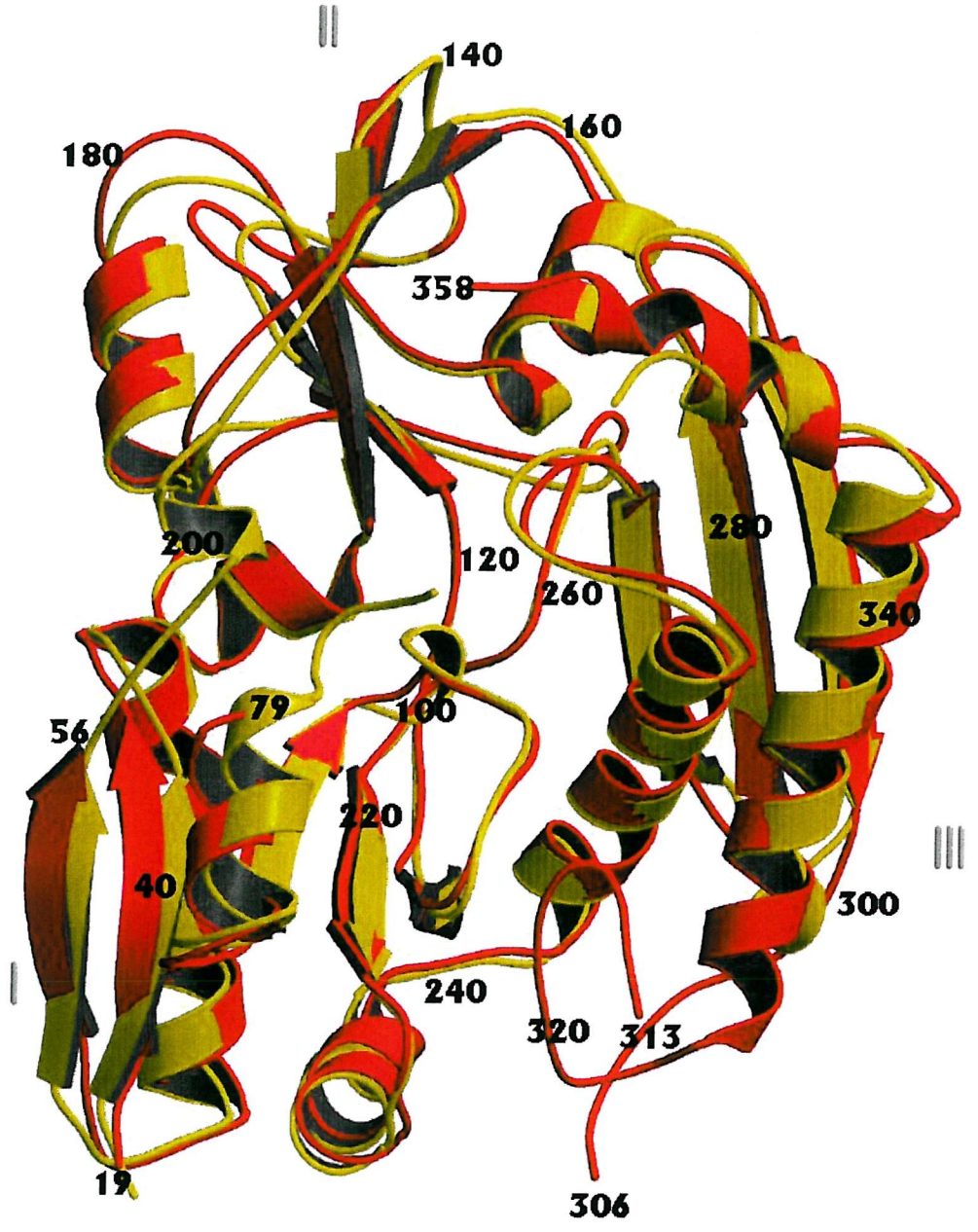


Figure 2.15: The superposition of the human R167Q-uPBGD (depicted in red) and *E. coli* PBGD (oxidised form depicted in yellow) structures (MOLSCRIPT and rendered with Raster3D). Note that the overall structure of human R167Q-uPBGD is very similar to that of *E. coli* PBGD, with the main structural differences (> 1.0 Å) occurring at several solvent exposed loop regions and at the large insertion (residues 297-324 in domain 3).

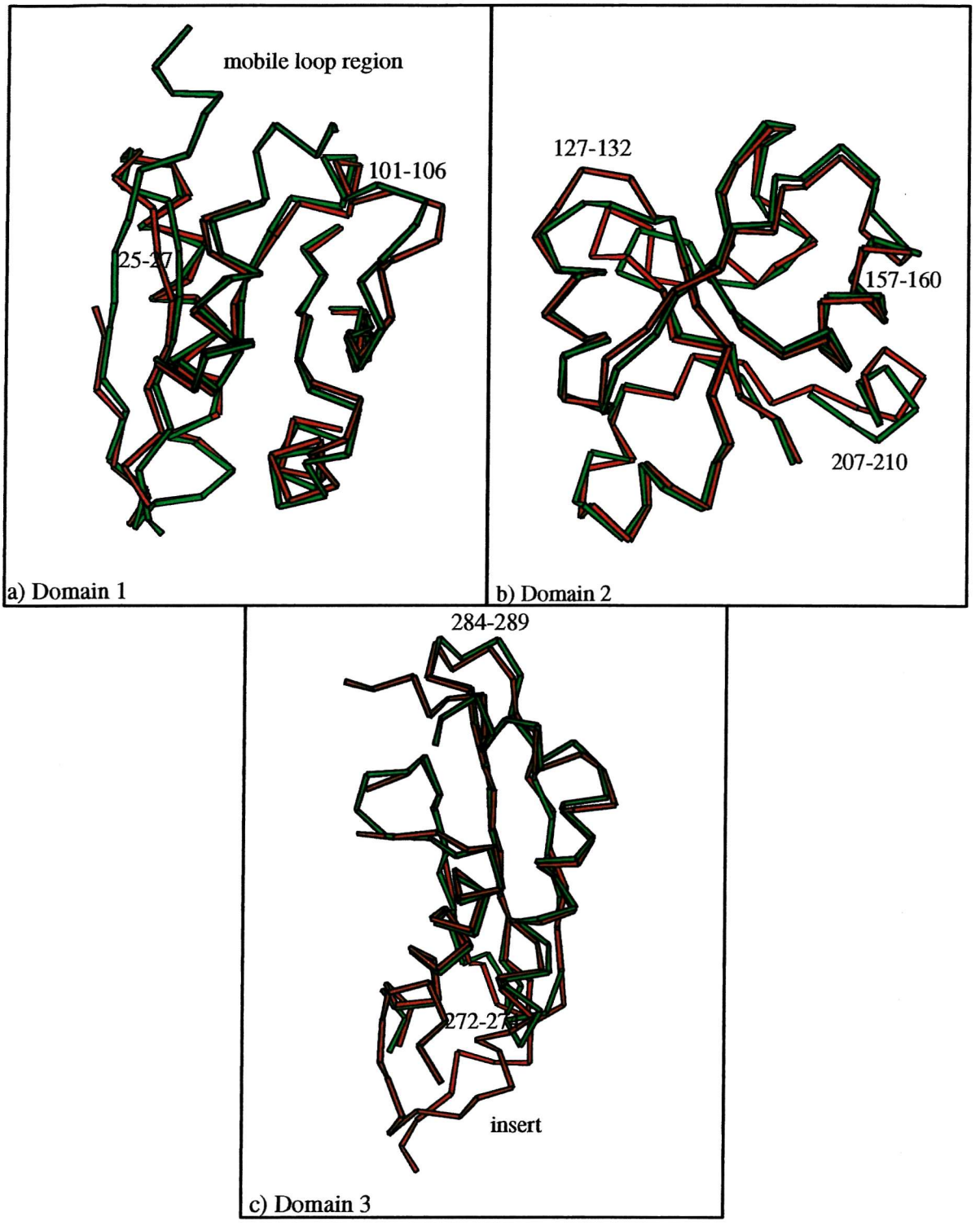


Figure 2.16: Superposition of the individual domains of human R167Q-uPBGD (red) and the oxidised form of *E. coli* PBGD (green) (MOLSCRIPT). For domain 1 the rms deviation is 0.86 Å for 93 equivalent C^{α} atoms that superpose within 3.5 Å. Note that the active site mobile loop region in domain 1 is more disordered in human R167Q-uPBGD than in *E. coli* PBGD. Domain 2 of both structures also superposes well yielding an rms deviation of 0.79 Å (for 89 C^{α} atoms). The large difference in domain 3 is mainly attributed to the insert following strand $\beta 3_3$ in human R167Q-uPBGD; the superposition of domain 3 yields an rms deviation of 0.89 Å (for 86 C^{α} atoms). The regions showing the largest structural differences (<1.0 Å) are mainly restricted to solvent exposed regions and have been labelled (human numbering).

OG1—Gln-314 NE2 2.78 Å and Thr-109 OG1—Gly-317 O 2.80 Å). In contrast, there are no hydrogen-bonding interactions between domains 2 and 3.

The sequence identity between the *E. coli* and human enzyme is 49 % and the majority of residue substitutions that occur perform compensatory functions. This is reflected at the hydrophobic interface between domain 1 and 2, where Met-82, Ala-225 and Val-228 in the *E. coli* structure are substituted by Ile-97, Leu-244 and Cys-247, respectively in human R167Q-uPBGD. The hydrophobic interface between domains 2 and 3 is also maintained through compensatory hydrophobic interactions (Cys-134, Leu-193 and Pro-266 of *E. coli* PBGD are substituted by Ala-152, Met-212 and Leu-285 in the human PBGD structure) and a bulky aromatic residue (Trp-283 in human PBGD equivalent to Gly-264 in *E. coli*), which makes numerous hydrophobic contacts.

2.6.2.3 Hydrophobic Core

As observed in all water soluble proteins, human R167Q-uPBGD and *E. coli* PBGD possess a hydrophobic core in each of the domains. In domains 1 and 2, the hydrophobic core is made up of apolar residues from the β-sheet and the flanking helices, which help to maintain the structural stability of the protein. This is reflected especially in domain 1 whereby a number of compensatory hydrophobic exchanges are observed (Leu-6, Ile-8, Ala-20, Val-36, Leu-38, Ile-77, Leu-93, Ile-203, Cys-205 and Leu-216 in the *E. coli* structure are replaced by Ile-21, Val-23, Thr-35, Phe-51, Ile-53, Leu-92, Phe-108, Val-222, Val-224 and Val-235 in the human structure).

The residues that make up the hydrophobic core in domain 2 of the human enzyme are larger and bulkier than those in the *E. coli* structure. For instance, Ile-145, Val-152, Tyr-164 and Leu-179 in the *E. coli* structure are replaced by Phe-163, Leu-170, Phe-183 and Trp-198 in the human structure. Furthermore, polar residues such as Arg-140 and Ser-110 that swing away from the core region in *E. coli* PBGD are substituted by Phe-158 and Phe-125 in human PBGD. As a result, the hydrophobic core in domain 2 of the human enzyme is more hydrophobic than that of the *E. coli* enzyme.

The majority of residues that contribute towards the hydrophobic core in domain 3 possess small aliphatic side chains that are conserved in both structures, occurring between two roughly parallel helices ($\alpha1_3$ and $\alpha2_3$). Moreover, there are two compensatory hydrophobic exchanges that occur in this region, whereby Thr-229 and Met-234 in the *E. coli* structure are replaced by Ile-248 and Phe-253 in the human structure. The hydrophobic core region is further stabilised by a series of hydrophobic interactions between the large insert (Ala-303, Ile-318, Ile-323, Pro-324 and Pro-327) and the base of helix $\alpha2_3$ (Leu-244 and Leu-245).

2.6.3 Crystal Contacts of Human R167Q-uPBGD

Regions that participate in crystal contacts include the loop following helix $\alpha1_1$, helix $\alpha3_1$, the large insert following strand $\beta3_3$, and the *C*-terminal of helix $\alpha2_3$. The contacts made by the loop following helix $\alpha1_1$ include several hydrophobic interactions (Tyr-46, Pro-47) and two hydrogen-bonding interactions. The crystal contacts contributed by helix $\alpha3_1$ are mainly hydrophobic mediated by Leu-234 and Val-237. The large insert in domain 3 also contributes to a number of hydrophobic interactions and several hydrogen bonding interactions involving His-305, Glu-306 and Asn-322. Finally, the *C*-terminal end of helix $\alpha2_3$ and the proceeding loop region contributes to the majority of the crystal contacts mediated by several hydrogen-bonding (Ser-337, Asn-340, Ser-344 and Lys-345) and hydrophobic (Ile-336, Leu-341, Leu-343 and Gly-346) interactions.

2.6.4 Dimer Interface of Human R167Q-uPBGD

The molecular surface area buried at the dimer interface of R167Q-PBGD is 1631 $Å^2$ per subunit. As each monomer is related by an approximate (NCS) 2-fold rotation, all interactions at the interface occur twice (Figure 2.17). The principal dimerisation contacts arise from one face of helix $\alpha1_1$ (residues 32, 36, 40, 43 and 44), the *N*-terminal of strand $\beta2_1$ (residues 50 and 53), and several loop regions (following helix $\alpha2_2$ (residues 178-181) and strand $\beta3_2$ (residues 195-202)). The interface is stabilised by a series of hydrophobic contacts (including Ile-33, Ile-53, Pro-127, Phe-129, Val-130 Met-196, Gly-197, Trp-198 and the aliphatic moieties of Asp-36, Arg-132 Lys-43, His-126, Arg-175, Asp-178, Glu-179 and Asn-200) and a series of salt bridges and hydrogen-bonds (listed in Table 2.11).

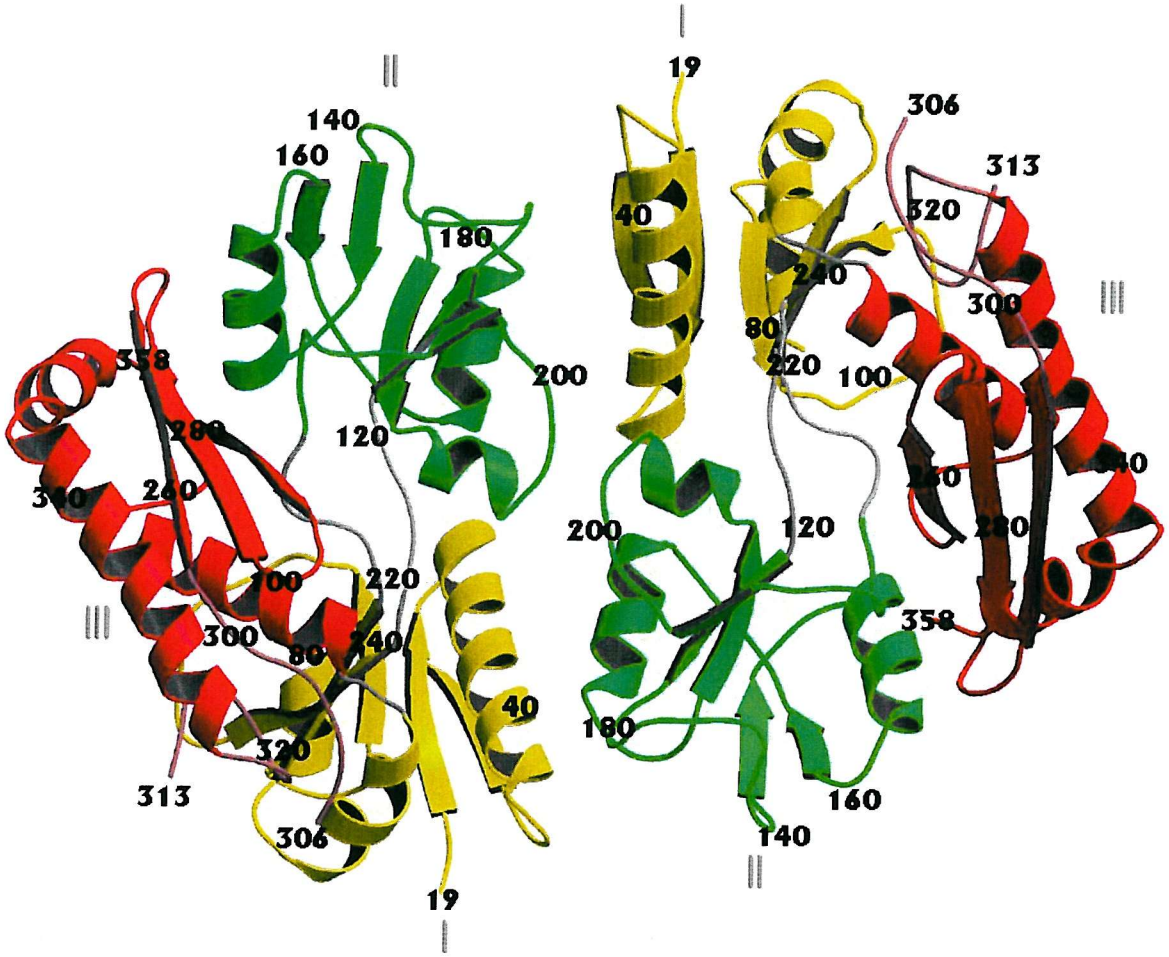


Figure 2.17: The human R167Q-uPBGD dimer (MOLSCRIPT and rendered with Raster3D). The principal dimerisation contacts arise from one face of helix $\alpha 1_1$ (residues 32, 36, 40, 43 and 44), the *N*-terminal of strand $\beta 2_1$ (residues 50 and 53), the loop following helix $\alpha 2_2$ (residues 178-181) and the loop following strand $\beta 3_2$ (residues 195-202). The dimer is stabilised by a series of hydrogen bonds and hydrophobic interactions. All interactions at the interface are duplicated because of the approximate (NCS) 2-fold.

Subunit A	Subunit B	Distance (Å)
Arg-32-NH1	Glu-179-OE2	2.4
Arg-32 -NH1	Glu-179-OE1	2.6
Asp-36-OD1	Glu-179-OE2	2.5
Gln-50-NE2	Gln-181-OE2	2.4
Asp-178-OD1	Arg-32-NH1	2.9
Asp-178 -OD1	Arg-32-NH2	2.6
Met-196-O	Gln-29-NE2	3.0

Table 2.11: The ionic interactions and hydrogen bonds that stabilise the human R167Q-uPBGD dimer.

2.6.5 The Active Site of Human R167Q-uPBGD

The active site of human R167Q-uPBGD is located in a cleft at the interface between domains 1 and 2 and is very similar to that of the *E. coli* enzyme, in terms of architecture and location. The *N*-termini of helices $\alpha 1_2$ and $\alpha 2_2$ and their connecting loops make up the "ceiling" of the active site cleft, whereas the rear wall is made up one face of helix $\alpha 3_2$, the interdomain cross-over strands and their connecting loops. The "floor" of the cleft is made up of the *C*-termini of strands $\beta 1_1$ and $\beta 3_1$. As expected, the active site residues are highly invariant between the human and *E. coli* PBGD structures.

The DPM cofactor is covalently attached to Cys-261, which is located on a loop connecting helices $\alpha 1_3$ and $\beta 1_3$ in domain 3. Based on the observed electron density, the DPM factor adopts the reduced conformation with ring C2 positioned towards the rear of the cleft (Figure 2.18). The reduced conformation has previously been observed as the minor conformer (20 % occupancy) in the oxidised form of *E. coli* PBGD (Louie *et al.*, 1992) and at 100 % occupancy in the selenomethionyl variant crystallised under reducing conditions (Hädener *et al.*, 1992).

The acetate and propionate moieties of the DPM contribute most of the ionic interactions and hydrogen bonds to neighbouring active site side chains (Table 2.12; Figure 2.18). The interactions between the carboxylate groups of ring C1 and PBGD involve several invariant arginine residues (Arg-149, Arg-150 and Arg-173), the carbonyl and amide

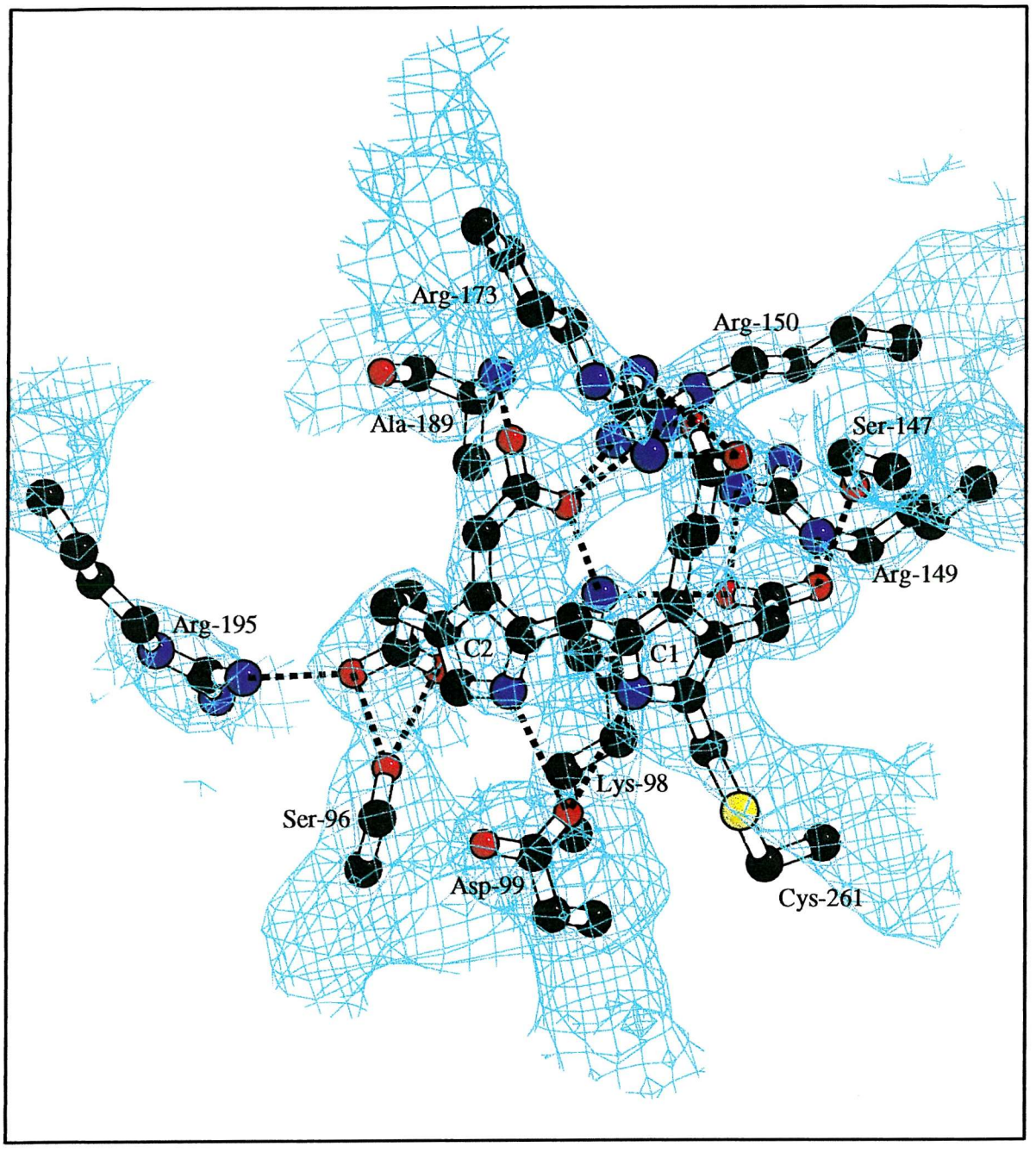


Figure 2.18: The interactions (shown in black) between the dipyrromethane (DPM) cofactor and protein side chains in the active site cleft of human R167Q-uPBGD (BOBSCRIPT). The DPM cofactor is covalently attached to Cys-261 and adopts the reduced conformation with ring C2 occupying a position at the rear of the cleft. The carboxylate groups of ring C1 form salt-bridges (Arg-149, Arg-150 and Arg-173) and hydrogen-bonds (the carbonyl and amide group of Ser-146 and the hydroxyl group of Ser-147). The carboxylate groups of ring C2 also form several salt bridges (Arg-150 and Arg-195) and hydrogen bonds (backbone amide nitrogens of Ala-189 and Gly-218). The importance of the invariant residues Lys-98 and Asp-99 is also apparent; Lys-98 forms salt-bridges with the acetate moieties, whereas one of the carboxylate oxygen atoms of Asp-99 forms hydrogen bonds to both pyrrole NH atoms. The 2Fo-Fc electron density map is contoured at the 1.3 σ level (depicted in sky-blue).

group of Ser-146 and the hydroxyl group of Ser-147. In contrast, ring C2 is positioned at the back of the cleft and interacts with Ser-96, Arg-150, Arg-195 and the backbone amide nitrogens of Ala-189 and Gly-218. The invariant Lys-98 and Asp-99 interact with both rings; Lys-98 forms salt-bridging interactions with the acetate moieties, whereas one of the carboxylate oxygen atoms of Asp-99 forms hydrogen bonds to both pyrrole NH atoms.

Protein-DPM interactions (Ring C1)	Inter-atomic distance (Å)	Protein-DPM interactions (Ring C2)	Inter-atomic distance (Å)
Lys-98 NZ — O1A	3.1	Ala-189 N — O1B	2.7
Arg-149 NH2 — O1A	3.0	Lys-98 NZ — O2B	2.7
Ser-147 OG — O2A	2.8	Arg-150 NH1 — O2B	3.1
Arg-149 NE — O2A	2.8	Arg-150 NH2 — O2B	2.8
Ser-146 O — O3A	3.1	Ser-96 OG — O3B	2.7
Ser-146 N — O3A	3.1	Arg-195 NH1 — O3B	2.6
Arg-173 NH2 — O3A	2.8	Ser-96 OG — O4B	3.2
Arg-173 NH1 — O3A	3.1	Gly-218 N — O4B	3.0
Arg-150 NH1 — O4A	3.0	Asp-99 OD1 — NB	3.2
Arg-173 NH1 — O4A	2.9		
Asp-99 OD1 — NA	3.2		

Table 2.12: The ionic interactions and hydrogen bonds between the DPM cofactor and active site amino acid side chains.

The side-chain of Gln-167 is disordered and therefore its precise location cannot be determined; there is electron density up to the CB atom which points towards the solvent, away from the DPM cofactor. The active site cleft also contains large *Fo-Fc* density (contoured at the 2 σ level), which appears to be closely associated with the DPM cofactor, but not covalently linked (Figure 2.19). It is possible that a PBG unit is present and aligned in a reversed orientation so that the position of the substrate exocyclic is relative to the free α-carbon of the DPM cofactor. In this orientation the substrate pyrrole nitrogen may form a hydrogen bond with Asp-99, whereas the acetate and propionate side chains may form hydrogen bonds and ion pairs with Ser-28, Arg-26 and Arg-173.

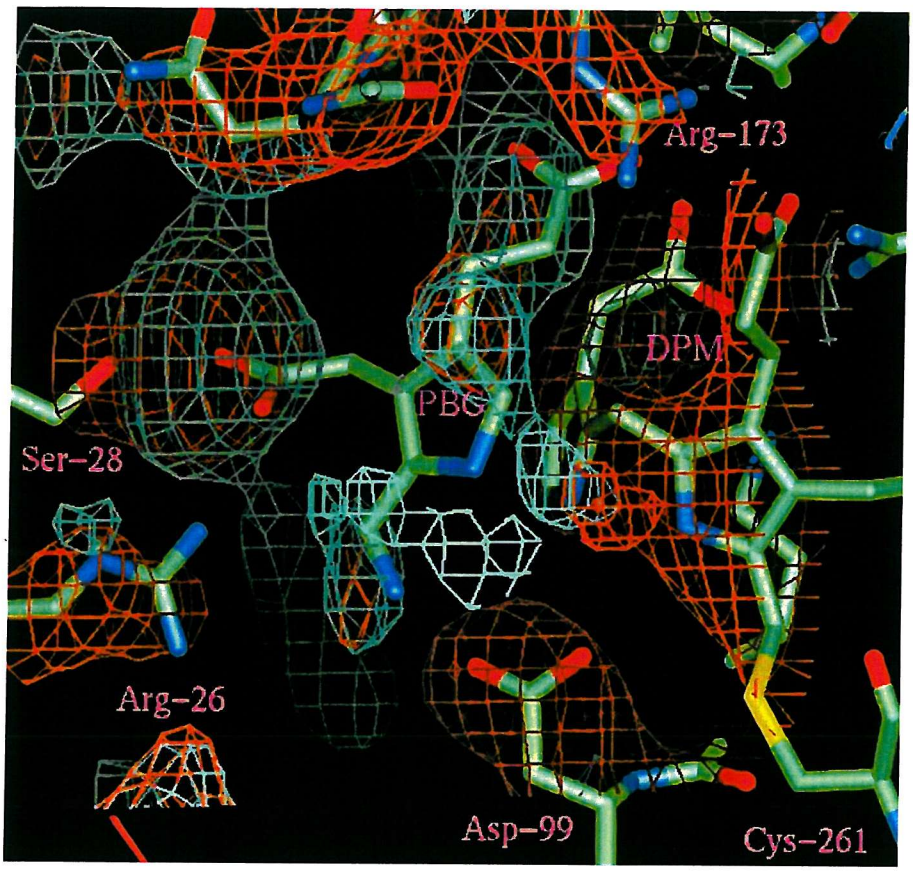


Figure 2.19: The active site cleft of R167Q-uPBGD (QUANTA). The large Fo-Fc density (contoured at the 2 σ level and depicted in green) is closely associated with the DPM cofactor, Arg-26, Ser-28 and Arg-173. The modelling of a PBG unit in a reversed orientation allows the exocyclic of the substrate to align in a position relative to the free α-carbon of the DPM cofactor. In this orientation the protonated amino group of substrate may form an ion pair with Asp-99, whereas the acetate and propionate side chains may form hydrogen-bonds and salt-bridges with Ser-28, Arg-26 and Arg-173. The *2Fo-Fc* electron density map is contoured at the 1 σ level (orange).

2.6.6 Conclusions

The three-dimensional structure of reduced human R167Q-uPBGD is the first human PBGD to be determined; previously the semi oxidised (Louie *et al.*, 1992; Louie *et al.*, 1996) and reduced (Hädener *et al.*, 1999) forms of *E. coli* PBGD were determined to 1.76 and 1.9 Å, respectively. More recently, the erythroid form of PBGD has been crystallised and solved by the molecular replacement method in this laboratory, using the human R167Q-uPBGD structure as the search model (Thompson; personal communication). The lack of a complete dataset resulted in a number of problems being encountered during the refinement of the human R167Q-uPBGD structure; the R-factor and R_{free} are still considerably higher than expected despite the building in of the majority of human PBGD residues (including the large insert) and the DPM cofactor.

Human R167Q-uPBGD is structurally similar to the *E. coli* enzyme with deviations mainly restricted to the large insert in domain 3 and several solvent exposed loop regions. Previous modelling studies of human PBGD based on the three-dimensional structure of *E. coli* PBGD suggested that the insert followed strand $\beta 2_3$ (Wood *et al.*, 1995). However, this is not the case for the human R167Q-uPBGD structure, as electron density for the insert is observed after strand $\beta 3_3$. The precise role of the insert is more likely structural rather than catalytic since it stabilises the interface between domains 1 and 3, contributes to the hydrophobic core of domain 3 and is involved in crystal contacts.

The active site architecture of human R167Q-uPBGD is very similar to that of *E. coli* PBGD since the residues of the active site cleft are highly conserved. As expected, the DPM cofactor is covalently linked to Cys-261, a residue located on a loop in domain 3. Based on the electron density for the cofactor and the colourless nature of the crystals, it is not surprising that the DPM cofactor is present in the reduced, active conformation. The ionic interactions of Arg-149 and Arg-150 with the DPM cofactor in human R167Q-uPBGD explain the results of previous site directed mutagenesis studies in which mutants R149Q and R150Q proved to be unstable apoenzymes with zero activity (Al-Dbass, 2001). The role of Lys-98 in the reaction catalysed by PBGD is uncertain although it forms ion pairs with both acetate moieties of the DPM cofactor. Mutation of the invariant lysine residue in *E. coli* PBGD prevents folding whatever amino acid is used. It is possible that

Lys-98 has a structural role in maintaining the cofactor in the correct conformation in the active site cleft.

Kinetic studies of the R167Q-uPBGD mutant reveal an enzyme that exhibits only 6-7 % of the activity of the wild-type enzyme (Al-Dbass, 2001). The disorder observed for the side chain of Gln-167 in the crystal structure may be explained by the conformational state of the DPM cofactor. It is possible that in the absence of ring C2 at the front of the cleft, Gln-167 may swing away from the cofactor and interact with the solvent. Evidence supporting this proposal is provided by structural studies of *E. coli* PBGD in the reduced form; Arg-149 (equivalent to Arg-167 in human PBGD) points away from the cofactor and forms a hydrogen bond with a water molecule (Hädener *et al.*, 1999). Since Gln-167 is disordered, the structure cannot confirm the proposed role of Arg-149/167 in binding to substrate and ES intermediates.

One of the properties of human R167Q-uPBGD that greatly differs from the wild type enzyme is the reduction in the pH optimum to 7 (normally 8.0-8.5). The PBG substrate possesses an internal ion pair between the acetate moiety and the protonated amine, which needs to be broken in order for the reaction to proceed. It is highly likely that the role of Arg-167 in the wild type enzyme involves distracting the carboxylate of the acetate group by forming a salt-bridge. As a result the internal ion pair is broken, which in turn frees the amino group and facilitates the initial deamination reaction (Jordan; personal communication). Unfortunately, the precise location of Gln-167 with respect to the large *Fo-Fc* density is unknown. Since the activity of R167Q-uPBGD is significantly reduced, in contrast to the wild-type enzyme, one can postulate that Gln-167 is much less effective at breaking the internal ion pair. The PBG internal ion pair interaction is possibly broken more easily at lower pH following the protonation of the acetate moiety of PBG, consistent with the reduction in the pH optimum.

Numerous attempts have been made to provide a greater insight into the nature of the disordered loop region which is located proximal to the active site. It has been proposed that the loop serves as a flexible lid that transiently shields the growing polypyrrole chain from solvent (Louie *et al.*, 1992). It was considered that this region would be more ordered in the crystal structure of R167Q-uPBGD if an ES complex were present.

However, the loop region is more disordered in the human structure (residues 57-78) at 2.65 Å than in the *E. coli* structure at 1.9 Å (residues 48-58). The disorder observed for the Phe-77 side-chain maybe correlated with the conformational state of the DPM cofactor; previous structural studies of *E. coli* PBGD have demonstrated that Phe-62 (equivalent to Phe-77 in human PBGD) is more ordered when the DPM cofactor is in the oxidised (inactive) state, rather than when it is in the reduced (active) form because it forms stacking interactions with ring C2.

The large *Fo-Fc* density present at the front of the cleft, close to the DPM cofactor, may arise from the presence of a PBG unit. It is possible that the oxidised form of DPM is present as a minor conformer and this would coincide with the C2 ring being positioned at the front of the active site cleft. Evidence for both DPM conformations being present in the crystal structure have been obtained previously with the oxidised form of *E. coli* PBGD (Louie *et al.*, 1992); data collected from a yellow coloured crystal contained a minor conformer in the reduced form (20 % occupancy) and a major conformer in the oxidised form (80 % occupancy). Moreover, one of the common problems associated with growing PBGD crystals under reduced conditions is that as the effects of DTT diminish with time, the crystals become subject to oxidation. However, when a PBG unit is modelled into the electron density at the front of the cleft (a position normally occupied by ring C2 in the oxidised state) the propionate moiety is too close to Ser-28.

Alternatively, the additional electron density may arise from a PBG unit docking at or near the S site. The R149H (in *E. coli* PBGD; Woodcock and Jordan, 1991) and R167Q (in human PBGD; Al-Dbass, 2001) mutants have been well characterised and are known to interrupt the assembly process following the formation of the ES complex. The accumulation of stable ES complexes with this mutant is most likely due to the interruption in the continuity of the positive charge by a glutamine residue. Therefore, it would not be entirely unexpected if an ES complex has accumulated in the active site cleft as a result of the R167Q mutation. However, there is no electron density linking the free α-carbon of the DPM cofactor with the putative substrate molecule, as one would expect if an ES complex were present. It is possible that the substrate molecule is present in the reversed orientation (with propionate and acetate side chain positions interchanged), so that the two reactive species are aligned in order for the reaction to take place.

The unusual behaviour of mutations associated with AIP can be better understood now that the structure of the actual human R167Q-uPBGD has been solved. Analysis of mutations that occur at the active site can easily be rationalised. For example, the R173W and R167Q (De Siervi *et al.*, 1999) mutations would be expected to disrupt the binding of the substrate and DPM cofactor. The mutations that occur away from the active site can be explained in terms of steric and electrostatic effects that most likely destabilise the three-dimensional fold of PBGD. For instance, the L177R and G111R (De Siervi *et al.*, 1999) mutations place a charged residue in the hydrophobic core of the protein that would be deleterious to its folding (Mgone *et al.*, 1992), whereas the V222M mutation (Mustajoki *et al.*, 1998) leads to disruption of the closely packed hydrophobic core of domain 1. The Q217L mutation (Schneider-Yin *et al.*, 2000) places a hydrophobic residue at the rear of the active site cleft leading to a disruption of the polar environment. The R116W (Solis *et al.*, 1999) and R201W (De Siervi *et al.*, 1999) mutations disrupt the salt-bridging interactions with Glu-250 and Asp-178 that help to stabilise the protein. However, in some cases the deleterious nature of the mutation is still not fully understood and requires further study.

2.7 Further Work

The final rounds of refinement and model building of the human R167Q-uPBGD are in progress. Differences between the two molecules within the asymmetric unit have been observed following the release of the NCS restraints. The structure also needs to be refined with the putative substrate molecule added. Larger crystals also need to be grown which hopefully will diffract to a higher resolution so that the difference density near the DPM cofactor is more interpretable. The structure of human erythroid PBGD is currently being refined to 2.8 Å and, together, these results will hopefully provide valuable insights into the mechanism of human PBGD. The structure of human R167Q-uPBGD in its active form also provides a platform for further studies of enzyme binding to substrate and/or inhibitors.

2.8 References

Akhtar, M. (1994) in *The Biosynthesis of the Tetrapyrrole Pigments,* (Ciba Foundation Symposium 180) 131-152, Wiley, Chichester, U.K.

Al-Dbass, A. (2001) Ph. D. Thesis, University of Southampton

Anderson, P.M. and Desnick, R.J. (1980) *J. Biol. Chem.* **255** 1993-1999

Attwood, T.K., Payne, A.W.R., Michie, A.D. and Parry-Smith, D.J. (1997) *EMBnet news.* **3** (3)

Bagust, J., Jordan, P.M., Kelley, M.E.M. and Kerkut, G.A. (1985) *Neurosci. Lett.* **21** S85

Bairoch, A. and Apweiler, R. (2000) *Nucleic. Acids. Res.* **28** 45-48

Battersby, A.R., Fookes, C.J.R. and Pandey, P.S. (1983) *Tetrahedron.* **39** 1919-1926

Beaumont, C., Porcher, C., Picat, C., Nordmann, Y. and Grandchamp, B. (1989) *J. Biol. Chem.* **264** 14829-14834

Bogorad, L. (1958) *J. Biol. Chem* **233** 501-509

Brownlie, P.D., Lambert, R., Louie, G.V., Jordan, P.M., Blundell, T.L., Warren, M.J., Cooper, J.B. and Wood, S.P. (1994) *Protein. Sci.* **3** 1644-1650

De Siervi, A., Roseetti, M.V., Parera, V.E., Astrin, K.H., Aizencang, G.I., Glass, I.A., Batlle, A.M.D. and Desnick, R.J. (1999) *Am. J. Med. Genet.* **86** 366-375

Desnick, R.J., Glass, I.A., Xu, M.W., Solis, S. and Astrin, K.H. (1998) *Seminars in Liver Disease* **18** 77-84

Ensouf, R. (1997) *J. Mol. Graphics* **15** 132

Ferreira, G.C., Franco, R., Lloyd, S.G., Moura, I., Moura, J.J.G. and Huynh, B.H. (1995) *J. Bioenerg. Biomembr.* **27 (2)** 221-229

Ferreira, G.C. and Gong, J. (1995) *J. Bioenerg. Biomembr.* **27(2)** 151-159

Grandchamp, B., de Verneuil, H., Beaumont, C., Chretien, S., Walter, O. and Nordmann, Y. (1987) *Eur. J. Biochem.* **162** 105-110

Hart, G.J., Leeper, F.J. and Battersby, A.R. (1984) *Biochem. J.* **222** 93-102

Hart, G.J., Miller, A.D., Leeper, F.J. and Battersby, A.R. (1987) *J. Chem. Soc. Chem. Commun.* 1762-1765

Hädener, A., Alefounder, P.R., Hart, G.J., Abell, C. and Battersby, A.R. (1990) *Biochem. J.* **271** 487-491

Hädener, A., Matzinger, P.K., Malashkevich, V.N., Louie, G.V., Wood, S.P., Oliver, P., Alefounder, P.R., Pitt, A.R., Abell, C. and Battersby, A.R. (1992) *Eur. J. Biochem.* **211** 615-624

Hädener, A., Matzinger, P.K., Battersby, A.R., McSweeney, S., Thompson, A.W., Hammersley, A.P., Harrop, S.J., Cassetta, A., Deacon, A., Hunter, W.N., Nieh, Y.P., Raftery, J., Hunter, N. and Helliwell, J.R. (1999) *Acta Crystallogr.* **D55** 631-643

Johnson, M.S. (1990) Modelling and Biocomputing group, Center for Biotechnology, P.O. Box 123, FIN-20521 Turku, Finland. MALIGN: ©1990 Kramsku, Finland

Jordan, P.M. and Woodcock, S.C. (1991) *Biochem. J.* **280** 445-449

Jordan, P.M. (1995) *J. Bioenerg. Biomembr.* **27** 191-195

Jordan, P.M. (1994) Royal Society of Chemistry Perkin, 4th Symposium on Organic Reactivity

Jordan, P.M. and Warren, M.J. (1987) *FEBS Lett.* **225** 87-92

Jordan, P.M. and Seehra, J.S. (1979) *FEBS Lett.* **104** 283-286

Kappas, A., Sassa, S., Galbraith, R.A. and Nordmann, Y. (1989) in *The Metabolic Basis of Inherited Diseases* 1305-1365, Eds. Scriver, C.R., *et al.*, McGraw-Hill, New York

Kannagra, C.G., Andersen, R.V., Pontopidan, B., Willows, R. and Von Wettstein, D. (1994) in *The Biosynthesis of the Tetrapyrrole Pigments* (Ciba Foundation Symposium 180) 3-20, Wiley, Chichester, U.K.

Keng, T., Richard, C. and Larocque, R. (1992) *Mol. Gen. Genet.* **234** 233-243

Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24** 946-950

Lambert, R., Brownlie, P.D., Woodcock, S.C., Louie, G.V., Cooper, J.B., Warren, M.J., Jordan, P.M., Blundell, T.L. and Wood, S.P. (1994) in *The Biosynthesis of the Tetrapyrrole Pigments* (Ciba Foundation Symposium 180) 97-105, Wiley, Chichester, U.K.

Lander, M., Pitts, A.R., Alefounder, P.R., Bardy, D., Abell, C. and Battersby, A.R. (1991) *Biochem. J.* **275** 447-452

Leeper, F.J. (1985) *Nat. Prod. Rep.* **2** 19-47

Leeper, F.J. (1994) in *The Biosynthesis of the Tetrapyrrole Pigments* (Foundation Symposium 180) 111-130, Wiley, Chichester, U.K.

Louie, G.V. (1993) *Curr. Opin. Struct. Biol.* **3** 401-408

Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C. and Jordan, P.M. (1992) *Nature* **359** 33-39

Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Malashkevich, V.N., Hadener, A., Warren, M.J. and Jordan, P.M. (1996) *Prot. Struct. Funct. Genet.* **25** 48-78

McCree, D.E. (1999) *J. Struct. Biol.* **125** 156-165

Mathews, B.W. (1968) *J. Mol. Biol.* **33** 491-497

May, A. and Bishop, D.F. (1998) *Haematologica* **83** 56-60

Meritt, E.A. and Bacon, D.J. (1997) *Meth. Enzymol.* **277** 505-524

Mgone, C.S., Lanyon, W.G., Moore, M.R. and Connor, J.M. (1992) *Human Genetics.* **90** 12-16

Miller, A.D., Hart, G.J., Packman, L.C. and Battersby, A.R. (1988) *Biochem. J.* **254** 915-918

Moore, M.R. (1993) *Int. J. Biochem.* **10** 1353-1368

Mustajoki, S., Pihlaja, H., Ahola, H., Peterson, N.E., Mustajoki, P. and Kauppinen, R. (1998) *Human Genetics.* **102 (5)** 541-548

Raich, N., Romeo, P.H., Dubart, A., Beaupain, D., Cohen-Sohal, M. and Goossens, M. (1986) *Nucleic Acids Res.* **14** 5955-5968

Sano, S. and Granick, S. (1961) *J. Biol. Chem.* **236** 1173-1180

Schneider-Yin, X., Bogard, C., Rufenacht, U.B., Puy, H., Nordmann, Y., Minder, E.I. and Deybach, J.C. (2000) *Human Heredity* **50 (4)** 247-250

Shemin, D. and Rusell, C.S. (1953) *J. Am. Chem. Soc.* **75** 4873

Solis, C., Lopez-Echaniz, I., Sefarty-Graneda, D., Astrin, K.H. and Desnick, R.J. (1999) *Molecular Medicine* **5 (10)** 664-671

Thomas, S.D. and Jordan, P.M. (1986) *Nucleic Acids Res.* **14** 6215-6226

Wang, A.L., Arredondovga, F.X., Giampiertro, P.F., Smith, M., Anderson, W.F. and Desnick, R.J. (1981) *Proceedings of the American Association for cancer research.* **22** 5734-5738

Wang, D., Driessen, H.P.C. and Tickel, I.J. (1991) *J. Mol. Graph.* **9** 50-52

Warren, M.J. and Jordan, P.M. (1988) *Biochemistry* **27** 9020-9030

Warren, M.J. and Scott, A.I. (1990) *Trends Biochem. Sci.* **180** 486-491

Warren, M.J., Gul, S., Aplin, R.T., Scott, A.I., Roessner, C.A., O'Grady, P.I and Jordan, P.M. (1995) *Biochemistry* **34** 11288-11295

Wood, S.P., Lambert, R. and Jordan, P.M. (1995) *Molecular Medicine Today.* **1 (5)** 232-239

Woodcock, S.C. and Jordan, P.M. (1994) *Biochemistry* **33** 2688-2695

Chapter 3

Structural Studies of Bovine Inositol Monophosphatase

3.1 Introduction to Inositol Monophosphatase

3.1.1 The Phosphatidylinositol (PI) Signalling Pathway

Following receptor mediated stimulation of phospholipase C, the membrane lipid, phosphatidylinositol 4,5-bisphosphate (PIP_2) is cleaved, generating secondary messengers, 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3) (Nishizuka, 1992). DAG directly activates protein kinase C, whereas intracellular signalling by IP_3 is achieved through binding to specific receptors located on the endoplasmic reticulum, leading to immobilisation of intracellular Ca^{2+} levels (Berridge, 1987).

The recycling of PIP_2 in the PI signalling pathway involves free inositol combining with cytidine monophosphorylphosphatidate (CMP-PA). CMP-PA is derived from DAG via the single intermediate, phosphatidic acid. In contrast, the pathway for recycling IP_3 and hence the termination of intracellular signalling is more complex involving a series of specific phosphatases and kinases (Figure 3.1). The phosphatases sequentially dephosphorylate inositol polyphosphate esters to their corresponding monophosphate esters (Berridge and Irvine, 1989). The intermediates produced during the recycling process may have specific physiological functions; inositol 1,3,4,5-tetrakisphosphate (IP_4) may control cellular Ca^{2+} fluxes (Irvine and Moor, 1986), whereas IP_5 and/or IP_6, are thought to be involved in extracellular signalling (Vallejo *et al.*, 1987).

The monophosphate esters serve as substrates for myo-inosoitol monophosphatase (IMPase, EC 3.1.3.25), a key enzyme in the PI signalling pathway. IMPase catalyses the dephosphorylation of D-inositol 1-phosphate (D-Ins(1)P), D-inositol 3-phosphate (D-Ins(3)P) and D-inositol 4-phosphate (D-Ins(4)P) to yield inorganic phosphate and inositol. The liberated inositol is reincorporated into phospholipids in order to maintain the PI signalling pathway (Berridge and Irvine, 1989). In addition, L-inositol 1-phosphate (L-Ins(1)P), an intermediate in the *de novo* synthesis of inositol from glucose-6-phosphate (Glu-6-P), also serves as a substrate for IMPase (Shears, 1989).

The generation of DAG and IP_3 is dependent upon a continuous supply of the precursor lipid PIP_2, which is maintained through continuous lipid resynthesis from high intracellular levels of inositol. Cells have access to three sources of inositol which include recycling by

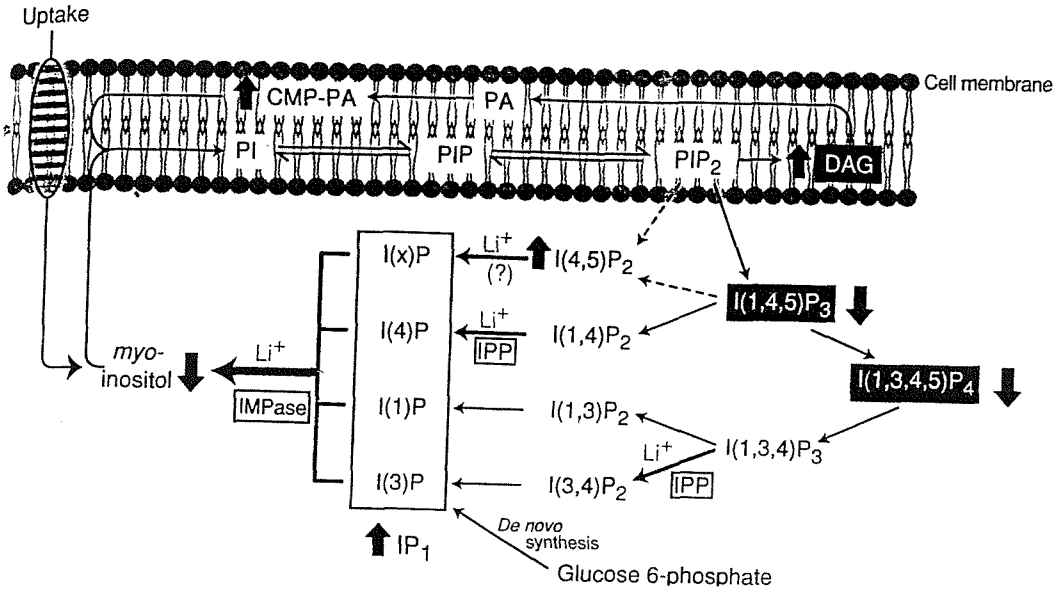


Figure 3.1: An overview of the phosphatidylinositol (PI) signalling pathway. The recycling of secondary messengers, inositol 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG) leads to the generation of inositol and cytidine monophosphorylphosphatidate (CMP-PA), which when combined form PI. Also shown are the cell's access to three sources of inositol (recycling via PI signalling pathway, uptake by plasma and de novo synthesis from glucose 6-phosphate). Pathways which are sensitive to lithium are labelled as Li^+ (Atack *et al.*, 1995b).

the PI signalling pathway, *de novo* synthesis from Glu-6-P and uptake from plasma (Figure 3.1). In contrast to the first two mechanisms, the latter is insensitive to the inhibitory effects of lithium (Berridge *et al.*, 1989).

3.1.2 Bipolar Disorder

Bipolar disorder is a mental illness characterised by extreme mood swings from mania to depression. The disorder can be classified into two main types, Bipolar I and Bipolar II. Diagnosis for type I bipolar disorder is achieved when an individual, lacking a previous history of depressive episodes, experiences a manic episode. Symptoms associated with manic episodes include increased self-esteem, euphoric or irritable mood, distractibility, agitation, decreased need for sleep and pressured speech and in severe cases, individuals become delusional and suffer from hallucinations (Karasu *et al.*, 2000).

The diagnosis for type II Bipolar disorder can be made when an individual experiences a depressive episode following or preceding a hypomanic episode. Patients with this type of disorder exhibit mild manic episodes and are often misdiagnosed as suffering from clinical depression. The symptoms of a hypomanic episode are similar to that of a manic episode with few exceptions; individuals experiencing a hypomanic episode do not exhibit psychotic features such as hallucinations or require hospitalisation. The depressive episode normally lasts up to two weeks and is characterised by a depressed mood, lack of interest, indecisiveness, sleep difficulties, excessive guilt and recurrent thoughts of death (Karasu *et al.*, 2000).

Despite extensive research, the exact etiology of bipolar disorder is unknown. At present, therapy for bipolar disorder involves treating the acute symptoms and preventing relapse pharmacologically. The most widely used drugs for stabilising manic-depressive mood swings are mood stabilisers which include lithium, sodium valproate and carbamazepine (Garza-Trevino, 1997).

3.1.3 Lithium Therapy

The Australian psychiatrist John Cade first pioneered the use of lithium, a small alkali metal, in the treatment of bipolar disorder in 1949. By maintaining lithium serum levels of 1 mM, manic-depressive episodes can be modulated (Rosenthal and Goodwin, 1982). Additional roles for lithium have more recently included regulating neurological disorders such as aggressive and self mutilating behaviour (Wickham and Reed, 1987). The precise manner in which lithium influences neurotransmission within the central nervous system is disputed. Popular opinion suggests that lithium modifies neurotransmission by disrupting the receptor-coupled PI signalling pathway in the brain (Emilien *et al.*, 1995). Alternative possibilities for the therapeutic action of lithium include its effects on electrolytes and ion transport and its effects on neurotransmitters and their release (Katzung, 1995).

Evidence supporting the proposal that lithium influences secondary messengers is provided by the ability of lithium to uncompetitively inhibit IMPase, a critical enzyme catalysing the rate limiting step of inositol recycling in the brain. Lithium inhibition effect of IMPase reduces intracellular levels of inositol required to maintain membrane inositol lipids, which in turn attenuates secondary messenger signalling. This phenomenon is commonly referred to as the 'inositol depletion hypothesis' (Berridge *et al.*, 1989). Further support for this hypothesis is provided by the fact that the permeability of the blood brain barrier towards inositol is low, thus blocking the supply of inositol to the brain. However, cells outside the brain are protected from the effects of lithium by access to plasma inositol. Therefore, by acting as a buffer, lithium reduces neural secondary messenger signalling to within its normal range without affecting other physiological mechanisms (Berridge *et al.*, 1989).

Owing to the side effects accompanying lithium therapy, lithium serum concentrations have to be monitored and maintained between a normal range (0.5-1.0 mM). Side effects include hand tremor, weight gain, thirst and increased urine production. At higher concentrations (2 mM) the side effects my include nausea, diarrhoea, confusion and slurred speech. Unfortunately, patients with lithium serum concentrations in excess of 2 mM may end up in a coma, eventually leading to death. In addition, there has been some concern that long-term lithium therapy may cause renal damage (American Psychiatric Association, 1994). These side effects prevent a large proportion of bipolar patients from tolerating lithium therapy. Non-compliance is a major cause of treatment failure in bipolar patients;

approximately 20-40 % fail to display an adequate anti-manic response to lithium following lithium therapy (Manji *et al.*, 1996). Therefore, a novel approach in treating this disorder involves designing drugs that mimic the therapeutic effects of lithium (Atack *et al.*, 1995b). Recent studies have focused on the use of anticonvulsants (valproate and carbamazepine) in the treatment of mania and depression. Valproate may mimic lithium by stimulating glutamate release, resulting in the activation of postsynaptic receptors and Ins(1,4,5)P_3 accumulation (Dixon and Hokin, 1997).

3.1.4 Functional Studies of IMPase

Chen and Charalampous (1966) first studied IMPase in yeast and established that the enzyme was involved in the *de novo* synthesis of inositol from Glu-6-P. The ability of IMPase from rat testes to hydrolyse both enantiomers of Ins(1)P was reported by Eisenberg (1967). These results were later confirmed for bovine brain IMPase (Hallcher and Sherman, 1980; Gee *et al.*, 1988). IMPase has since been isolated and purified from a variety of sources including rat brain (Takimoto *et al.*, 1985), human brain (McAllister *et al.*, 1992), bovine brain (Gee *et al.*, 1988; Attwood *et al.*, 1988) and rat testes (Parthasarathy *et al.*, 1993). The bovine and human enzymes are made up of homodimers with a subunit molecular weight of 30 kDa (Gee *et al.*, 1988).

The release of inositol from IMPase has been demonstrated in the presence of lithium, following Ins(1)P hydrolysis. This suggests that lithium binds to the enzyme after the phospho ester bond is cleaved, forming interactions with either the phosphate bound form of the enzyme or a phosphorylated enzyme (E-P) intermediate (Shute *et al.*, 1988). However, attempts to trap an E-P intermediate have been unsuccessful, supporting the hypothesis that an activated water molecule rather than an active site residue, serves as the nucleophile. Limited proteolysis studies on IMPase carried out in the presence of trypsin (Greasley *et al.*, 1993) and endoproteinase Lys-C (Whiting *et al.*, 1990) have demonstrated that the bond between Lys-36 and Ser-37 was susceptible to hydrolysis resulting in a simultaneous loss of enzyme activity. The rate at which this bond was cleaved decreased in the presence of substrates and ligands, implying that a conformational change is transmitted to this region of the molecule upon ligand binding.

The Mg^{2+} dependence of IMPase has been well documented (Hallcher and Sherman, 1980; Gee *et al.*, 1988). At relatively high concentrations, Mg^{2+} is an uncompetitive inhibitor with respect to the substrate (Ganzhorn and Chanal, 1990), similar to lithium, whereas at lower concentrations, Mg^{2+} is a noncompetitive activator (Hallcher and Sherman, 1980). The behaviour of Mg^{2+} at different concentrations indicates two different mechanisms of binding. Previous kinetic studies in the presence of Ins(1)P provided Hill coefficients of 1.9 for IMPase, demonstrating that the binding of Mg^{2+} is co-operative (Ganzhorn and Chanal, 1990). Furthermore, the binding of Mg^{2+} to IMPase was not dependent on the presence of the substrate (Pollack *et al.*, 1993). In contrast, Mg^{2+} was required in order for phosphate to bind to the enzyme (Greasley and Gore, 1993). Together these observations imply that binding of Mg^{2+} must precede and follow substrate binding, consistent with a two metal mechanism.

The existence of two metal binding sites was confirmed by fluorescence studies on bovine IMPase labelled with pyrene-maleimide (a thiol directed fluorescent reagent). By monitoring the fluorescence upon binding of Mg^{2+} to the pyrene labelled enzyme, a high affinity site (K_d=300 μm) was identified (Greasley *et al.*, 1994). This established the existence of two Mg^{2+} binding sites, since a lower affinity site (K_m=3 mM) had been determined previously (Gore *et al.*, 1993). Further investigations of metal binding to the high affinity site were carried out using stopped flow fluorescence spectroscopy. By swiftly mixing Mg^{2+} with the pyrene labelled enzyme, an initial rapid increase of fluorescence intensity, possibly induced by the formation of the E-Mg^{2+} complex, was observed followed by a slower increase in fluorescence. The slow phase was independent of Mg^{2+} concentration and provided evidence for a possible rate limiting conformational change following the binding of Mg^{2+} at the high affinity site (Thorne *et al.*, 1996).

Two metal binding sites in bovine IMPase were also detected using circular dichroism (CD) spectroscopy (Rees-Milton *et al.*, 1997). Spectral changes in the near ultra violet region (275 nm) were observed following the binding of Mg^{2+} to the high affinity site (site 1). In contrast, spectral perturbations in the far ultra violet region (225 nm) demonstrated Mg^{2+} binding to the lower affinity site (site 2). In addition, high lithium concentrations perturbed the far ultra violet CD signal, confirming that lithium binds at the lower affinity site (Rees-Milton *et al.*, 1997).

3.1.5 Sequence Studies of IMPase and Related Proteins

Sequence alignments of IMPase from a variety of sources demonstrate a high degree of sequence homology (Figure 3.2). Sequence identity between the human and bovine enzymes is high (89 %), whereas homology between these enzymes and tomato IMPase is considerably less (43 %). Sequence homology between IMPase and other phosphatases is uncommon, although small stretches of conserved sequences have been identified (motif A (87-100), motif B (219-233) and motif C (69-72)) (Atack *et al.*, 1995a). These conserved motifs are comprised of residues that are critical for substrate and metal binding as well as water activation.

These invariant regions have also been identified in other phosphatases (fructose 1,6-bisphosphatase and inositol polyphosphate 1-phosphatase) and the products of several diverse genes. The latter include the qa-x protein of *Neurospora crassa*, the qutG protein of *Aspergillus nidulans* and the suhB protein of *E. coli* (Bone *et al.*, 1992). These proteins possess homologous hydrophobic regions as well as conserved motif regions A and B. Also present in some of the proteins is a conserved glutamate residue (213) responsible for substrate binding (Bone *et al.*, 1994a). Based on these observations it is apparent that these phosphatases are members of the same structural family and contain a conserved active site.

3.1.6 Crystallographic Studies of Human Recombinant IMPase

The X-ray structure of human recombinant IMPase was determined by MIR to 2.1 Å resolution, complexed with inhibitory gadolinium (Gd^{3+}) and sulphate (Bone *et al.*, 1992). Human IMPase is an α/β protein composed of two subunits, each consisting of 273 residues. Each IMPase monomer is comprised of alternating layers of α helices and β sheet forming a core penta layered $\alpha\beta\alpha\beta\alpha$ sandwich. The interactions between the interface of each subunit are extensive; the dimer is stabilised by 18 hydrogen bonds between the two subunits (Bone *et al.*, 1992).

The X-ray structure revealed the presence of a single Gd^{3+} binding site per subunit, with 6-9 potential metal ligands. In order to determine the exact location of the inositol binding site, crystals of human IMPase complexed with D and L-Ins(1)P were grown (Bone *et al.*,

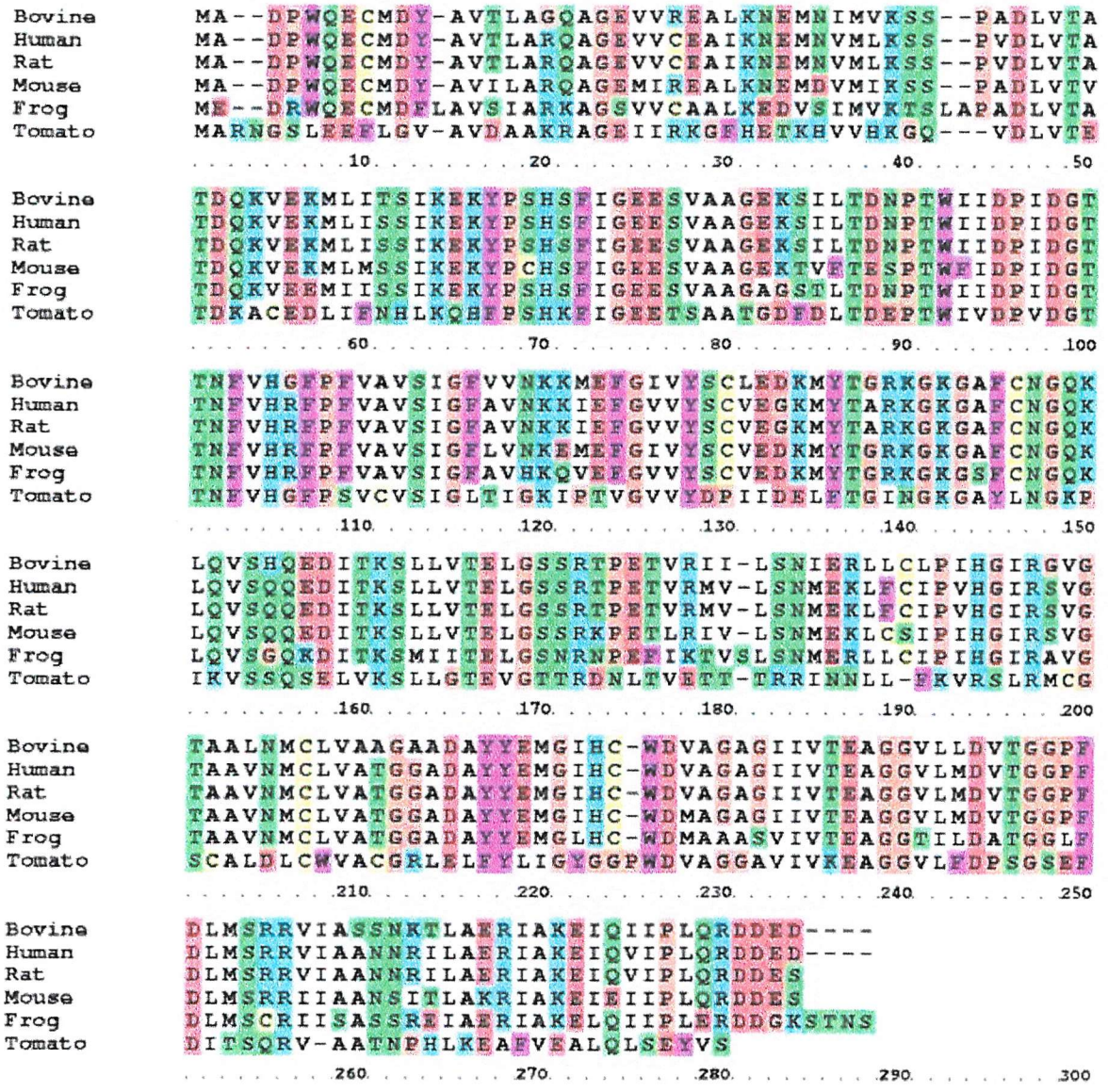


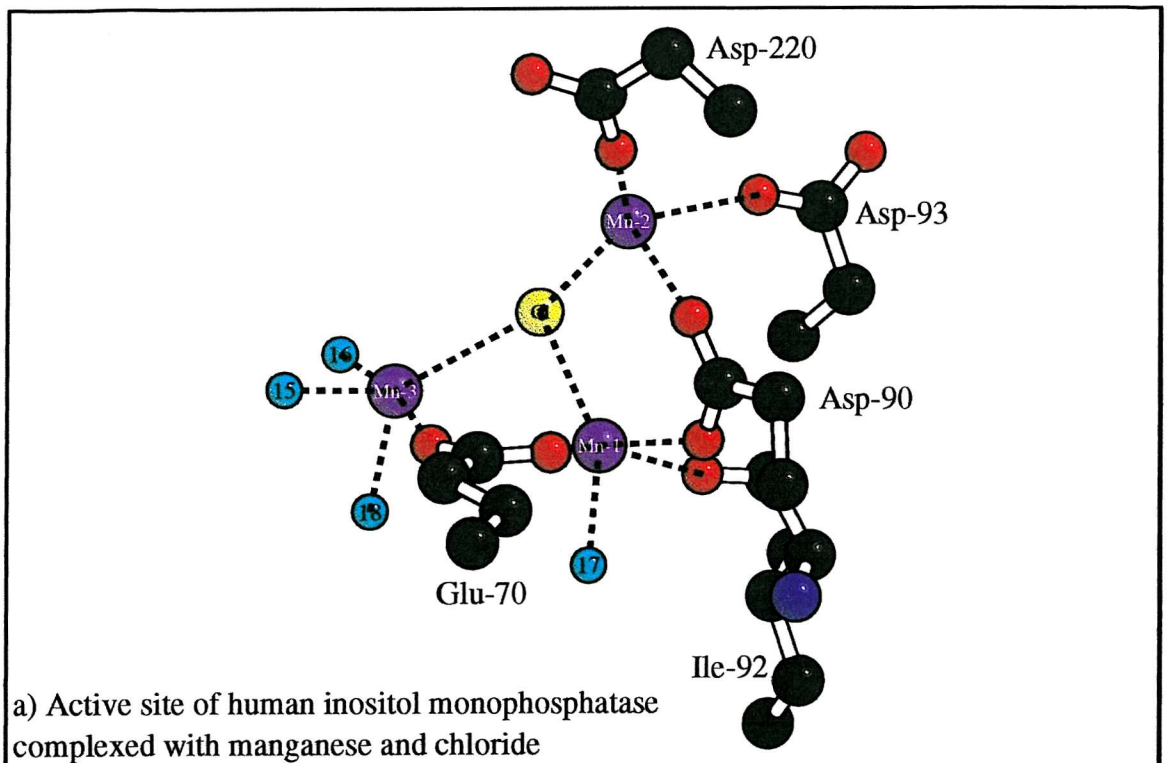
Figure 3.2: A primary sequence alignment of inositol monophosphatase (IMPase) from several species (CINEMA 2.1). The alignment was performed using MALIGN. The colour key for residue types is the following: polar positive (blue), polar negative (red), polar neutral (green), non-polar aliphatic (white), non-polar aromatic (purple), glycine and proline (brown) and cysteine (yellow). The sequences were obtained from the SWISS-PROT/TrEMBL protein sequence database.

1994a). In the presence of D-Ins(1)P, Asp-93 and Ala-196 form hydrogen bonds with the axial 2-hydroxyl group (2-OH), while Asp-220 interacts with the 6-hydroxyl group (6-OH). Furthermore, the highly conserved Glu-213 hydrogen bonds with the 4-hydroxyl group (4-OH). The interactions between L-Ins(1)P and the enzyme are similar with the exception that the 6-OH group rather than the 2-OH group hydrogen bonds to Asp-93 and Ala-196 (Bone *et al.*, 1994a).

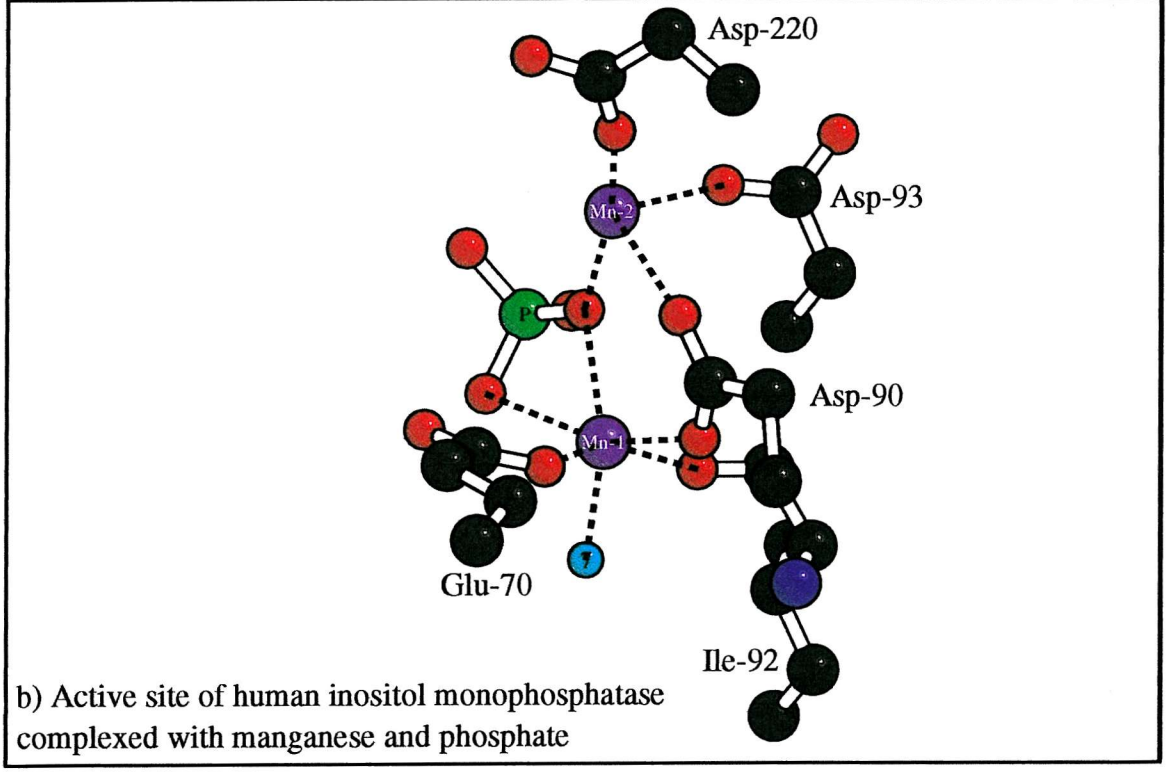
Unfortunately these structures proved to be ambiguous in demonstrating the number of metal binding sites present within the active site. This was due to the fact that Gd^{3+} is a larger trivalent metal, typically with 2-3 additional metal ligands than Mg^{2+}, and, hence, would not be expected to bind at equivalent sites. Furthermore, the existence of a single Gd^{3+} binding site per subunit suggested that the high lithium concentration present in the crystallisation may have promoted lithium occupation at other metal sites. However, direct observation of the lithium binding site within these structures is difficult since lithium possesses only two electrons and is practically invisible in an X-ray crystallographic structure except at atomic resolution (Bone *et al.*, 1992; Bone *et al.*, 1994a).

To address this concern and further elucidate the metal binding sites, the structure of human IMPase complexed with manganese (Mn^{2+}) was solved to 2.6 Å resolution (Bone *et al.*, 1994b). The human IMPase-Mn^{2+} structure identified 3 Mn^{2+} and a single chloride ion within the active site of each subunit (Figure 3.3a). Metal co-ordination site 1 (Mn-1) was located near the carboxylates of Glu-70 and Asp-90 and the carbonyl group of Ile-92 (previously identified as the Gd^{3+} binding site). Metal co-ordination site 2 (Mn-2) was located near the carboxylates of Asp-90, Asp-93 and Asp-220. The third metal co-ordination site (Mn-3) was located in the proximity of Glu-70.

The metal interactions were further investigated by determining the structure of human IMPase in the presence of phosphate and Mn^{2+}. This structure revealed the displacement of Mn-3 upon phosphate binding, suggesting that this site is artifactual (Figure 3.3b). However, Mn-1 and Mn-2 were retained with octahedral and tetrahedral geometry, respectively (Bone *et al.*, 1994b). More recently, the human IMPase structure has been solved in the presence of Ca^{2+} and D-Ins(1)P and possesses three metal binding sites (Ca-1, Ca-2 and Ca-3) (Ganzhorn and Rondeau, 1997).



a) Active site of human inositol monophosphatase complexed with manganese and chloride



b) Active site of human inositol monophosphatase complexed with manganese and phosphate

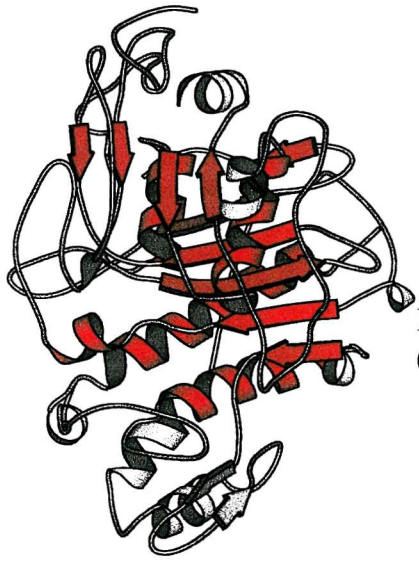
Figure 3.3a: The Mn^{2+} complex contains 3 Mn^{2+} ions (depicted in purple; Mn-1, Mn-2 and Mn-3) and one chloride ion (Cl; depicted in yellow) bound in each active site. The chloride is a ligand to each of the Mn^{2+} ions. Further ligands to Mn-1 (previously identified as the gadolinium binding site) include the side chains of Glu-70 and Asp-90, the carbonyl group of Ile-92, and a water molecule (Wat-17). Ligands to Mn-2 include the side chains of Asp-90, Asp-93 and Asp-220 (a total of 4 ligands including chloride). Mn-3 is co-ordinated to Glu-70, Wat-15, Wat-16 and Wat-18 (a total of 5 ligands including the chloride). b: When soaking in phosphate (depicted in green), the chloride and Mn-3 ions are displaced leaving the metal ions with octahedral (Mn-1) and tetrahedral (Mn-2) co-ordination geometry. The water molecules are depicted in blue (BOBSCRIPT).

Structural comparisons of human IMPase with other phosphatases deposited in the Brookhaven PDB demonstrate similar topology to two other phosphatases, namely inositol polyphosphate 1-phosphatase (IPP) (York *et al.*, 1994) and fructose 1,6-bisphosphatase (F1,6-BP) (Ke *et al.*, 1989). All three proteins have the same fold consisting of alternating layers of α helices and β sheet, forming a core five-layered αβαβα sandwich (Figure 3.4). Furthermore, both IPP and F1,6-BP are Mg^{2+} dependent with structurally conserved metal binding sites and are sensitive to lithium inhibition.

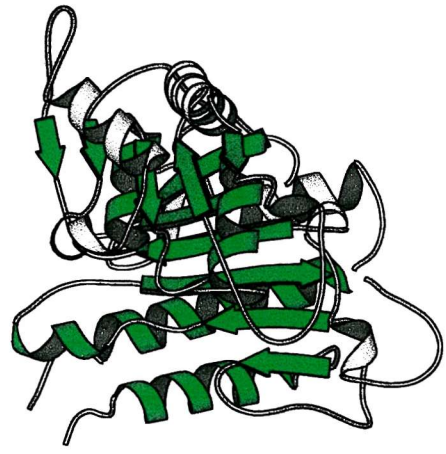
3.1.7 Mechanism of IMPase

The most widely accepted mechanism begins with the binding of Mg^{2+} ion to metal site 1 resulting in the activation of the enzyme (Figure 3.5). The substrate Ins(1)-P upon binding to the enzyme interacts with the bound Mg^{2+} at site 1 via the phosphate group. Following substrate binding, a second Mg^{2+} ion binds to metal binding site 2, allowing an activated water nucleophile to attack the substrate phosphorous causing cleavage of the phospho ester bond. The water molecule may be activated by forming hydrogen-bonding interactions with metal site 1, Glu-70 and possibly Thr-95.

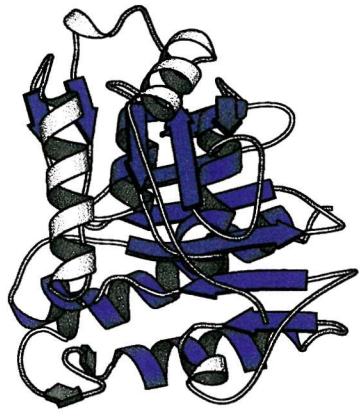
Following cleavage of the phosphate bond the inositol dissociates, leaving phosphate ligated to both Mg^{2+} ions. Leaving group expulsion is aided by the stabilisation of the negative charge by the second Mg^{2+} ion as the phosphate ester is cleaved. Following dissociation of Mg^{2+} at metal site 2, phosphate is released thereby regenerating IMPase for the next catalytic cycle. However, in the presence of lithium, metal site 2 is occupied by lithium, inhibiting the dissociation of the phosphate to form an enzyme-Mg^{2+}-phosphate-lithium complex. This complex is very stable and as a result the phosphate is trapped, effectively inhibiting IMPase activity (Attack *et al.*, 1995a).



Inositol Polyphosphate 1-Phosphatase (bovine)



Fructose-1,6-Bisphosphatase (porcine)



Inositol Monophosphatase (human)

Figure 3.4: The three-dimensional structures of phosphatases (MOLSCRIPT). Each shares a common core structure comprised of 5 α helices and 11 β strands (depicted in red for bovine inositol polyphosphate 1-phosphatase, green for porcine fructose-1,6-bisphosphatase and blue for human inositol monophosphatase.

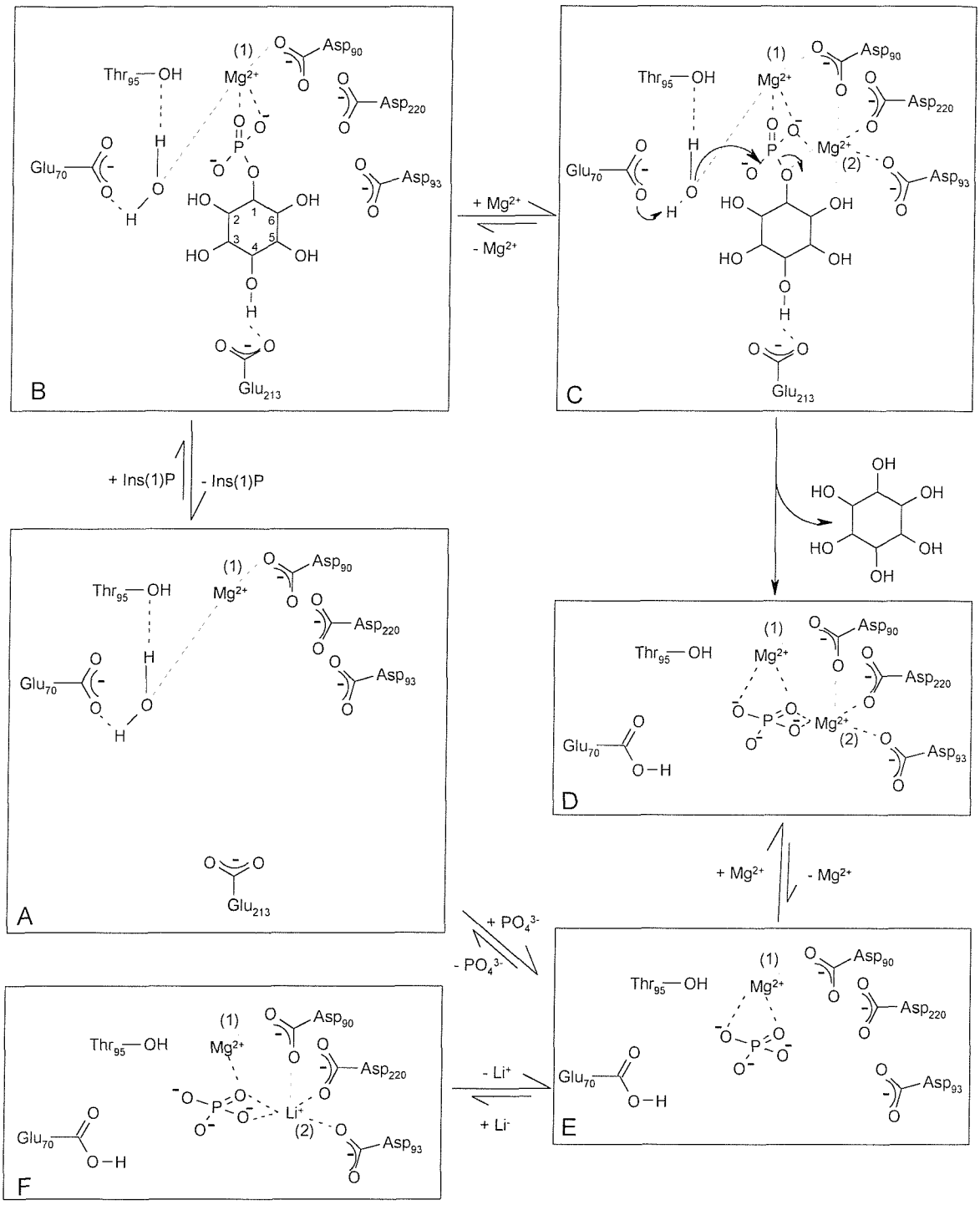


Figure 3.5: The most widely accepted mechanism for the reaction catalysed by inositol monophosphatase (Atack *et al.*, 1995a). A) Mg^{2+} ion is present at site 1 in the absence of substrate. B) Substrate (Ins(1)P) binds. C) The binding of second Mg^{2+} at site 2 leads to an activated water molecule attacking the phosphate phosphorous, resulting in phosphate bond cleavage. D) Following inositol debinding, phosphate forms ligands to both Mg^{2+} ions. E) Mg^{2+} at site 2 dissociates leaving an enzyme phosphate complex in which phosphate is ligated to Mg^{2+} at site 1. F) Normally, phosphate debinds regenerating IMPase with a single Mg^{2+} at site 1 ready for the next catalytic cycle. However, in the presence of lithium, metal site 2 becomes occupied with lithium leaving the phosphate trapped. This complex is very stable resulting in the inhibition of the enzyme.

3.1.8 Research Aims

IMPase plays a particularly important role in the brain since plasma inositol cannot transverse the blood-brain barrier, and its inhibition by physiological levels of lithium (1 mM) has made it a likely candidate for the target of lithium therapy in the treatment of bipolar disorder. However, the side affects accompanying lithium therapy and non-compliance are major causes of treatment failure in bipolar patients. Several substrate-based inhibitors have been developed in an attempt to search for better therapeutic agents for bipolar disorder. To provide a structural basis for rational drug design, the three-dimensional structure of bovine IMPase has been determined in complex with its natural ligand, Mg^{2+}.

3.2 Expression, Purification and Crystallisation of Bovine IMPase

Bovine IMPase was expressed in *E. coli* and purified as described by Rajji Badyal (2000). The protein was crystallised at room temperature using the hanging drop vapour diffusion method. Hanging drops were obtained by mixing equal volumes (2 μl) of protein solution (20 mg/ml bovine IMPase in 20 mM Tris HCl pH 7.5 containing 20mM $MgCl_2$) with reservoir solution (0.1 M sodium acetate, 0.1 M sodium HEPES pH 8.5 and 15 % PEG 4000).

3.3 Data Collection and Processing of Bovine IMPase

X-ray data were collected from a single bovine IMPase crystal (0.6 x 0.4 x 0.1 mm^3) using a Mar CCD detector at the ESRF (Grenoble, France) beamline ID14-EH3 (λ= 0.931 Å). The crystal was flash cooled initially in liquid ethane and data were collected at 100 K using an Oxford Cryosystem cryostream cooler. The cryoprotectant was determined to be mother liquor containing 30 % (v/v) glycerol. Successful flash-freezing was obtained when the crystal was transferred from 0 % to 30 % glycerol in steps of 7.5 %. The crystal diffracted X-rays to a resolution of 1.65 Å (Figure 3.6) and was found to belong to the triclinic space group P1 (a=43.1Å, b=54.8Å, c=59.3Å, α=68.5°, β=69.2°, γ=80.7°). A total of 180 1°-oscillation frames were collected with an exposure time of 60 s per frame and a crystal to detector distance of 100 mm. Intensity data were processed with MOSFLM, and sorted, scaled and merged using the CCP4 package of programmes. The data processing statistics are listed in Table 3.1.

Total number of reflections	125405
Number of unique reflections	53444
Resolution (Å)	1.65
Completeness (%)	94.2
R_{merge} (%) (outer resolution shell)	8.9 (25.3)
Multiplicity	1.9
Average I/σ(I) (outer resolution shell)	5.1 (1.4)

Table 3.1: Relevant data processing statistics for bovine IMPase crystal.

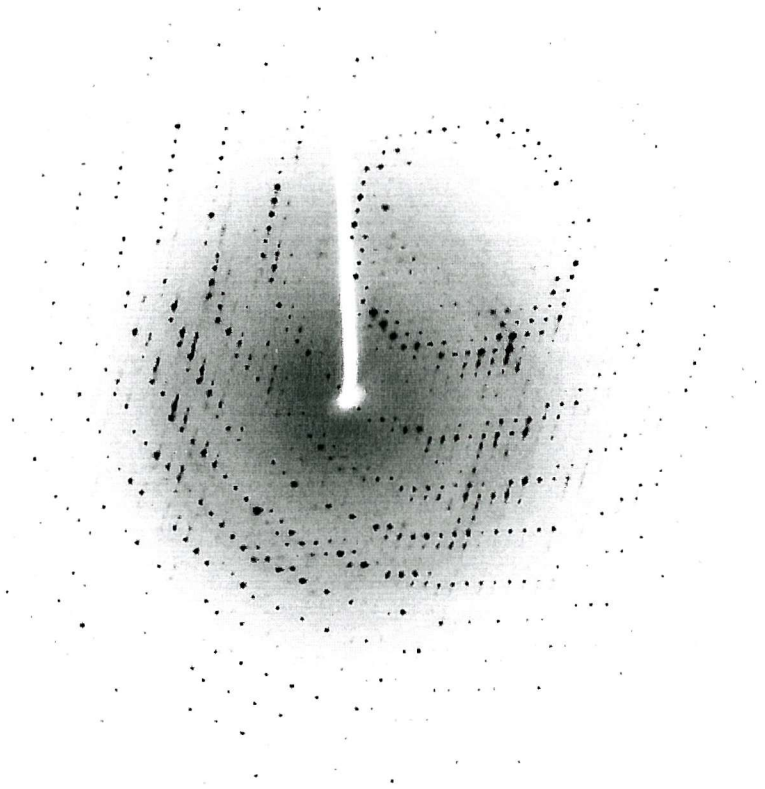


Figure 3.6: A 1° oscillation image of bovine IMPase crystal collected at 100 K with a Mar CCD detector at the ESRF (Grenoble, France) beamline ID14-EH3 (λ= 0.931 Å). The crystal diffracted X-rays to a resolution of 1.65 Å and crystallises in the triclinic (P1) space group. The exposure time was 60 s with the crystal-to-detector set at 100 mm.

3.4 Molecular Replacement Studies of Bovine IMPase

3.4.1 Introduction

Assuming there are two IMPase subunits (60 kDa) per asymmetric unit (V_m=2.55 $Å^3$/Da), a solvent content of 52 % was calculated. This assumption is consistent with the crystal form obtained for human recombinant IMPase, which also contains a dimer within the asymmetric unit (Bone *et al.*, 1992). The human IMPase dimer (accession code 2HHM; Bone *et al.*, 1992) refined to 2.1 Å resolution was used as the search model in the structural analysis of bovine IMPase.

3.4.2 Cross Rotation Search

The search model (dimensions x=72 Å, y= 48 Å and z=58 Å) with the metal ion co-ordination sphere omitted was placed in an artificial P1 cell. Data in the 10 to 3.5 Å resolution range were used in the reciprocal space cross rotation calculations performed with the program AMoRe. The cross rotation search calculated with a radius of integration of 30 Å (Table 3.2) generated a strong rotation function solution (α=205.82°, β=13.00°, γ=54.94°). This solution, with a peak height of 10 σ above the noise (with next peak being at 5 σ), corresponded to the orientation of the bovine IMPase dimer in the target asymmetric unit. Several other choices of integration radii gave the same highest peak. The translation search was unnecessary since the choice of the origin is arbitrary in P1.

Peak number	α (°)	β (°)	γ (°)	Peak height (σ)
1	205.82	13.00	54.94	10.0
2	94.00	34.91	218.50	5.2
3	112.06	52.16	100.30	5.0
4	47.17	71.08	97.89	5.0
5	77.62	43.06	133.36	4.8

Table 3.2: The 5 highest peaks following the cross rotation search performed with AMoRe. Data between 6.0 Å and 3.5 Å were used.

3.4.3 Cross Validation of Cross Rotation Function Solution Using X-PLOR

The cross rotation search was then repeated using X-PLOR (Brünger, 1992) to see if this molecular replacement package could validate the solution found using AMoRe. Reflections within the resolution range of 10 Å to 2.5 Å were used in the cross rotation calculations, with a radius of integration of 35 Å. The search model was placed in an artificial P1 cell with dimensions a=102 Å, b=78 Å, c=88 Å, $\alpha=\beta=\gamma=90°$. The 5 highest peaks from the cross rotation search are listed in Table 3.3.

Peak Number	θ_1 (°)	θ_2 (°)	θ_3 (°)	RF-function
1	157.06	157.85	73.34	2.07
2	220.37	16.62	239.32	2.05
26	139.02	85.85	17.67	1.49
44	96.43	58.15	113.31	1.36
51	99.21	99.69	167.61	1.31

Table 3.3: The 5 highest peaks following a real space cross rotation search calculated with X-PLOR. θ_1, θ_2, and θ_3 correspond to X-PLOR Eulerian angles.

The top 40 peaks from the cross rotation search were used as input for PC refinement. The advantage of the PC method is that as it refines the rotation function peaks, individual domains are treated as rigid bodies and allowed to move. The PC values for each rotation function solution, before and after PC refinement were plotted using PCPLOT (Figure 3.7). In order to compare the rotation peaks, the XPLOR Eulerian angles (θ_1, θ_2, θ_3) were converted to CCP4 format (α, β, γ) as follows

$\theta_1=270-\gamma$, $\theta_2=\beta$, $\theta_3=90-\alpha$

X-PLOR Eulerian angles can also be converted to CCP4 format and vice versa using ROTMAT, a program from the CCP4 suite. Therefore, peaks 1 (157.06°, 157.85°, 73.34°) and 2 (220.37°, 16.62°, 239.32°), expressed as X-PLOR Eulerian angles, corresponded to (α=16.66°, β=157.85°, γ=112.94°) and (α=210.8°, β=16.62°, γ=49.63°) in CCP4 format. The latter orientation is consistent with the orientation calculated using AMoRe (α=205.82°, β=13.00°, γ=54.94°).

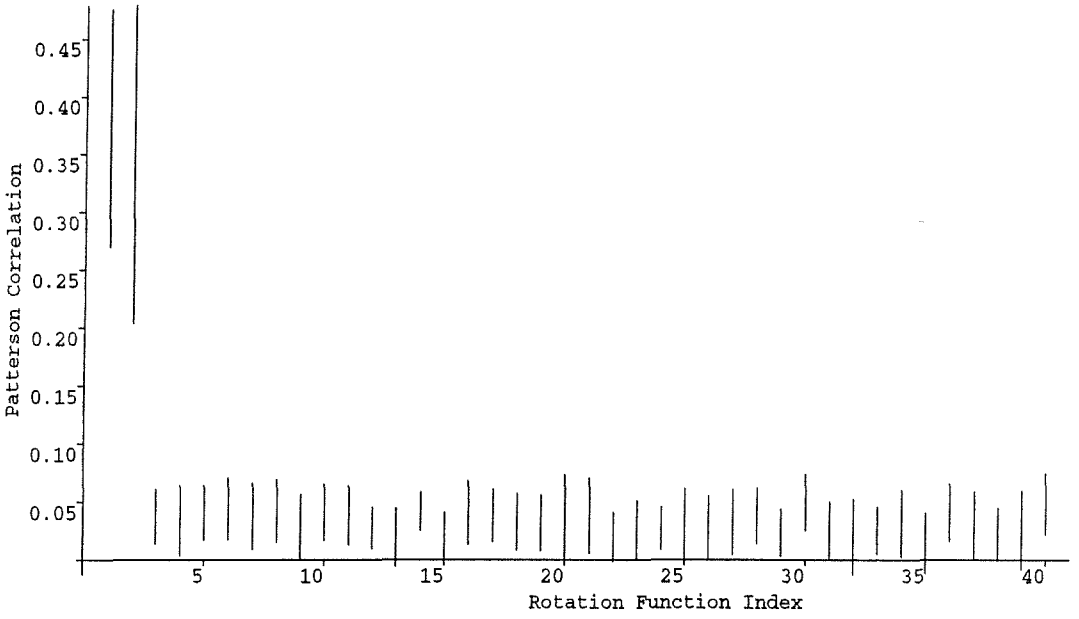


Figure 3.7: The results of the Patterson correlation refinement performed on peaks obtained from the cross rotation function in X-PLOR. There are two significant peaks related by symmetry, corresponding to the orientation (ϕ_1=157.06°, ϕ_2=157.85°, ϕ_3=73.34° for peak1 and ϕ_1= 220.37°, ϕ_2=16.62°, ϕ_3=239.32° for peak 2). The conversion of peak 2 from X-PLOR Eulerian angles to CCP4 Eulerian angles corresponds to α=210.8°, β=16.62°, γ=49.63° which is consistent with the orientation calculated using AMoRe (α=205.82°, β=13.00°, γ=54.94°).

3.5 Refinement of the Bovine IMPase Structure

The crystal packing of the molecular replacement solution was viewed using MOLPACK and seen to be feasible. The refinement of the molecular model was performed with CNS, using 53444 reflections within a resolution range of 20 to 1.65 Å. A test set of reflections (5 % of data, 2703 reflections) was used for R_{free} calculations. NCS restraints were maintained between the two monomers during the early stages of refinement. The initial round of refinement consisted of rigid body refinement, simulated annealing (slow cool from 2500 to 300 K in 25 K decrements) and grouped B-factor refinement. This reduced the R-factor from 44.0 % (R_{free}=44.6 %) to 31.2 % (R_{free}=33.6 %). The resulting model was used to calculate sigmaA-weighted electron density maps that were subjected to two-fold NCS averaging.

Model visualisation and rebuilding was performed using the program QUANTA. Examination of the averaged maps revealed *Fo-Fc* density for bovine IMPase side chains differing from their equivalents in the human IMPase search model (Figure 3.8). This confirmed that the calculated maps were not model biased and provided further evidence for the choice of a correct molecular replacement solution. The maps were of excellent quality allowing most of the bovine IMPase residues to be manually fitted to the *Fo-Fc* density during the initial round of model building. The model was subsequently subjected to several alternating cycles of refinement and model building. Each cycle consisted of positional refinement and individual isotropic B-factor refinement, during which the NCS restraints were released.

Once the R-factor (and the R_{free}) had dropped below 30 %, solvent molecules were placed at stereochemically acceptable sites following visual verification of the *2Fo-Fc* map. Solvent atoms were only accepted if they appeared in *Fo-Fc* maps contoured at 3 σ, reappeared in subsequent *2Fo-Fc* maps, formed hydrogen bonds with chemically reasonable groups and had B-factors less than 60 $Å^2$. These water molecules were added in successive steps and were included in the subsequent refinement cycles. The final model has an R-factor of 18.7 % (R_{free}= 20.7 %) and consists of 4068 protein atoms, seven Mg^{2+} ions and 350 solvent molecules. Inspection of the Ramachandran plot using PROCHECK (Laskowski *et al.*, 1993) demonstrated acceptable geometry for all main chain parameters with no outliers. The refinement statistics are listed in Table 3.4.

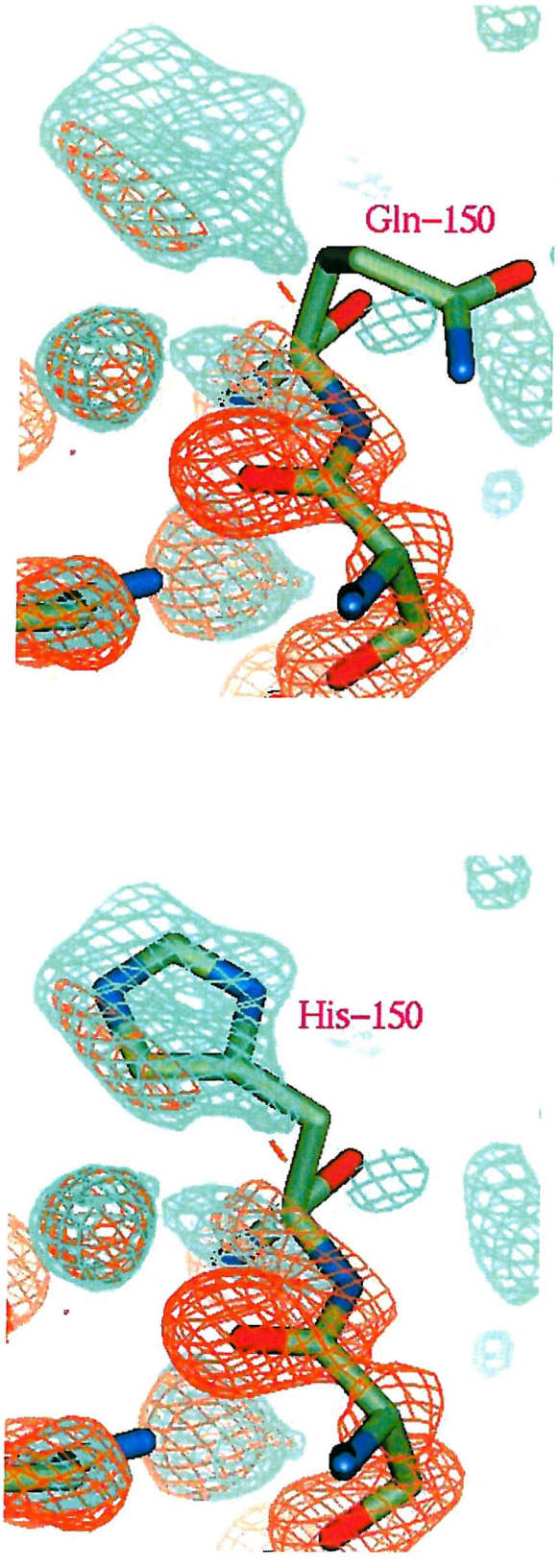


Figure 3.8: Confirmation of a correct molecular replacement solution for bovine IMPase. Features that are present in the bovine IMPase crystal but absent from the human IMPase model are displayed (QUANTA). In the bovine structure Fo-Fc density for the histidine side chain can be observed at position 150 (glutamine in the human enzyme). The Fo-Fc map (depicted in turquoise) is contoured at the 2 σ level, whereas the 2Fo-Fc map (depicted in orange) is contoured at the 1 σ level.

Resolution range (Å)	20-1.65
R-factor (%)	18.9
R-free (%)	20.7
Number of reflections in working set	53444
Number of reflections in test set	2703
Number of protein atoms	4068
Number of solvent atoms	350
Rms bond length deviation (Å)	0.0048
Rms bond angle deviation (°)	1.35
Mean B-factor for main chain : Subunit A	11.0
Mean B-factor for main chain : Subunit B	13.6
Mean B-factor for side chain : Subunit A	13.6
Mean B-factor for side chain : Subunit B	15.5
Mg^{2+} B-factors ($Å^2$) Mg-1,Mg-2,Mg-3a : Subunit A	9,20,19
Mg^{2+} B-factors ($Å^2$) Mg-1,Mg-2,Mg-3b : Subunit B	9,13,22

Table 3.4: Refinement statistics for bovine IMPase structure.

3.6 The Structure of Bovine IMPase

3.6.1 Overall Structure

As expected, the overall fold of bovine IMPase is very similar to human IMPase, with each subunit composed of 9 α-helices and 13 β-strands that make up a penta-layered αβαβα sandwich (Figure 3.9). The first layer comprises of two anti-parallel six and five-turn helices (α1, Trp-5—Ala-26 and α2, Ala-44—Tyr-62) connected by a β-hairpin (β1, Asn-32—Lys-36 and β2, Asp-41—Val-43). The second layer is a large six stranded anti-parallel β sheet (β3, Ser-66—Glu-70, β4, Pro-85—Asp-90, β5, Val-105—Val-113, β6, Lys-116—Ser-124, β7, Asp-128—Lys-135 and β8, Lys-137—Asn-142) orientated parallel with the first helical layer. Also present in the second layer is a two-turn helical segment following strand β3 (α3, Glu-70—Ala-75) and an unusual kink structure (Asp-90—Thr-95) that extends into another two-turn helix (α4, Thr-95—Gly-101).

The third layer is composed of two three-turn parallel helices (α7, Thr-195—Ala-205 and α8, Cys-218—Glu-230) that are aligned perpendicular to the second layered β sheet. The fourth layer is made up of mostly parallel five-stranded β sheet (β10, His-188—Arg-191, β9, Ser-157—Thr-161, β11, Asp-209—Gly-215, β12, Val-234—Leu-236 are parallel, while β13, Arg-248—Ser-254, positioned between strands β11 and β12, is orientated anti-parallel). Also present in this layer is a one-turn helical segment that precedes strand β7 (α5, Asn-153—Ser-157). Finally, two anti-parallel three and four-turn helices (α9, Asn-255—Ileu-266 and α6, Thr-168—Cys-184) that pack against each other and the fourth layered β sheet make up the fifth layer.

3.6.2 Conformational Differences Between Subunits A and B of Bovine IMPase

A superposition of the C-alpha atoms of subunits A and B of the bovine IMPase structure yields an rms deviation of 0.27 Å (Figure 3.10a). With the exception of the *N* and *C* termini, the molecular region showing the largest structural difference in the two molecules is the *C*-terminal end of helix α9 and the following loop region (residues 263-270) (Figure 3.11a). This difference can be ascribed to crystal contacts present in subunit A, mediated by a Mg^{2+} ion co-ordinated to six water molecules (1.9 to 2.2 Å). Five of the water ligands

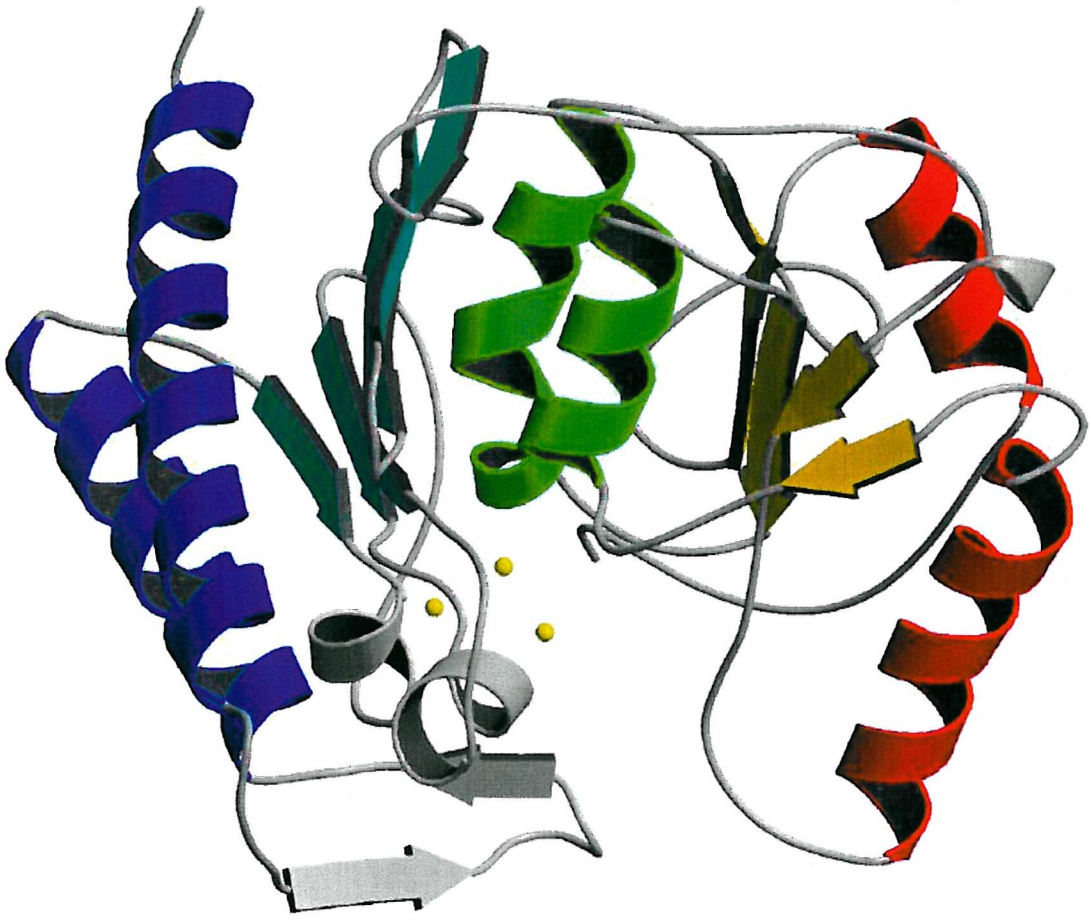


Figure 3.9: The structure of Subunit A of bovine inositol monophosphatase (MOLSCRIPT and rendered using Raster3D). The overall fold in each subunit is composed of 9 α-helices and 13 β-strands that make up a penta-layered αβαβα sandwich (each type of the secondary structural elements are coloured differently). The magnesium ions located in the active site are depicted in yellow.

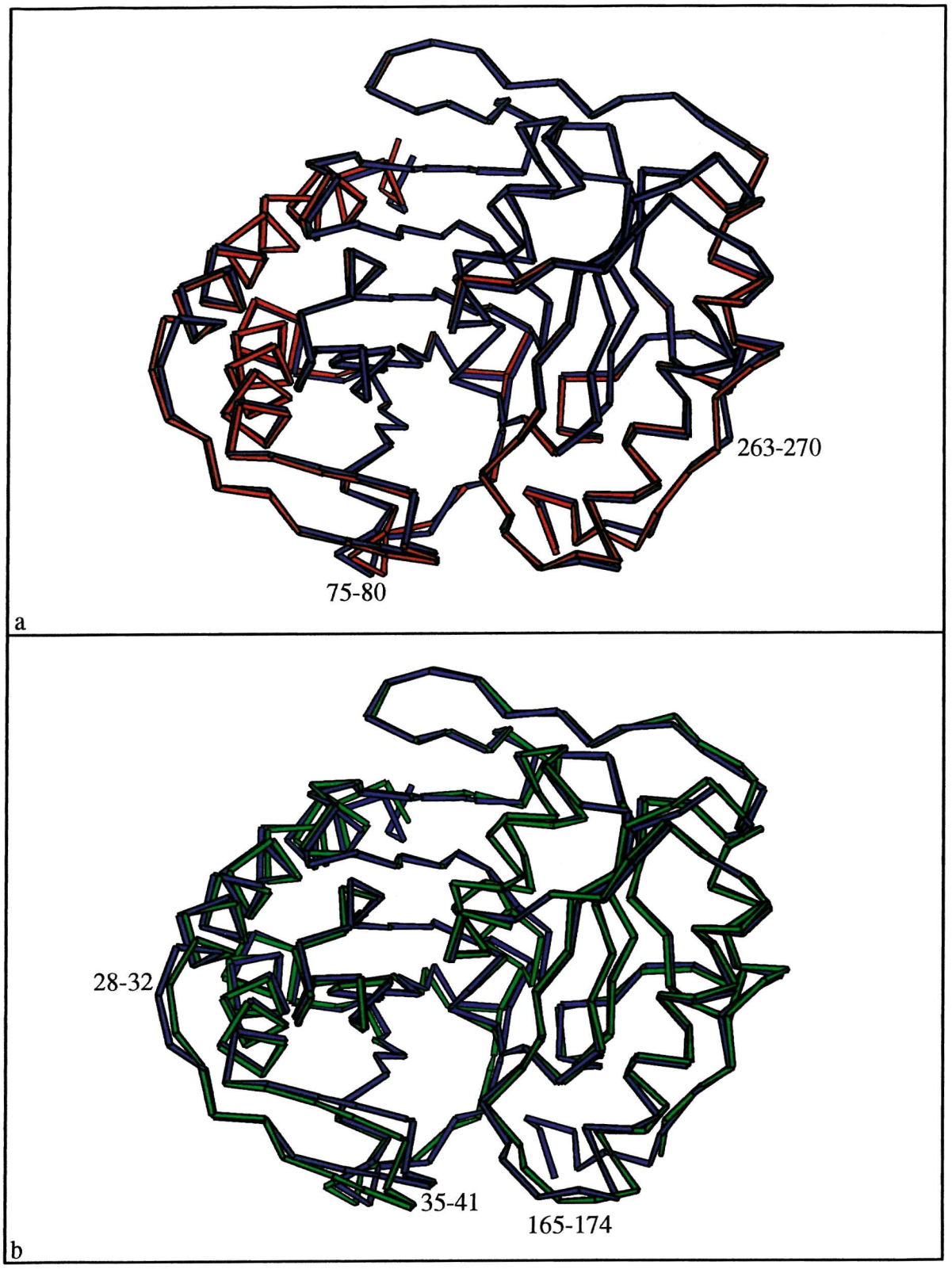


Figure 3.10a) Superposition of subunits A (blue) and B (red) of bovine IMPase. The rms deviation is 0.28 Å for 274 equivalent C-alpha atoms that superpose within 3.5 Å. b) Superposition of subunits A of bovine IMPase (blue) and human IMPase complexed with manganese and phosphate (green). The rms deviation is 0.40 Å for 272 equivalent C-alpha atoms that superpose within 3.5 Å. The molecular regions displaying the largest structural differences have been labelled in both diagrams (MOLSCRIPT).

form multiple hydrogen-bonding interactions to residues 263-270, while the sixth water is from a crystallographic symmetry related molecule. In contrast, the corresponding region in subunit B forms fewer crystal contacts. Further regions with large conformational differences (>1.0 Å in the least squares superposition) are restricted to surface residues including Lys-28 (loop residue following helix α1), Glu-60 (helix α2), Ala-75 to Ile-80 (loop region following helix α3) and Glu-170 (helix α6) (Figure 3.11a).

The average main chain temperature factors (B-factors) for subunits A and B are 11 $Å^2$ and 13.6 $Å^2$, respectively. There are three regions with significantly higher than the mean temperature factor in subunit A and five in subunit B (Figure 3.11b). Two of the regions are at the *N* and *C* terminal end of each subunit, reflecting the mobility of the protein tails. The β-hairpin connecting helices α1 and α2 in subunit B (residues 34-42) is also very mobile, and is associated with poor electron density. The equivalent region in subunit A is less disordered because it forms hydrophobic crystal contacts with a neighbouring molecule.

Additional regions with above average temperature factors include two solvent exposed loop regions (residues 75-80 and 238-241) and the *C*-terminal end of helix α9 (residues 263-270). The loop encompassing residues 75-80 is highly mobile in both subunits displaying similar temperature factor trends. In contrast, region 238-241 (loop connecting strands β12 and β13) is more disordered in subunit B compared to subunit A. The higher temperature factors observed for subunit B can be rationalised by the extensive crystal contacts made by the equivalent region in subunit A (hydrophobic interactions and multiple hydrogen-bonding interactions between Thr-238 and a neighbouring molecule). Finally, the extra mobility of helix α9 and the ensuing loop region in subunit B can also be attributed to crystal contacts (via Mg^{2+} and solvent molecules) present in the equivalent region of subunit A.

3.6.3 Conformational Differences Between Bovine and Human IMPase

The results of the superpositions of the bovine structure and the human structure complexed with various metal ions and ligands are listed in Table 3.5. Since most of the human structures are similar, only the dimer complex with Mn^{2+} and PO_4 (accession code

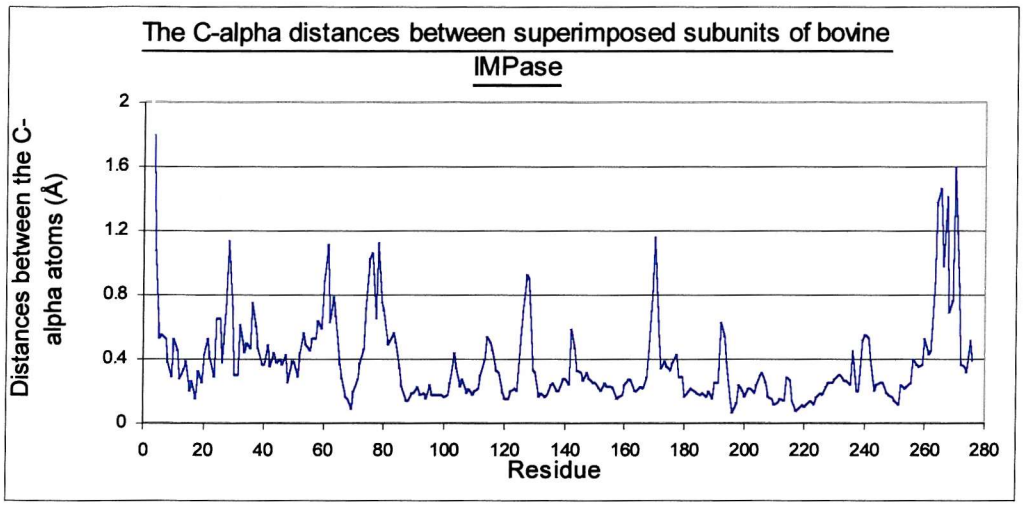


Figure 3.11a: The C-alpha distances between superimposed subunits A and B of bovine IMPase (y-axis) as a function of residue (x-axis). There are 4 molecular regions with deviations >1.0Å in the least squares superposition (residues 75-80, 263-270 and the protein tails).

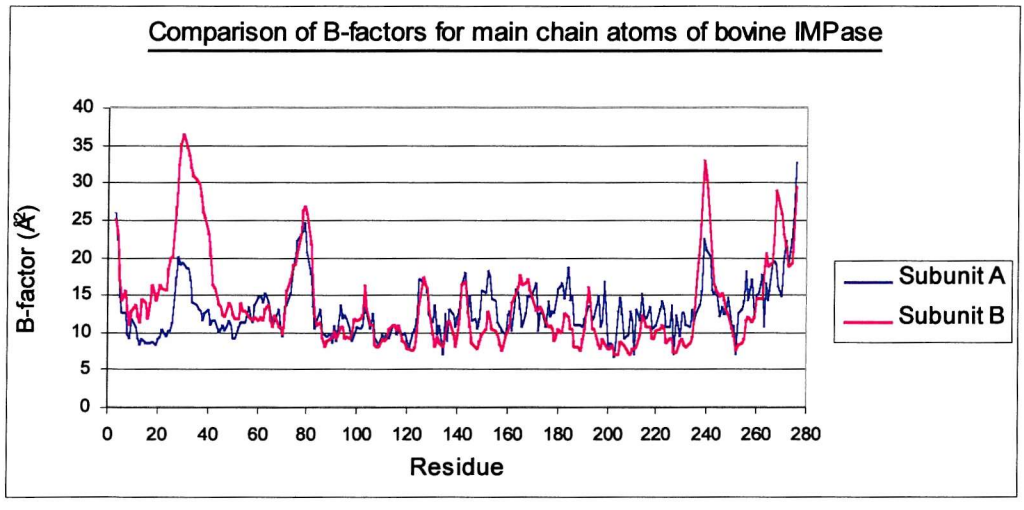


Figure 3.11b: The average isotropic B-factors per residue for the refined bovine IMPase structure (y-axis) plotted against residue number (x-axis). With the exception of the N and C termini, there are 2 regions with above average B-factors in subunit A (residues 75-80 and 238-241) and 3 in subunit B (residues 34-42, 74-80 and 238-241). Intermolecular contacts between crystallographic symmetry related molecules account for the majority of differences in the average B-factors of the two molecules.

1IMD; Bone *et al.*, 1994b) will be used in comparisons with the bovine enzyme. When superimposing subunits A of the bovine and human enzyme (Figure 3.10b), significant differences in the atomic positions are observed in three regions; 28-32 and 35-41 (both within the β hairpin connecting helices α1 and α2) and 166-174 (loop following β9 and the *N*-terminal end of helix α6) (Figure 3.12a). With the exception of region 35-41, crystal contacts present in bovine IMPase are largely responsible for these differences. The conformational difference in region 35-41 is mainly due to the indirect interaction of Asp-41 to metal binding site 3a (via Wat-163) in the bovine enzyme, resulting in this region being pulled towards the active site.

Deviations between C-alpha atoms for the bovine and human enzymes in subunit B are localised to four regions; 24-30 (the *C*-terminal end of helix α1 and the proceeding loop region), 36-41 (as in subunit A), 165-174 (as in subunit A) and 264-270 at the *C*-terminus of the monomer. In both structures region 24-30 is not involved in crystal contacts but the side chains are on the protein surface, completely exposed to the solvent. There are multiple crystal contacts observed in region 36-41 for the bovine monomer, but the corresponding contacts for the human monomer are limited to residues 34 and 35. The mobility of region 165-174 is reduced in the bovine enzyme compared to the human enzyme (as in subunit A), and this is due to presence of crystal contacts. Finally, the reason for the differences at the *C*-terminal end of subunit B 264-270 is not clear since a similar number of crystal contacts are observed in both subunits.

Comparisons of temperature factors for subunits A of the bovine and human enzyme are shown in Figure 3.12b. In general, the bovine enzyme displays lower temperature factors with respect to the human enzyme owing to resolution differences; bovine IMPase was solved to a resolution of 1.65 Å, whereas human IMPase complexed with Mn^{2+} and PO_4 was solved to a resolution of 2.6 Å. There are several regions with significant differences in which the human enzyme exhibits higher B-factors (for regions 57-67, 113-116 and 164-174; in both subunits), and these correlate with crystal contacts lowering the mobility of the corresponding regions in the bovine enzyme. Furthermore, regions 28-40 and 74-80 in subunit B of both enzymes are highly mobile, whereas the loop encompassing residues 238-242 is more mobile in subunit B of the bovine enzyme.

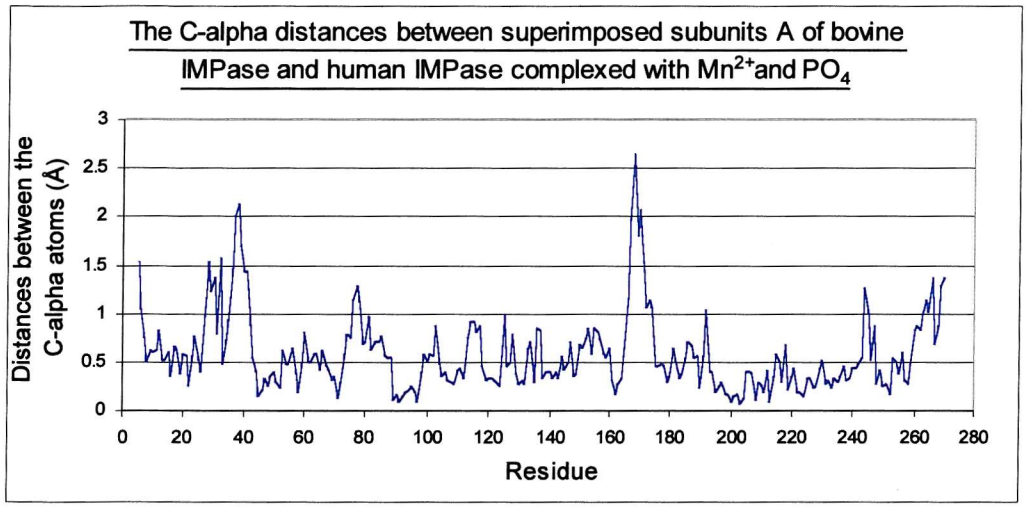


Figure 3.12a: The C-alpha distances between superimposed subunits A of bovine IMPase and human IMPase complexed with Mn^{2+} and PO_4 (y-axis) as a function of residue number (x-axis). Large deviations are localised to 3 regions (residues 28-32, 35-41 and 165-174). The differences in region 35-41 are due to Asp-41 interacting indirectly with Mg^{2+} at site 3 in the bovine structure resulting in the region being pulled towards the active site. Differences in region 165-174 are ascribed to crystal contacts present in the bovine enzyme but absent in the corresponding region of the human enzyme.

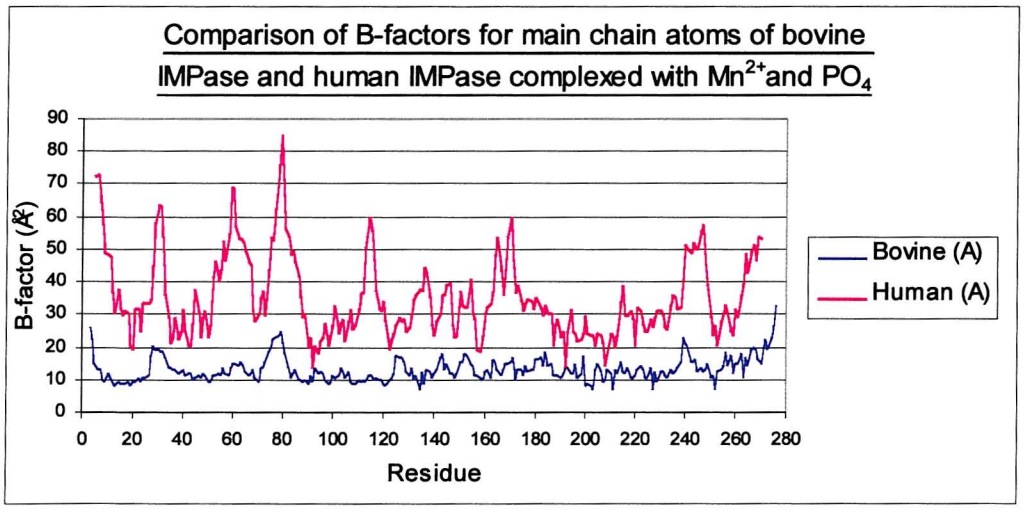


Figure 3.12b: The average isotropic B-factors per residue for the refined structures of bovine IMPase and human IMPase complexed with Mn^{2+} and PO_4 (y-axis) are plotted against the residue number (x-axis). In general, the human IMPase structure exhibits higher B-factors compared to the bovine IMPase structure due to the large difference in resolution. Taking this into account, there are still several regions of the human enzyme that exhibit high B-factors (residues 57-67, 113-116 and 164-174), which can be accounted for by crystal contacts between crystallographic symmetry related molecules in the corresponding regions of the bovine enzyme.

Human IMPase	RMS Deviation (Å)	
	Subunit A	Subunit B
Complex with Gd^{3+} and sulphate	0.34	0.36
Complex with Gd^{3+} and D-Ins(1)P	0.36	0.44
Complex with Gd^{3+} and L-Ins(1)P	0.39	0.46
Complex with Ca^{2+}	0.35	0.43
Apoenzyme	0.53	0.57
Complex with Mn^{2+} and chloride	0.37	0.46
Complex with Mn^{2+} and phosphate	0.40	0.48
Complex with Ca^{2+} and D-Ins(1)P	0.37	0.44

Table 3.5: Summary of the superposition results of bovine IMPase and human IMPase complexed with various metal ions.

3.6.4 Dimer Interface of Bovine IMPase

The dimer interface of bovine IMPase is extensive (2296 $Å^2$), with regions throughout both chains making contributions. Since the dimer is formed via an approximate (NCS) 2-fold rotation, all contacts in the interface are duplicated. The interface has a more hydrophobic character than polar, with principal dimerisation contacts involving several hydrophobic clusters. In the largest cluster, residues involved in these interactions (3.8-4.5 Å) include Phe-102, Pro-103, Phe-104, Leu-158, Val-193 and the aliphatic moieties of His-100 and Arg-191 (Figure 3.13). The hydrophobic interactions between the bovine subunits differ from those in the human dimer in two respects. Firstly, the presence of smaller residues (Ala-40 and Gly-192) in the bovine structure (equivalent to Val-40 and Ser-192 in the human enzyme) result in fewer contacts made at the interface. Secondly, a compensatory hydrophobic exchange involving Phe-183 of human IMPase and Leu-183 in the bovine enzyme has taken place.

The dimer interface of bovine IMPase is further stabilised by two salt bridges (between His-100 and Asp-209 and Arg-173 and Glu-180), five hydrogen bonds and a network of water molecules that make contact with both subunits (Figure 3.13). The five hydrogen bonds have an average distance below 3.3 Å. The side chain oxygen atom of Asn-97 forms hydrogen-bonding interactions with the NE and NH2 atoms of Arg-191. Also, the NE and NH2 atoms of Arg-167 form hydrogen bonds with the carbonyl groups of Ile-187 and His-

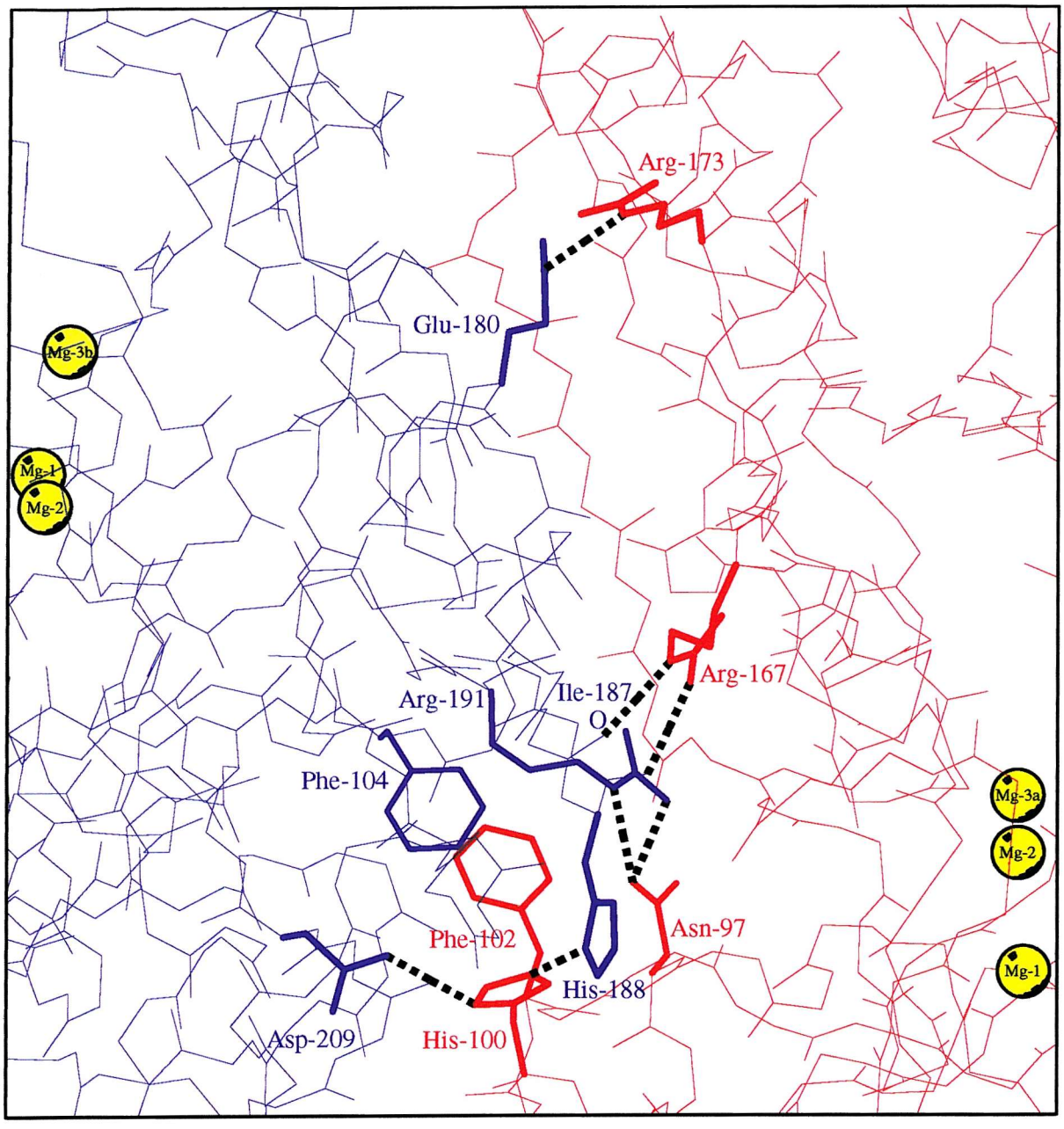


Figure 3.13: The dimer interface of bovine IMPase (MOLSCRIPT). Subunit A is depicted in red and subunit B in blue. Residues involved in stabilising the dimer are displayed in sticks. The dimer is stabilised by 5 hydrogen bonds (Arg-167-NE—Ile-187-O, Arg-167-NH2—His-188-O, Arg-191-NE—Asn-97-OD1, Arg-191-NH2—Asn-97-OD1 and His-100-NE2—His-188-ND1), two salt bridges (Asp-209-OD2—His-100-ND1 and Glu-180-OE2—Arg-173-NE) and several hydrophobic interactions (Phe-102, Phe-104 and the aliphatic moieties of Arg-191 and His-188). Note that all interactions are duplicated because of the 2-fold non-crystallographic symmetry (giving a total of 10 hydrogen bonds and 4 salt bridges). The Mg^{2+} binding sites for each subunit are also depicted (shown in yellow and labelled as Mg-1, Mg-2, Mg-3a (for subunit A) or Mg-3b (for subunit B).

188. The latter also forms a hydrogen bond with His-100. By contrast, as well as the above interactions, the human interface is stabilised by three additional hydrogen bonds. Firstly, the NE atom of Arg-167 is within hydrogen bonding distance (3.11 Å) to the carbonyl group of Phe-183 (3.8 Å in bovine). Secondly, both carboxyl oxygen atoms of Glu-180 interact with Arg-193 rather than one. Finally, the carbonyl group of Ser-192 forms a hydrogen bond with its equivalent; the equivalent residue in bovine IMPase (Gly-192) does not participate in dimerisation contacts.

3.6.5 Structural Comparisons with Human IMPase

The bovine enzyme shares 89 % amino acid sequence identity with the human enzyme, with substitutions occurring in regions not involved in the active site structure. There are several compensatory residue substitutions present in the bovine structure that maintain core hydrophobic contacts (such as Ile-27, Val-33, Ile-117, Val-121, Val-126 and Val-175 of human IMPase, which are substituted by Leu-27, Ile-33, Met-117, Ile-121, Leu-126 and Ile-175, respectively, in bovine IMPase). Analysis of the crystal structure of human IMPase has shown that there is a disulphide bridge (Cys-24-Cys-125), which stabilises helix α1 and the loop following strand β7 (Bone *et al.*, 1992). In the bovine enzyme Cys-24 is replaced by Arg-24, which forms a hydrogen bond between the guanidinium group and the main-chain carbonyl group of Cys-125. From a structural point of view the Arg-24—Cys-125 interaction may compensate, in part, for the absence of the disulphide bridge.

There are several hydrogen bonds in the bovine structure that are absent in the human structure, owing to residue substitutions. For instance, Arg-24 forms an intramolecular salt bridge with Glu-21 (Arg-24 NE2—Glu-21 OE2 3.18 Å) as well as forming a compensatory hydrogen bonding interaction with Cys-125. Furthermore, Ser-253 in the bovine enzyme (equivalent to Ala-253 in human) stabilises the protein structure by forming multiple hydrogen-bonding interactions with the main-chain carbonyl and amide groups of Asn-255 (Ser-253 OG—Asn-255 N 3.19 Å; Ser-253 OG—Asn-255 O 2.82 Å). Also, the side chain oxygen atom of Thr-257 (equivalent to Ile-253 in human) forms a hydrogen bond with Glu-152 (Thr-257 OG1—Glu-152 OE2 3.15 Å).

3.6.6 The Active Site of Bovine IMPase

The active site of bovine IMPase is located at the intersection of several secondary structure elements (α2, β3, α8 and the unusual 'kinked structure'). Each subunit possesses three Mg^{2+} binding sites in a hydrophilic cavity near the surface of the molecule. Several ordered water molecules in the active site serve as metal ligands and perform a structural role by forming hydrogen bonds to residues that make up the active site. The metal binding sites were identified on the basis of several criteria. Firstly, *Fo-Fc* maps calculated at 1.65 Å displayed clear peaks ranging from 5 to 10 σ (metal ions from the human model had been omitted prior to molecular replacement calculations). Secondly, the co-ordination sphere of each metal site consisted of water molecules and carboxylate groups (listed in Table 3.6) all within the accepted range for magnesium-oxygen distances (1.9 to 2.3 Å). Thirdly, the co-ordination sphere of the majority of metal sites refined to octahedral geometry, consistent with the expected co-ordination for Mg^{2+} (Huheey, 1978). Finally, all the Mg^{2+} ions had low temperature factors (Table 3.4).

In both subunits the co-ordination sphere of Mg^{2+} at site 1 (Mg-1) consists of several conserved residues (Glu-70, Asp-90 and Ile-92). The ligand sphere of Mg-1 is completed with co-ordination to three water molecules (Wat 8, 13 and 90 in subunit A and Wat 5, 50 and 351 in subunit B), forming an octahedral co-ordination site (Figures 3.14 and 3.15). Based on previous structural studies of the human enzyme, Wat-13 (Wat-50 in subunit B) is the most likely candidate for the role of the activating water nucleophile in the bovine structure, since it forms hydrogen bonds with Thr-95 and Glu-70. Mg-1 can be superposed onto the site 1 binding sites of the human structures complexed with Gd^{3+} (Bone *et al.*, 1992), Mn^{2+} (Bone *et al.*, 1994b) and Ca^{2+} (Ganzhorn and Rondeau, 1997).

The co-ordination geometry of Mg^{2+} at site 2 (Mg-2) differs in the two subunits; as well as forming ligands to the carboxylate oxygens of Asp-90, Asp-93 and Asp-220 in each subunit, metal 2 is co-ordinated to two water molecules in subunit A (trigonal bipyramid co-ordination) and three in subunit B (octahedral co-ordination). Despite an agreement with the co-ordination geometry of Mg-2, the Mg^{2+}-water distances for Wat-96 (2.53 Å) and Wat-322 (2.42 Å) are significantly longer than expected for a direct co-ordination bond (Harding, 1999). This site is equivalent to metal binding site 2 of the human IMPase structures complexed with Mn^{2+} and Ca^{2+}. However, the co-ordination geometry of Mg-2

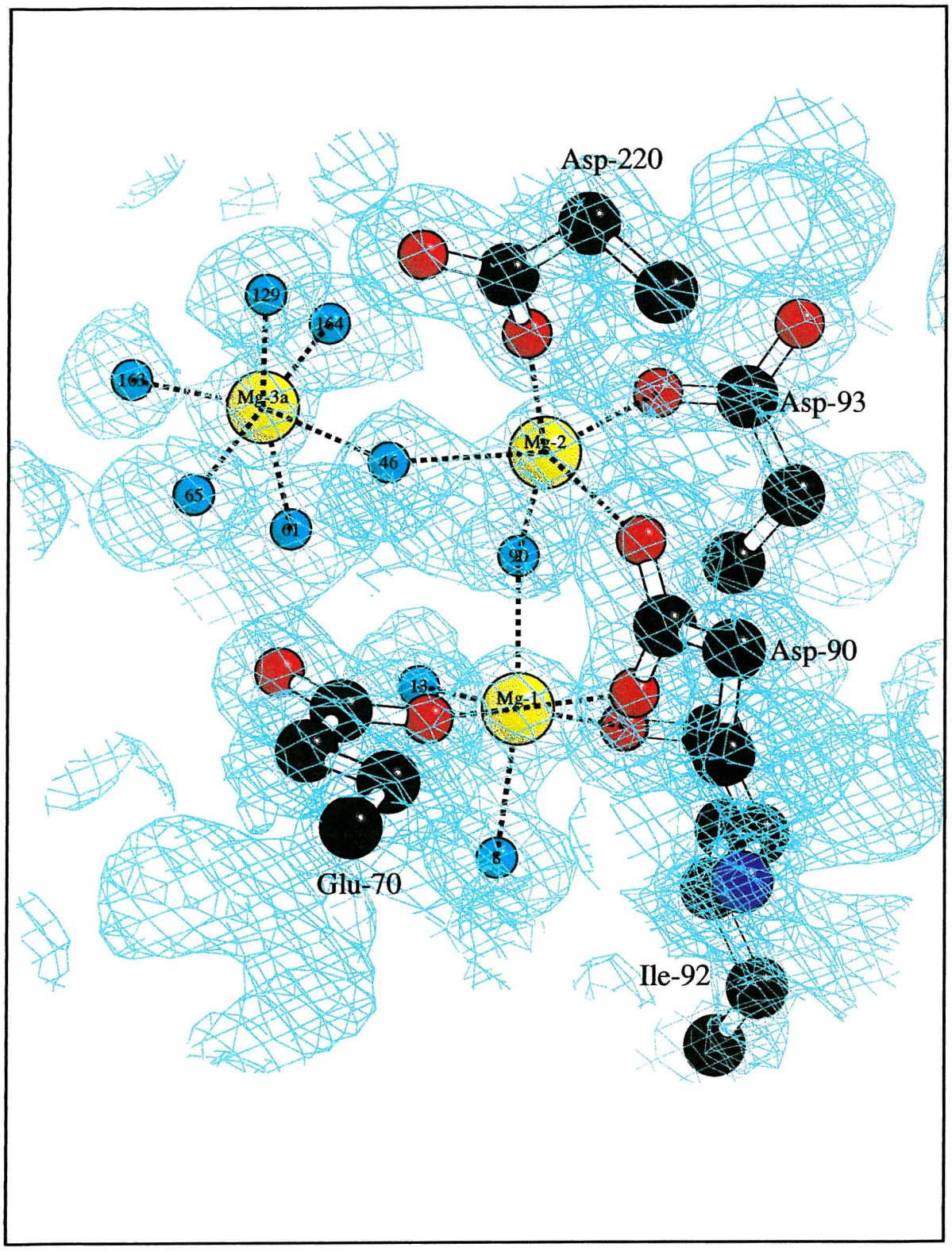


Figure 3.14: The active site interactions of subunit A of bovine IMPase (BOBSCRIPT). Three Mg^{2+} ions (depicted in yellow and labelled as Mg-1, Mg-2 and Mg-3a) form ligands to several conserved residues (carboxylates of Glu-70, Asp-90, Asp-93, Asp-220 and the carbonyl group of Ile-92) and water molecules (depicted in blue). The co-ordination geometry for Mg-1 and Mg-3a is octahedral, whereas for Mg-2 it is trigonal bipyramid. The putative activating water molecule (Wat-13) forms interactions with side chains of Glu-70 and Thr-95 and Mg-1. The 2Fo-Fc map is contoured at the 1 σ level (sky-blue).

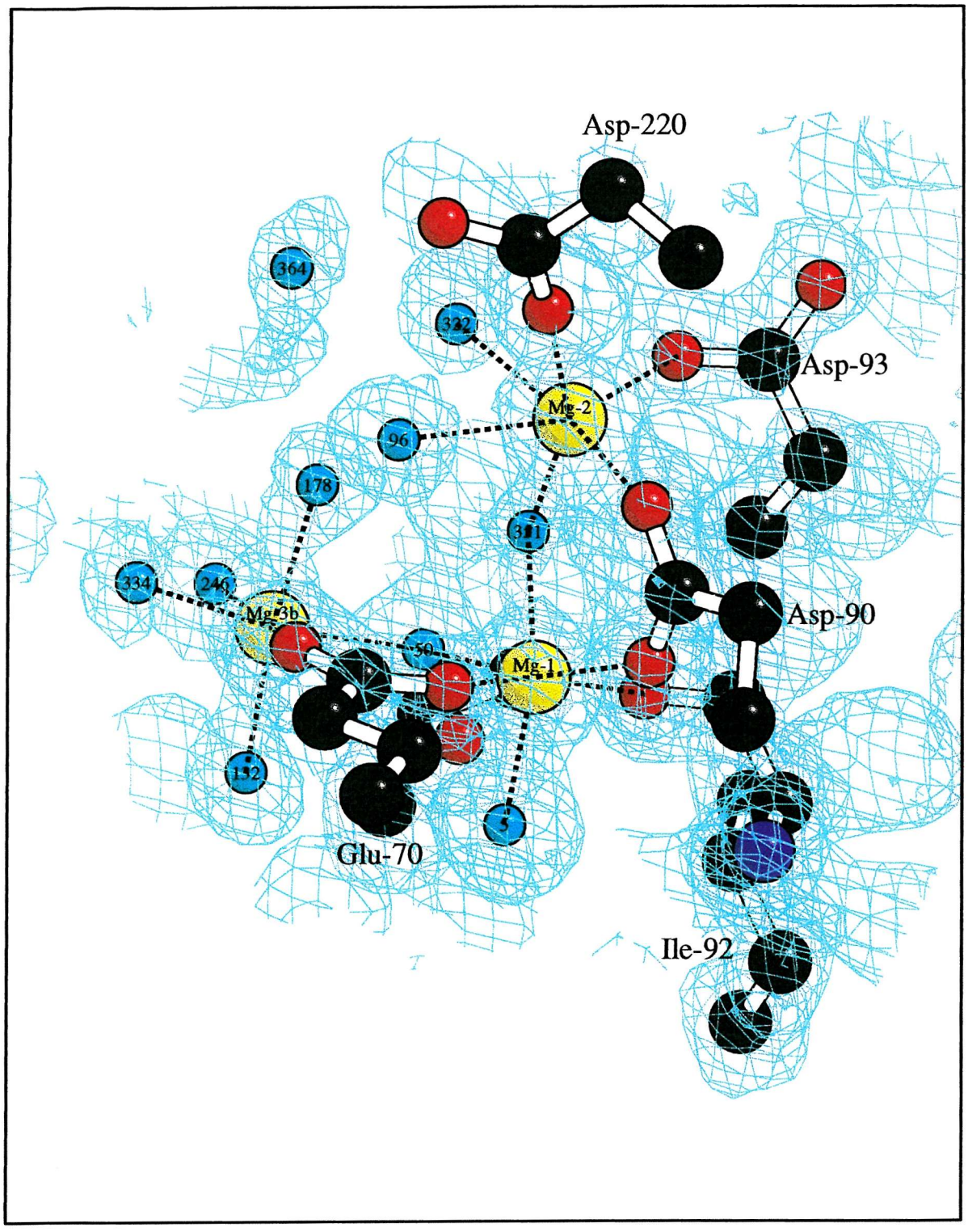


Figure 3.15: The active site interactions of subunit B of bovine IMPase (BOBSCRIPT). The interactions between the Mg^{2+} ions (depicted in yellow and labelled as Mg-1, Mg-2 and Mg-3b), active site residues and water molecules (depicted in blue) are similar to those in subunit A. Mg-3b is located at a different position in comparison to subunit A, forming a ligand with the carboxylate oxygen atom of Glu-70. All 3 Mg^{2+} binding sites display octahedral co-ordination geometry. The putative activating water molecule (Wat-50) forms interactions with Glu-70, Thr-95, Mg-1 and Mg-3b. The 2Fo-Fc map is contoured at the 1 σ level (sky-blue).

in subunit B of the bovine structure differs from any of the previous human structures, being octahedral rather than trigonal bipyramid. The distance between Mg-1 and Mg-2 is 3.6 Å (same in subunit B), with Asp-90 and Wat-90 (Wat-351 in subunit B) co-ordinating to both sites (Figure 3.15).

The Mg^{2+} ion occupying site 3 in subunit A (Mg-3a) is co-ordinated to six water molecules and is absent in the corresponding region of subunit B. The third Mg^{2+} binding site in subunit B (Mg-3b) is equivalent to Mn-3 and Ca-3 of human IMPase although its position has moved by ~ 1.3 Å. The co-ordination sphere of Mg-3b consists of one protein ligand (Glu-70) and five water molecules. There is spherical *Fo-Fc* density (contoured at the two σ level) in the proximity of Glu-70 (2.0 Å) in subunit A, indicating a possible Mg^{2+} binding site; the putative Mg^{2+} binding site is located in an equivalent position to Mg-3b in subunit B. However, following the refinement of Mg^{2+} ion at this site, examination of the subsequent *2Fo-Fc* maps revealed little density, suggesting that the Mg^{2+} ion is most likely present at low occupancy.

Mg^{2+} Sites	Subunit A		Subunit B	
Mg-1	Glu70-OE1	(2.07)	Glu70-OE1	(2.04)
	Asp90-OD1	(2.11)	Asp90-OD1	(2.08)
	Ile92-O	(2.11)	Ile92-O	(2.14)
	Wat-8	(2.09)	Wat-5	(2.05)
	Wat-13*	(2.09)	Wat-50 *	(2.15)
	Wat-90	(2.14)	Wat-351	(1.99)
Mg-2	Asp90-OD2	(2.16)	Asp90-OD2	(2.03)
	Asp93-OD1	(2.09)	Asp93-OD1	(1.99)
	Asp220-OD1	(2.13)	Asp220-OD1	(2.00)
	Wat-46	(2.29)	Wat-96	(2.53)
	Wat-90	(1.79)	Wat-351	(1.87)
			Wat-322	(2.42)
Mg-3a and 3b	Wat-46	(2.29)	Glu70-OE2	(2.15)
	Wat-61	(2.12)	Wat-178	(2.04)
	Wat-65	(2.15)	Wat-132	(2.03)
	Wat-129	(2.02)	Wat-50*	(2.15)
	Wat-163	(2.00)	Wat-246	(2.25)
	Wat-164	(2.06)	Wat-334	(2.08)

Table 3.6: The co-ordination sphere of the Mg^{2+} ions in bovine IMPase (distances in Å).

3.6.7 Modelling Studies of D-Ins(1)P

The bovine IMPase structure is similar to that of human IMPase complexed with Ca^{2+} and D-Ins(1)P, with an rms deviation of 0.44 Å (between superimposed C-alpha atoms of subunit B). By superposing subunit B of both structures, possible interactions between the substrate and the metal ions were determined (Figure 3.16). This analysis indicates the activating water nucleophile (Wat-50) is retained in the presence of substrate following a slight adjustment in its position. The non-bridging phosphate oxygen of substrate (O7) would displace Wat-351 and co-ordinate to Mg-1 (O7—Mg-1 = 2.35 Å) and Mg-2 (O7—Mg-2 = 2.06 Å). Several ligands are also likely to be retained at site 1 (Glu-70, Asp-90, Ile-92 and Wat-5), generating distorted octahedral co-ordination geometry.

The ester oxygen of substrate (O1) would also displace one of the metal-bound water molecules (Wat-322) at Mg-2 (O1—Mg-2 = 2.30 Å). As a result, the substrate ligands (O1 and O7) together with Asp-90, Asp-93, Asp-220 and Wat-96 would generate an octahedral co-ordination binding site. Interestingly, Mg-3b would not be displaced by the phosphate moiety of the substrate, as demonstrated previously in the crystal structure of human IMPase-Mn^{2+}-PO_4 complex (Bone *et al.*, 1994b); several ligands would be retained (Glu-70, Wat-50, Wat-246, Wat-132 and Wat-334), whereas Wat-178 would most likely be displaced by O8 of the phosphate group (O8—Mg-3b = 2.76 Å) to form an octahedral co-ordination binding site.

3.6.8 Conclusions

The structure of bovine IMPase has been solved with its natural metal ligand, Mg^{2+}, under inhibitory conditions (in the presence of 20 mM $MgCl_2$). Previous human IMPase structures have been determined in the presence of Gd^{3+} (Bone *et al.*, 1992), Mn^{2+} (Bone *et al.*, 1994b) and Ca^{2+} (Ganzhorn and Rondeau, 1997). The active site of each bovine IMPase subunit possesses three Mg^{2+} binding sites, although the location of the third metal binding site differs in each subunit. In subunit B, Mg-3b forms a solitary protein interaction with Glu-70, as observed in the human structures complexed with Mn^{2+} (Bone *et al.*, 1994b) and Ca^{2+} (Ganzhorn and Rondeau, 1997). By contrast, Mg-3a in subunit A is located in a different position within the active site, forming ligands to six water molecules.

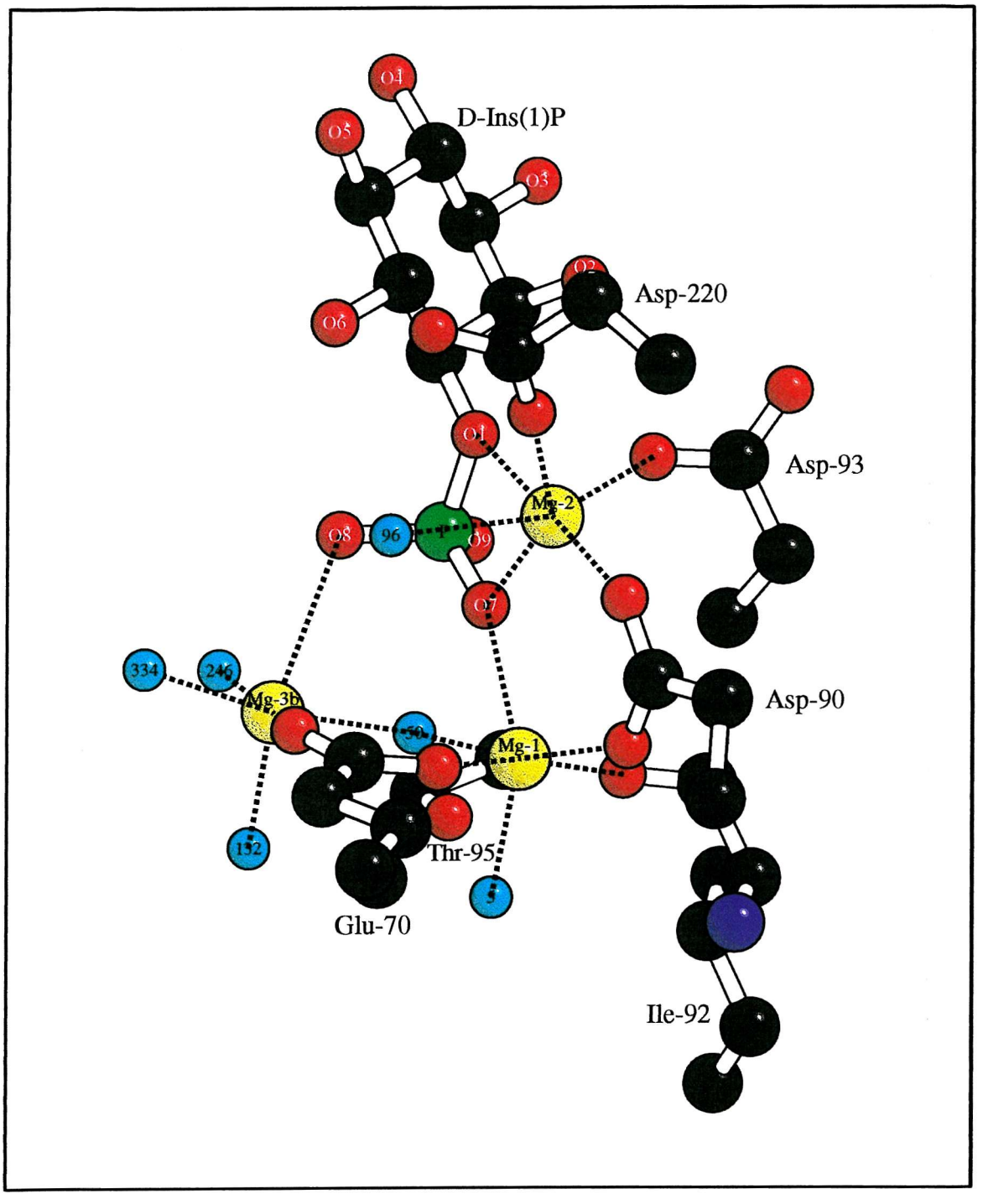


Figure 3.16: Possible interactions between D-myo-inositol-1-phosphate (D-Ins(1)P) and Mg^{2+} ions in subunit B of bovine IMPase (BOBSCRIPT). The substrate phosphoryl group would be able to bind without displacing the proposed activating water nucleophile (Wat-50). The non-bridging phosphate oxygen of substrate (O7) would displace a metal bound water molecule (Wat-351) and co-ordinate to Mg-1 (O7—Mg-1 = 2.35 Å) and Mg-2 (O7—Mg-2 = 2.06 Å). The ester oxygen (O1) of the substrate phosphate group would also co-ordinate to Mg-2 (O1—Mg-2 = 2.30 Å) displacing Wat-322. Finally, the O8 of the substrate phosphate group would displace Wat-178, forming a ligand with Mg-3b (O8—Mg-3b = 2.76 Å). The octahedral co-ordination geometry at all three sites is retained.

The third metal binding site in each subunit is presumably a low affinity site present only at high Mg^{2+} concentrations.

Although three metal binding sites are observed in each subunit of bovine IMPase structure, a two-metal mechanism is proposed in accordance with CD spectroscopic, kinetic and fluorescence studies. The reaction mechanism of bovine IMPase is deemed to be similar to that of the human enzyme, since the active site architecture and metal binding sites are highly conserved. Based on the human IMPase structure, the most likely contender for the role of the nucleophile is a metal-bound water molecule (Wat-50), which directly attacks the phosphate ester of the substrate (Bone *et al.*, 1992). Each subunit of bovine IMPase possesses a water molecule located at an equivalent position that forms interactions with Mg-1, Mg-3b (only in subunit B), Glu-70 and Thr-95. The water molecule is most likely activated through hydrogen-bonding interactions with Thr-95 (2.7 Å) and Glu-70 (3.2 Å) as observed in the human enzyme.

Based on the modelling of the substrate D-Ins(1)P into the active site, a mechanism for the reaction catalysed by bovine IMPase can be envisaged (Figure 3.17). The Mg^{2+} at site 1 interacts with the phosphate moiety of the substrate, implying that its primary role is to orientate the phosphate into the appropriate conformation for attack by the nucleophile. The specific role of Glu-70 most likely involves activating the water molecule by abstracting a proton, whereas Thr-95 participates in a secondary role by positioning the activating water molecule in line with the leaving group (inositol moiety) during the transition state. The important roles played by Glu-70 and Thr-95 have been confirmed by engineering site directed mutants of human IMPase (E70Q and T95A). The metal binding capabilities of the E70Q mutant was unaffected, although the k_{cat} decreased by 7000-fold, suggesting that Gln-70 does not function to activate the water molecule. Similarly, the k_{cat} for the T95A mutant was reduced by 14,000-fold while increasing the K_m for Mg^{2+} 7-fold (Pollack *et al.*, 1993).

Following the attack by the activated water molecule upon the phospho ester bond of the substrate, the expulsion of the leaving group could occur via two different mechanisms. The first mechanism posits that one of the carboxyls of Asp-93 or Asp-220 (both are within 3.2 Å of the ester oxygen) donate a proton to the ester oxygen. The second

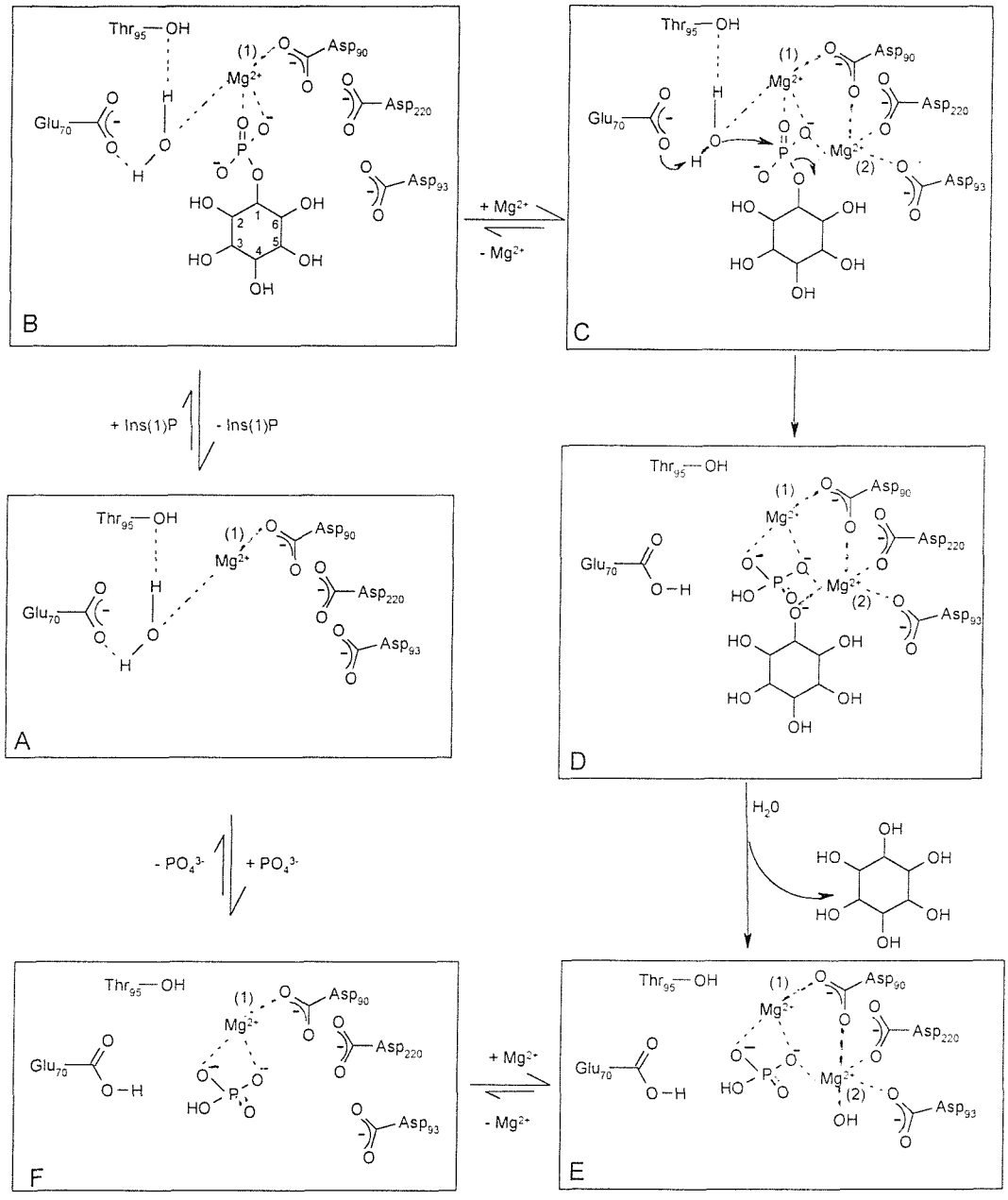


Figure 3.17: The proposed mechanism for the reaction catalysed by bovine inositol monophosphatase. Following the binding of Ins(1)P, the Mg^{2+} at site 1 orientates the phosphate moiety of the substrate into the appropriate conformation (B). A metal bound water molecule serves as a nucleophile (activated by forming hydrogen bonds with Glu-70 and Thr-95) and attacks the phosphate phosphorous following the binding of a second Mg^{2+} ion at site 2 (C). As the phosphate ester bond is cleaved, the developing negative charge on the ester oxygen is stabilised by the Mg^{2+} at site 2 to form a Mg^{2+} oxyanion (D). Expulsion of the leaving group (inositol) is aided by the attack of water from the medium (E). The Mg^{2+} at site 2 then dissociates, leaving an enzyme phosphate complex (F). The phosphate group is then released to regenerate the enzyme ready for the next catalytic cycle (A). In the presence of high Mg^{2+} concentrations, a third Mg^{2+} ion binds at site 3 and co-ordinates to Glu-70. As a result, Glu-70 can no longer serve as a base and fails to activate the water molecule.

mechanism posits that the Mg^{2+} at site 2 co-ordinates to the ester oxygen and stabilises the negative charge as the phosphate ester bond is cleaved. The first mechanism is less favoured because the carboxylates of Asp-93 and Asp-220 form ligands to the Mg^{2+} at site 2 and thus would be unsuitable for donating a proton. In contrast, support for the second mechanism (Figure 3.17) is provided by modelling studies with substrate, whereby Mg^{2+} at site 2 can form a ligand with the ester oxygen to form a Mg^{2+} oxyanion. The inositol moiety can then be expelled following attack by water from the medium. In the final stages, Mg^{2+} at site 2 dissociates and the phosphate is released to render the enzyme ready for another catalytic cycle.

A possible explanation for enzyme inhibition at high concentrations of activating metals has been provided by the human IMPase complex with Mn^{2+} and chloride; human IMPase co-crystallised with 5 mM $MnCl_2$ possesses a third metal binding site (Mn-3) that displaces the metal bound water nucleophile (Bone *et al.*, 1994b). However, the water nucleophile is clearly observed in the bovine IMPase complex with Mg^{2+} and the human IMPase complexes with Ca^{2+} and D-Ins(1)P, even though each of the structures possess a third metal binding site. It is possible that Glu-70 no longer has a negative charge following its co-ordination to the third metal binding site and thus fails to activate the water molecule. Another possibility is that at high concentrations of Mg^{2+}, site 2 is always saturated and therefore traps the phosphate moiety in an analogous manner to lithium inhibition.

The various effects of different metal ions on IMPase activity can be rationalised in terms of the size of a particular metal ion. The enzyme is most active in the presence of its natural metal ligand Mg^{2+} (ionic radius = 0.65 Å). There is an obvious correlation between the ionic radius of a metal ion and IMPase activity, since the likelihood of enzyme inhibition is increased in the presence of larger metal ions; the loss of enzyme activity in the presence of Gd^{3+} (ionic radius = 1.04 Å) or Ca^{2+} (0.99 Å) is most likely due to the presence of large atoms that disrupt the natural metal co-ordination geometry. In contrast, smaller divalent metal ions such as Zn^{2+} and Mn^{2+} (with an ionic radius of 0.81 and 0.74 Å, respectively) support 50 and 25 % IMPase activity, respectively (Pollack *et al.*, 1994). The metal-ligand co-ordination geometry for these divalent metal ions (normally octahedral) is presumably close to that of Mg^{2+}, resulting in the retention, to some extent, of IMPase activity.

3.7 Further Work

The final stage of refinement of the bovine IMPase structure is currently in progress. The occupancy of each of the metal binding sites needs to be refined. Furthermore, for completeness there are a number of residues with alternative conformations that have to be modelled and refined. Solving the bovine IMPase structure in the presence of active Mg^{2+} concentrations (2 mM) will provide further insight into the mechanism of the enzyme. X-ray data from crystals of G69S, E70Q and D90N mutants have been collected recently, and thus may elucidate the exact roles played by Gly-69, Glu-70 and Asp-90 in the bovine IMPase mechanism. Most importantly, since these crystals diffract to atomic resolution it is possible that the lithium binding site may be identified.

3.8 References

American Psychiatric Association. (1994) *Am. J. Psych* **151 (12)** 1S-36S

Atack, J.R., Broughton, H.B. and Pollack, S.J. (1995a) *FEBS. Lett.* **361** 1-7

Atack, J.R., Broughton, H.B. and Pollack, S.J. (1995b) *Trends Neurosci.* **18** 343-349

Attwood, P.V., Ducep, J.B. and Chanal, M.C. (1988) *Biochem. J.* **253 (2)** 387-394

Badyal, R. (2000) Ph. D. Thesis, University of Southampton

Berridge, M.J. (1987) *Annu. Rev. Biochem.* **56** 159-193

Berridge, M.J., Downes, P.C. and Hanley, M.R. (1989) *Cell* **59** 411-419

Berridge, M.J. and Irvine, R.F. (1989) *Nature* **341** 197-204

Bone, R., Springer, J.P. and Atack, J.R. (1992) *Proc. Natl. Acad. Sci.* **89** 10031-10035

Bone, R., Frank, L., Springer, J.R., Pollack, S.J., Osborne, S., Atack, J.R., Knowles, M.R., McAllister, G., Ragan, I.C., Broughton, H.B., Baker, R. and Fletcher, S.R. (1994a) *Biochemistry* **33** 9460-9467

Bone, R., Frank, L., Springer, J.R. and Atack, J.R. (1994b) *Biochemistry* **33** 9468-9476

Cade, J.F.J. (1949) *Med. J.* (Australia) **36** 349-352

Chen, I.W. and Charalampous, F.C. (1966) *J. Biol. Chem.* **241** 2194-2199

Dixon, J.F. and Hokin, L.E. (1997) *Proc. Nat. Acad. Sci.* **94** 4757-4760

Eisenberg, F., Jr. (1967) *J. Biol. Chem.* **242** 1375-1382

Emilien, G., Maloteaux, J.M., Seghers, A. and Charles, G. (1995) *Archives Internationales de Pharmacodynamie et de Therapie* **330 (3)** 251-278

Garza-Trevino, E.S. (1997) *Drugs of Today* **34 (4)** 295-305

Ganzhorn, A.J. and Chanal, M.C. (1990) *Biochemistry* **29** 6065-6071

Ganzhorn, A.J. and Rondeau, J.M. (1997) *Protein Engineering* **10** (s) 61

Gee, N.S., Ragan, C.I., Watling, K.J., Aspley, S., Jackson, R.G., Reid, G.G., Gani, D. and Shute, J.K. (1988) *Biochem. J.* **249** 883-889

Gore, M.G., Greasley, P.J. and Ragan, C.I. (1992) *J. Biochem. Biophys. Methods* **25** 55-60

Gore, M.G., Greasley, P.J., Ragan, C.I. and McAllister, G. (1993) *Biochem. J.* **296** 811-815

Greasley, P.J., Gore, M.G., Rees-Milton, K.J. and Ragan, C.I. (1993) *FEBS Lett.* **319** 49-53

Greasley, P.J. and Gore, M.G. (1993) *FEBS lett.* **331** 114-118

Greasley, P.J., Hunt, L.G. and Gore, M.G. (1994) *Eur. J. Biochem.* **222** 453-460

Hallcher, L.H. and Sherman, W.R. (1980) *J. Biol. Chem.* **255** 10896-10901

Harding, M.M. (1999) *Acta. Crystallogr.* **D55** 1432-1443

Huheey, J.E. (1978) in *Inorganic Chemistry*, second edition, Harper and Row, New York

Irvine, R.F. and Moor, R.M. (1986) *Biochem. J.* **240** 917-920

Karasu, T.B., Gelenberg, A., Wang, P *et al* (2000) *Am. J. Psychiat.* **157** 1-45

Katzung, B.G. (1995) in *Basic and Clinical Pharmacology*, Conneticut, Appleton and Lance

Ke, H.M., Thorpe, C.M., Seaton, B.A., Marcus, F. and Lipscomb, W.N. (1989) *Proc. Natl. Acad. Sci.* **86 (5)** 1475-1479

Laskowski, R.A., McArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Cryst.* **26** 283-291

Manji, H.K., Chen, G., Hsiao, J.K., Risby, E.D., Masana, M.I. and Potter, W.Z. (1996) *J. Clin. Psych.* **57 (13)** 34-46

McAllister, G., Whiting, P., Hammond, E.A., Knowles, M.R., Atack, J.R., Bailey, F.J., Maigetter, R. and Ragan, C.I. (1992) *Biochem. J.* **284 (3)** 749-754

Nishizuka, Y. (1992) *Science* **258** 607-614

Parthasarathy, L., Vadnal, R.E., Ramesh, T.G., Shyamaladevi, C.S. and Parthasarathy, R. (1993) *Arch. Biochem. Biophys.* **304 (1)** 94-101

Pollack, S.J., Atack, J.R., Knowles, M.R., McAllister, G., Ragan, I.C., Baker, R., Fletcher, S.R., Iversen, L.L. and Broughton, H.B. (1994) *Proc. Natl. Acad. Sci.* **91** 5766-5770

Pollack, S.J., Knowles, M.R., Atack, J.R., Broughton, H.B., Ragan, C.I., Osborne, S. and McAllister, G. (1993) *Eur. J. Biochem.* **217** 281-287

Rees-Milton, K., Thorne, M., Greasley, P., Churchich, J. and Gore, M.G. (1997) *Eur. J. Biochem.* **246** 211-217

Rosenthal, N.E. and Goodwin, F.K. (1982) *Annu. Rev. Med.* **33** 555-568

Shears, S.B. (1989) *Biochem. J.* **260 (2)** 313-324

Shute, J.K., Baker, R., Billington, D.C. and Gani, D. (1988) *J. Chem. Soc. Commun.* **9** 626-628

Silva, N. (1999) Ph. D. Thesis, University of Southampton

Takimoto, K., Okada, M., Matsuda, Y. and Nakagawa, H. (1985) *J. Biochem.* **98 (2)** 363-370

Thorne, M.R., Greasley, P.J. and Gore, M.G. (1996) *Biochem. J.* **315** 989-994

Vallejo, M., Jackson, T., Lightman, S. and Hanley, M.R. (1987) *Nature* **330** 656-658

York, J.D., Ponder, J.W., Chen, Z.W., Mathews, F.S. and Majerus, P.W. (1994) *Biochemistry* **33 (45)** 13164-13171

Whiting, P., Gee, N.S., Potter, J., Howell, S. and Ragan, C.I. (1990) *Biochem. J.* **272** 465-468

Wickham, E.A. and Reed, J.V. (1987) *Int. Clin. Psychopharmacol.* **2** 181-190

Chapter 4

Structural Studies of D303E Methanol Dehydrogenase from *Methylobacterium extorquens*

4.1 Introduction to Methanol Dehydrogenase

4.1.1 Methylotrophs

Micro-organisms which utilise C1 compounds such as methanol, methane and methylamine for growth are referred to as methylotrophs. Methylotrophic bacteria can be subdivided into two groups depending on their substrate requirements; obligate methylotrophs and facultative methylotrophs. The former have an absolute requirement for C1 compounds, whereas the latter can also grow on multi-carbon substrates such as pyruvate and succinate. *Methylobacterium extorquens* (a pink facultative methylotroph) is one of the most commonly used and characterised methylotrophs that can utilise methanol and methyalamine but not methane for growth.

4.1.2 Electron Transport Chain in Methylotrophs

In most methylotrophic bacteria, the oxidation of methanol to formaldehyde is catalysed by methanol dehydrogenase (MDH, EC 1.1.99.8), a periplasmic quinoprotein. The oxidation reaction is characterised by the release of two protons into the periplasm and the transfer of two electrons to the electron acceptor, cytochrome c_L (Anthony, 1986). These electrons are subsequently transferred to cytchrome c_H and finally to cytochrome aa_3, the terminal oxidase. During this process cytochrome aa_3 accepts two protons from the cytoplasmic side and oxygen is reduced to water (Figure 4.1). The separation of electric charge across the membrane sets up an electric field, generating a protonmotive force that can be utilised by H^+ translocating ATP synthase to produce ATP.

4.1.3 Functional Studies of MDH

MDH from *M. extorquens* was initially purified and characterised by Anthony and Zatman in 1964. Since then MDH from various methylotrophic bacteria have been characterised displaying similar properties. The periplasmic enzyme utilises a novel prosthetic group, pyrrolo-quinoline quinone (PQQ) to oxidise primary alcohols such as methanol and ethanol (Anthony and Zatman, 1965; Anthony, 1986). MDH is also capable of oxidising formaldehyde to formate, although this reaction is undesirable *in-vivo*, since formaldehyde is essential for assimilation into the bacterial cell wall (Anthony, 1982).

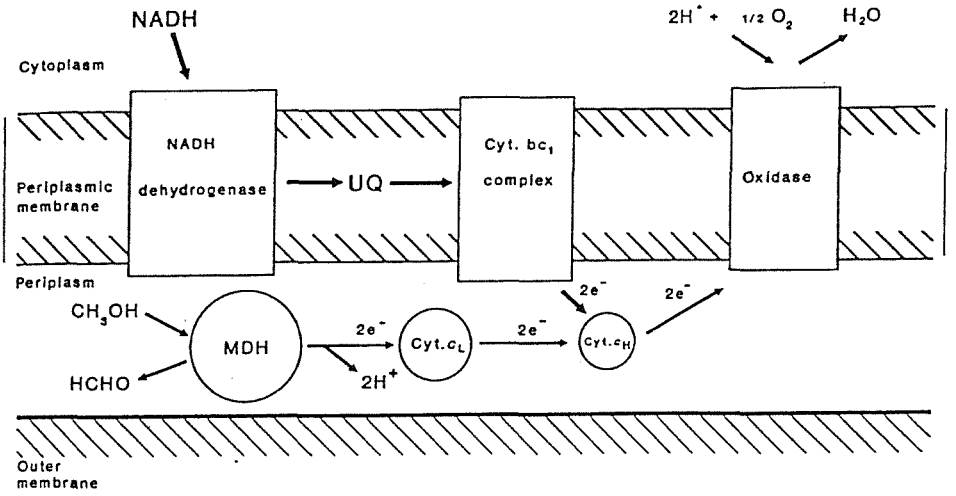


Figure 4.1: The electron transport chain for methanol oxidation. Following the reaction of MDH with cytochrome c_L (in two successive stages), two protons are released into the periplasm and two protons (per two electrons) are consumed on the cytoplasmic side of the oxidase, thus establishing a protonmotive force (Anthony, 1992a).

The affinity of MDH for formaldehyde is regulated by M-protein (Bolbot and Anthony, 1980), a modifier periplasmic protein complex made up of 3 subunits (45 kDa each). In the presence of M-protein, the affinity of MDH for formaldehyde is reduced (Long and Anthony, 1991). Furthermore, M-protein increases the affinity of MDH for poor substrates such as 1,2-propanediol, hence regulating formaldehyde oxidation (Page and Anthony, 1986). MDH activity can be assayed in the presence of artificial electron acceptors (phenazine ethosulphate (PES) and Wurster's blue) at pH 9.0, with ammonia or methylamine serving as activators. In contrast, ammonia activation of MDH at pH 7.0 with the natural electron acceptor (cytochrome c_L), is not essential (Duine *et al.*, 1979). The mechanism of MDH activation by ammonia is uncertain, although it may involve the formation of an iminoquinone complex (Ohshiro and Itoh, 1993).

4.1.4 Structure of the α and β Subunits

The three dimensional structure of MDH from *M. extorquens* has been determined at 1.94 Å resolution using the MIR method (Ghosh *et al.*, 1995). The first reported structures of a PQQ-containing MDH were determined at 2.6 Å resolution from *Methylophilus methylotrophus* and *Methylophilus* W3A1 (Xia *et al.*, 1992). MDH is an $\alpha_2\beta_2$ tetramer with each of the α and β subunits consisting of 600 and 74 amino acid residues, respectively. The mass of each α subunit is 66 kDa and that of each β subunit is 8.5 kDa, yielding a mass of 149 kDa for the assembled protein. Each α subunit ligates a non-covalently bound prosthetic group PQQ, as well as one Ca^{2+} ion which may have a dual structural and catalytic role.

The α subunit of MDH has a superbarrel structure consisting of 8 identical four stranded twisted anti-parallel β sheets (W-shaped) which are arranged radially around a pseudo 8-fold axis running through the centre of the subunit. Each 'W' motif resembles a propeller blade and hence the structure is often referred to as a propeller fold (Figure 4.2). Propeller blade structures were first identified in the influenza virus neuraminidase (Varghese *et al.*, 1983), and have subsequently been identified in several other proteins including haemopexin (Faber *et al.*, 1995), methylamine dehydrogenase (Vellieux *et al.*, 1989), galactose oxidase (Ito *et al.*, 1994) and nitrite reductase (Fulop *et al.*, 1995).

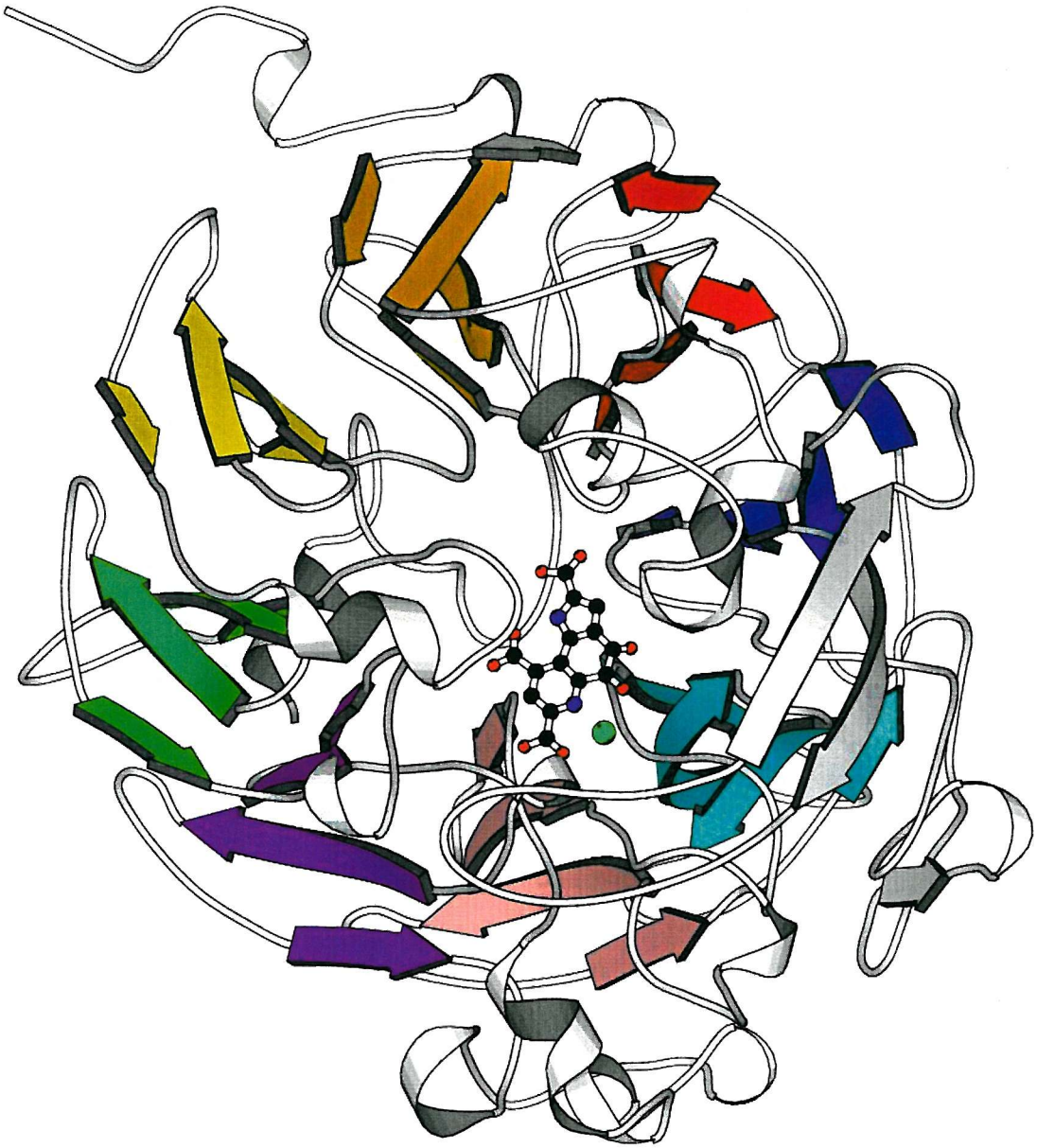


Figure 4.2: The α subunit of MDH from *M. extorquens* (Ghosh *et al.*, 1995) showing the β strands of the 'W' motifs (MOLSCRIPT). Each of the propeller blades are coloured differently. The pyrrolo-quinoline quinone (PQQ) prosthetic group and the Ca^{2+} ion (green) are depicted in CPK representation.

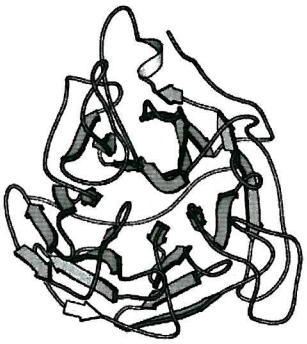
The number of propeller blades present within protein superbarrel structures varies; there are four in haemopexin, six in neuraminidase, seven in methylamine dehydrogenase and galactose oxidase and eight in nitrite reductase (Figure 4.3). The eight bladed β propeller structure of the latter is similar to that of MDH, although low sequence identity exists between the two proteins. The propeller blades of the α subunit of MDH are stabilised by a unique set of tryptophan docking interactions (Ghosh *et al.*, 1995), which are absent in the aforementioned β propeller blade structures. This is surprising, since the eight bladed superbarrel domain of nitrite reductase can be closely superimposed on that of MDH (Anthony and Ghosh, 1998).

The *N*-terminal region of the β subunit consists of one intra-chain disulphide bridge (Cys-6 and Cys-12) and a proline rich segment (residues 14 -20), whereas the *C*-terminal region consists of a single 7 turn α helix comprising 30 residues. Overall, the β chain can be represented as a 'J' shaped unit, with a long α helix lying on the surface of the larger α subunit. Interactions between the two subunits predominantly consist of ion pairs (40 % of residues present in the β subunit are charged). The two subunits are further stabilised through hydrogen bonding and hydrophobic interactions. The initial function proposed for the β subunit involved forming docking interactions with cytochrome c_L (Nunn *et al.*, 1989). However, chemical cross-linking studies demonstrated that MDH and cytochrome c_L dock together via electrostatic interactions, between lysine residues on the α subunit of MDH and carboxylates on cytochrome c_L (Anthony, 1992b). It is possible that the β subunit stabilises the fold of the larger α chain (Anthony and Ghosh, 1998), although its absence from other quinoproteins strongly suggests that it may play an additional specific role in MDH.

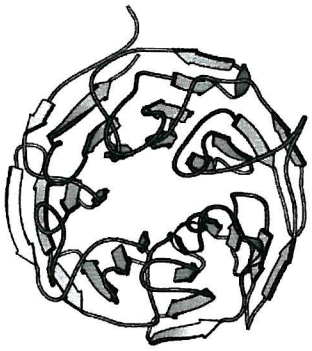
4.1.5 The Active Site of MDH

The active site is located in the α subunit and can be accessed through a shallow funnel composed of mainly hydrophobic residues, which leads to a narrow entrance to the cavity containing the PQQ and the Ca^{2+} ion. There are no interactions between the active sites of both α subunits. The PQQ moiety interacts with the enzyme via axial and equatorial interactions, the former involving two non-polar interactions, which result in the

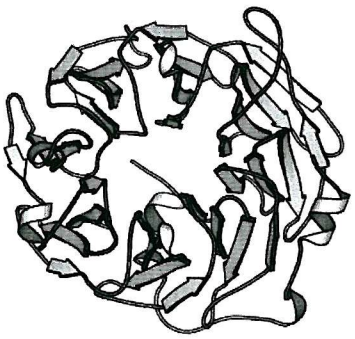
Haemopexin from *Oryctolagus cuniculus*



Neuraminidase from *Influenza B virus*



Galactose Oxidase from *Dactylium dendroide*



Nitrite Reductase
from *Paracoccus denitrificans*

Figure 4.3: The β-propeller domain of four representative proteins (MOLSCRIPT).

'sandwiching' of PQQ between the indole ring of Trp-243 and a novel disulphide ring structure (Cys-103 and Cys-104). The equatorial interactions include hydrogen bonding and ion pairing between PQQ and residues, present in the A strands of the 'W' motifs (Figure 4.4) (Ghosh *et al.*, 1995). In addition, the O-4 and O-5 atoms of PQQ are hydrogen bonded to the NH1 and NH2 atoms of Arg-331, respectively. The O-4 atom also forms a weak hydrogen bond with the amide NH2 of Asn-394. As well as interacting with the PQQ, Arg-331 forms a hydrogen bond with the carboxylate group of Asp-303.

The Ca^{2+} ion forms three metal ligands to PQQ (Figure 4.4), involving the O-5, O-7 and N6 ring atoms. The metal ion also interacts with three protein atoms (the OD2 atom of Asn-261 and the OE1 and OE2 atoms of Glu-177), giving a total of six metal ligands with distances varying from 2.4-2.8 Å. Comparisons of the two available MDH holoenzyme structures, has not allowed determination of the substrate binding site; the solvent molecules which may occupy the same space as the methanol group are in different positions (Ghosh *et al.*, 1995; Xia *et al.*, 1996). Although the latest published structure of *Methylophilus* W3A1 (W3A1-MDH) refined to 1.9 Å claimed that the methanol binding site is proximal to Asp-297 (equivalent to Asp-303 in *M. extorquens*) and the O-5 of PQQ (see Figure 4.11a) (Xia *et al.*, 1999), this interpretation has recently been retracted (Zheng *et al.*, 2001).

4.1.6 The Catalytic Cycle of MDH

The catalytic cycle of MDH (Figure 4.5) has been difficult to elucidate since the enzyme can only be isolated in the semi-reduced quinone form (MDH_{sem}) and is inactivated in the presence of an artificial electron acceptor (Anthony and Zatman, 1967; Duine and Frank, 1980). Protection against inactivation in the absence of substrate is provided by cyanide, a competitive inhibitor that allows isolation of the oxidised form of MDH (MDH_{ox}) with PQQ intact. The reduced form of MDH (MDH_{red}) with the bound quinol ($PQQH_2$) can be generated by the addition of substrate to MDH_{ox} (Duine and Frank, 1980). After the substrate binds the proposed mechanism involves simultaneous reduction of PQQ and release of the aldehyde product. During the reoxidation stage ($PQQH_2$ oxidised back to quinone), two electrons are transferred by an unknown mechanism to the electron acceptor cytochrome c_L while generating a semiquinone intermediate (Frank *et al.*, 1988). The

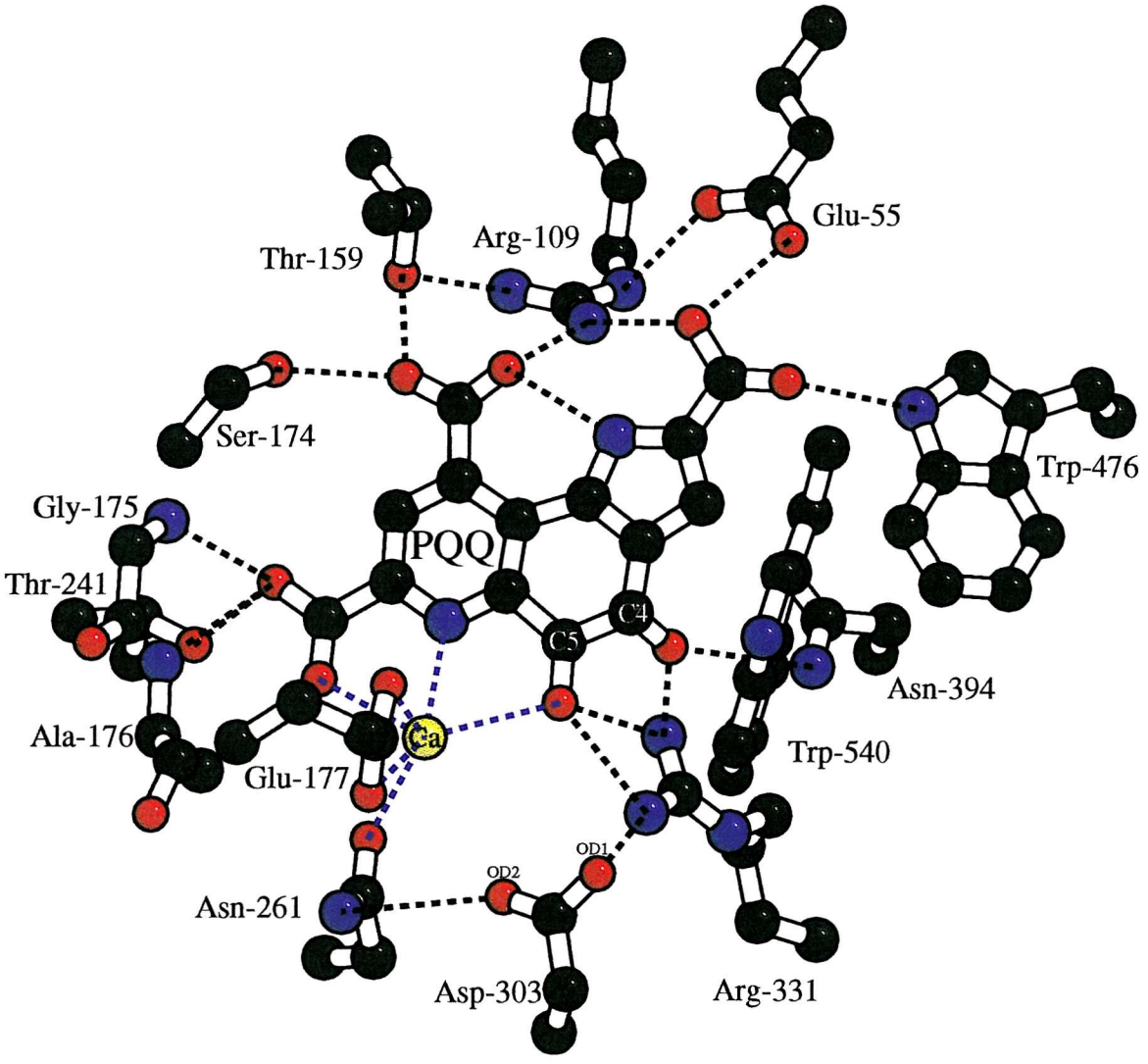


Figure 4.4: The active site equatorial interactions of PQQ (depicted in black) and the co-ordination ligand sphere of Ca^{2+} (depicted in blue) in MDH from *M. extorquens* (BOBSCRIPT). The Ca^{2+} ion is hexa-coordinated to three PQQ atoms (O-5, O-7 and N6) and three protein atoms (OD2 atom of Asn-261 and both carboxylate oxygen atoms of Glu-177). The C-5 carbonyl oxygen atom (O-5) is in the plane of the ring, whereas the C-4 oxygen (O-4) is out of the plane by 40° (Ghosh *et al.*, 1995).

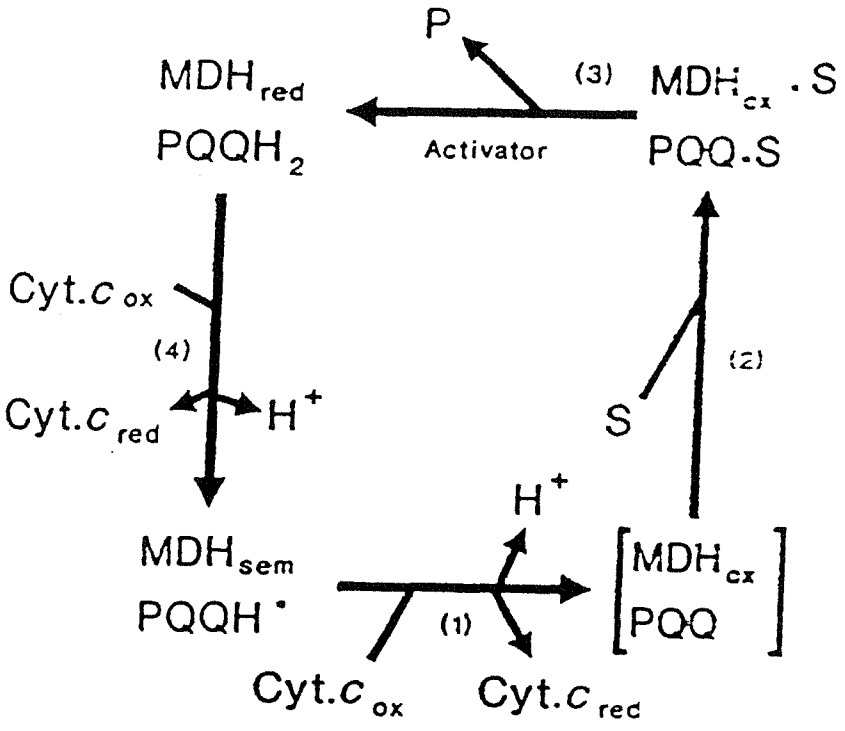


Figure 4.5: The catalytic cycle of MDH. The addition of methanol (S) to the oxidised form of MDH (MDH_{ox}) gives reduced MDH containing the quinol ($PQQH_2$). The oxidation of the quinol generates two protons and two electrons. The latter are transferred individually to electron acceptor cytochrome c_L (Cyt.c), during which the reduced quinol is oxidised back to quinone (PQQ) via a semiquinone intermediate ($PQQH^{\bullet}$). The enzyme is isolated as MDH_{sem} (Anthony and Ghosh, 1998).

electron transfer is inhibited in the presence of EDTA, although the electrostatic interactions between MDH and cytochrome c_L remain intact. This suggests that the electrostatic interactions are vital for holding the two proteins together, but must subsequently rearrange for optimal electron transfer (Dales and Anthony, 1995).

4.1.7 The Mechanism of MDH

The C-5 carbonyl group of PQQ is highly reactive towards nucleophilic reagents and adducts are formed readily with methanol, aldehydes, ketones and ammonia (Ohshira and Itoh, 1993). Further evidence supporting the involvement of the carbonyl group in the reaction is provided by the ability of MDH to oxidise cyclopropanol, which must proceed via the formation of a C-5 propanal complex. It is possible that the mechanism is initiated by proton abstraction by an active site base, allowing the resultant carboanion to attack the C-5 of PQQ (Frank *et al.*, 1989). A similar mechanism may account for the oxidation of methanol by MDH, with proton abstraction being followed by the formation of a C-O bond to give a hemiketal intermediate (Figure 4.6a).

X-ray structural studies of MDH from *M. extorquens* have identified Asp-303 as a possible candidate for the catalytic base (Ghosh *et al.*, 1995), since it is located in close proximity to the C-5 carbonyl oxygen of PQQ (see Figure 4.4). The reaction could be initiated by abstracting a proton from methanol. The resultant oxyanion may in turn attack the electrophilic C-5, leading to the generation of the proposed hemiketal intermediate (Figure 4.6a). Following reduction of PQQ, expulsion of the aldehyde product could be aided by the ionisation of the hemiketal complex, which would probably involve the N1 atom of the pyrrole ring. An alternative mechanism has been proposed which involves a simple acid/base catalysed hydride transfer, with Asp-303 again serving as the catalytic base (Figure 4.6b).

In both proposed mechanisms, Ca^{2+} functions as a Lewis acid by co-ordinating to the C-5 carbonyl oxygen of PQQ, and stabilising the electrophilic C-5 for attack by either a hydride or an oxyanion (Anthony *et al.*, 1994; Anthony 1996, 1997). Evidence for the involvement of Ca^{2+} in the reaction catalysed by MDH is provided by the ability of *M. extorquens* to grow in high concentrations of strontium (Sr^{2+}) (Harris and Davidson, 1994). In addition,

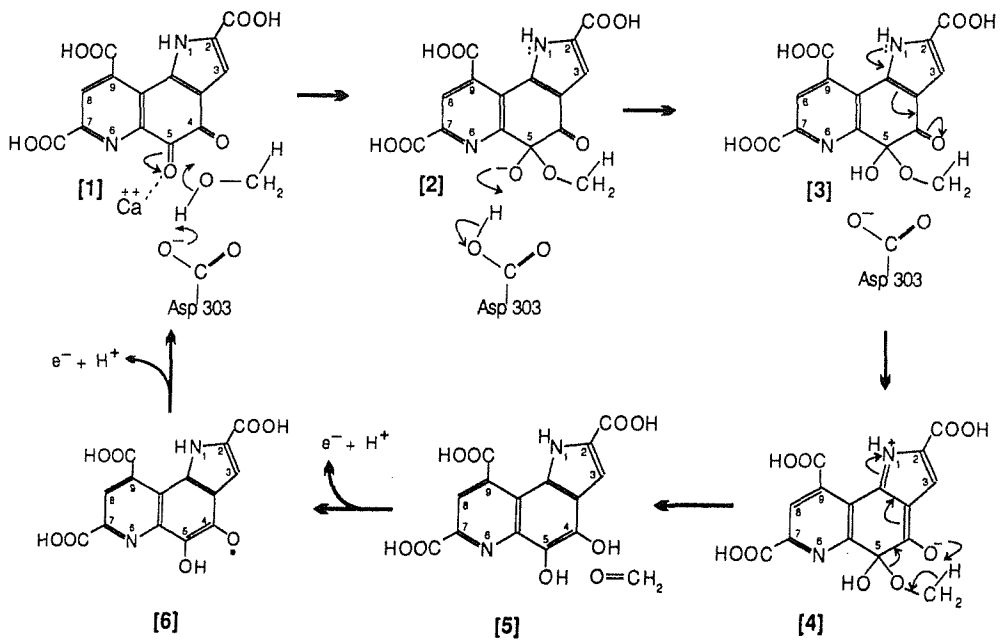


Figure 4.6a: Mechanism for MDH involving formation of a hemiketal intermediate. Following proton abstraction from methanol by the base (Asp-303), the Ca^{2+} ion promotes attack by the oxyanion on the electrophilic C-5, generating the hemiketal intermediate (Anthony, 1996).

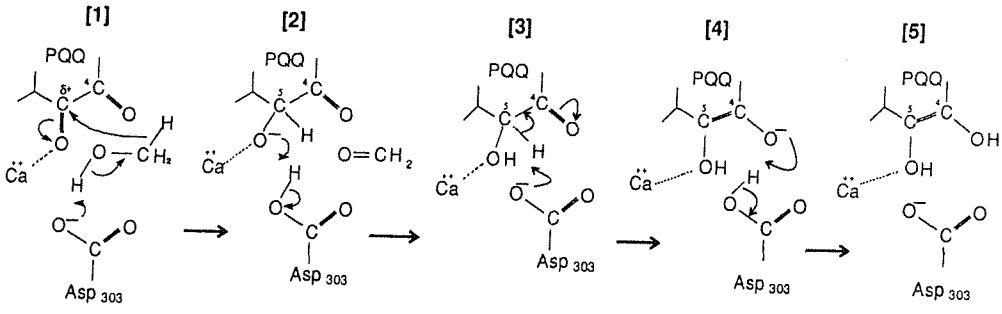


Figure 4.6b: Mechanism of MDH involving an acid/base hydride transfer mechanism. Following proton abstraction by the active site base (Asp-303), the electrophilic C-5 is directly involved in removing the methyl hydrogen as hydride (Anthony, 1996).

investigations involving the replacement of Ca^{2+} with barium (Ba^{2+}) have generated a Ba^{2+} containing MDH, which has a lower affinity for methanol but a higher catalytic rate (Goodwin and Anthony, 1996).

4.1.9 Comparisons of MDH with other PQQ Containing Proteins

The term quinoprotein was first coined in 1980 in describing MDH and glucose dehydrogenase (GDH) which both contain PQQ as their prosthetic group (Duine *et al.*, 1979, 1980). This term has since been used to describe other PQQ containing enzymes such as alcohol dehydrogenase (ADH), glycerol dehydrogenase and aldehyde dehydrogenase.

ADHs can be classed into one of 3 groups depending on their structure: (1) quinoprotein ADH (2) quinohaemoprotein ADH and (3) quinohaemoprotein-cytochrome c ADH. The first group is similar to MDH, in that it contains one PQQ molecule and a Ca^{2+} ion in each α subunit (Mutzel and Görisch, 1991), has a high pH optimum (9.0), and is activated by ammonia in the presence of PES. The second group has an in built electron acceptor in the form of haem C. Finally, the third group is made up of three different subunits (α, β and γ); the α subunit contains haem C and PQQ; the β subunit is cytochrome c and the function of the γ subunit remains to be determined. The sequence identity between the α subunit of quinohaemoprotein ADH and the α subunit of MDH is 31 %. As a result, the structure of quinohaemoprotein ADH has been modelled using MDH as a template (Cozier *et al.*, 1995). Despite differences in external loop regions involved in the formation of the shallow funnel, the region of the active site is highly conserved; all the axial and most of the equatorial interactions to PQQ are present in MDH. The conservation of the active site base and the ligands to the Ca^{2+} ion indicates that the reaction mechanism of ADH is similar to that of MDH.

GDH is a monomeric membrane protein that catalyses the oxidation of pyranose to lactone and utilises ubiquinone as the electron acceptor (Matsushita *et al.*, 1986). Comparisons of the primary sequence indicate that the *N*-terminal region may contain the ubiquinone binding site as well as forming a membrane anchor. The sequence identity between the periplasmic region of GDH (residues 155-796) and the α subunit of MDH is 26 %, which

has allowed the GDH structure to be modelled using the co-ordinates of MDH (Cozier and Anthony, 1995). The major differences in the model structure include the replacement of the disulphide ring with histidine which maintains the configuration of PQQ in the active site and supports a prediction that the residue is important for PQQ binding (Imanaga, 1989). The conservation of the proposed active site base and the co-ordination of the presumed Ca^{2+} ion suggest that the reaction mechanism catalysed by GDH is also similar to that of MDH (Cozier and Anthony, 1995).

4.1.9 Research Aims

Previous structural studies of MDH from *M. extorquens* have identified an aspartate residue (Asp-303) as the putative nucleophile. The role of the proposed nucleophile was further investigated by generating a site directed mutant (D303E) (Majekodunmi, 1997). Kinetic studies of D303E-MDH reveal an active enzyme with a 80,000-fold decrease in affinity for methanol (K_m up from 3 μM to 250 mM) and a decrease in the activation energy (from 35.4 to 20.2 kJ mol^{-1}), with a corresponding increase in V_{max} (from 0.81 to 1.25 μmol/min/mg), compared to the wild type enzyme (WT-MDH) (Afolabi, 1999). It was decided to determine the three-dimensional structure of D303E-MDH in order to gain a better insight into the role of the potential catalytic base and to account for its kinetic properties.

4.2 Expression, Purification and Crystallisation of D303E-MDH

The *mxaF* mutant strain of *M. extorquens* was grown on methanol in minimal culture medium as described by Day and Anthony (1990). The purification of D303E-MDH was based on a modified method also described by Day and Anthony (1990). Crystals of D303E-MDH were grown in the absence of methanol using the 'hanging drop' vapour diffusion method. Hanging drops were obtained by mixing equal volumes of protein solution (20 mg/ml D303E-MDH in 5 mM Tris-HCl buffer, pH 9.2) and reservoir solution (14.5 % PEG 6000, 20 mM Tris-HCl buffer (pH 9.5) and 10 mM $CaCl_2$). These were suspended above 1ml of reservoir solution in a sealed well and allowed to equilibrate by vapour diffusion. Crystals (1.0 x 0.2 x 0.1 mm^3) were obtained after incubation in the dark for one month at 4°C (Afolabi, 1999).

4.3 Data Collection and Processing of D303E-MDH

X-ray data from a single D303E-MDH crystal were collected 'in-house' at room temperature using a 300 mm MAR Image Plate X-ray detector, mounted on a ENRAF NONIUS FR591 rotating anode X-ray generator, which was operated at 50 kV and 100 mA. A total of 180 images, each covering an oscillation angle of 1°, were collected with an exposure time of 20 minutes per frame and a crystal to detector distance of 230 mm. The crystal initially diffracted X-rays to 2.7 Å but this had been reduced to 3.0 Å upon completion of data collection. Intensity data were processed with MOSFLM, and scaled and merged using programs of the CCP4 suite. The data processing statistics are listed in Table 4.1.

Total number of reflections	134792
Number of unique reflections	48502
Resolution (Å)	3.0
Completeness (%)	91.5
R_{merge} (%) (outer resolution shell)	9.4 (26.5)
Multiplicity	2.8
Average I/σ(I) (outer resolution shell)	6.9 (2.6)

Table 4.1: Data processing statistics for D303E-MDH crystal.

Initial estimates of the unit cell dimensions (a=101.9 Å, b=61.0 Å, c=212.7 Å and β=92.8°) were determined using the REFIX algorithm of MOSFLM. In order to overcome the space group ambiguity (P2 or $P2_1$), the data was reprocessed in P2 and pseudo precession pictures were generated using HKLVIEW to check for systematic absences. Unfortunately data for the 0K0 zone were missing, and unequivocal space group determination could only be made following a correct translation function solution during molecular replacement.

4.4 Molecular Replacement Studies on D303E-MDH

4.4.1 Introduction

Assuming the D303E-MDH crystals have two $\alpha_2\beta_2$ tetramers (298 kDa) per asymmetric unit, a solvent content of 44 % was calculated (V_m=2.22 $Å^3$/Da). Alternatively, if the crystals contain only one $\alpha_2\beta_2$ tetramer (149 kDa) per asymmetric unit, a solvent content of 72 % was calculated (V_m=4.44 $Å^3$/Da). In contrast, WT-MDH crystallised in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions a=66.8 Å, b=108.9 Å, c=189.2 Å and one $\alpha_2\beta_2$ tetramer per asymmetric unit. The calculated unit cell volume of WT-MDH is similar to that of D303E-MDH, and therefore it is likely that there are two $\alpha_2\beta_2$ tetramers within the asymmetric unit of the latter, giving a total of 4 $\alpha_2\beta_2$ tetramers in the unit cell.

4.4.2 Cross Rotation Search

Initial phases for D303E-MDH were obtained by molecular replacement using the program MOLREP. The search model consisted of a single $\alpha_2\beta_2$ tetramer of WT-MDH (dimensions of 69 x 94 x 96 Å), with the cofactor, metal ions and water molecules removed (Ghosh *et al.*, 1995). Reflections in the 20 to 3.0 Å resolution range were used in the cross rotation calculations, using a range of integration radii values (30-70 Å). The cross rotation function yielded two significant peaks at 12.6 σ and 10.7 σ, at a radius of integration of 55 Å (Table 4.2). These rotation function solutions (α=22.58°, β=65.86°, γ=297.91° and α=0.00°, β=71.83°, γ=271.09°) corresponded to the orientation of the two $\alpha_2\beta_2$ tetramers

within the target asymmetric unit and confirmed the solvent content calculations.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	22.58	65.86	297.91	12.57
2	0.00	71.83	271.09	10.74
3	180.38	90.00	150.91	8.57
4	291.46	84.58	202.77	8.21
5	269.12	84.13	350.16	8.08

Table 4.2: The 5 highest peaks following cross rotation calculations performed with MOLREP using data between 20 and 3 Å.

4.4.3 Translation Search

The initial translation search was performed in space group P2 and subsequently repeated in space group $P2_1$. Each rotation function solution yielded large translation function peaks of 24.7 σ and 17 σ (Tables 4.3 and 4.4) when searching in space group $P2_1$. No significant translation function peak was observed for solution 3 (α=180.38°, β=90°, γ=150.91°) of the cross rotation, providing further evidence for the existence of only two $\alpha_2\beta_2$ tetramers per asymmetric unit (Table 4.5).

Peak Number	X	Y	Z	Dens/σ
1	0.334	0.000	0.207	24.67
2	0.314	0.000	0.203	4.59
3	0.834	0.000	0.195	4.04
4	0.079	0.000	0.455	3.40
5	0.121	0.000	0.171	3.34

Table 4.3: The 5 highest translation function peaks calculated in space group $P2_1$, using the first orientation of the cross rotation (α=22.58°, β=65.86°, γ=297.91°).

Peak Number	X	Y	Z	Dens/σ
1	0.264	0.000	0.360	16.96
2	0.285	0.000	0.363	6.14
3	0.021	0.000	0.111	3.84
4	0.315	0.000	0.453	3.73
5	0.230	0.000	0.437	3.43

Table 4.4: The 5 highest translation function peaks calculated in space group $P2_1$, using the second orientation of the cross rotation (α=0.00°, β=71.83°, γ=271.09°).

Peak Number	X	Y	Z	Dens/σ
1	0.292	0.000	0.167	4.06
2	0.243	0.000	0.117	3.89
3	0.071	0.000	0.317	3.77
4	0.052	0.000	0.021	3.68
5	0.286	0.000	0.104	3.51

Table 4.5: The 5 highest translation function peaks calculated in space group $P2_1$, using the third orientation of the cross rotation (α=180.83°, β=90.00°, γ=150.91°).

The position of the first tetramer was fixed and the translation search was repeated to determine the position of the second tetramer with respect to the same origin. Since space group P2 is polar, the translation functions calculated with just one rotation function solution are two-dimensional. However, the final non-crystallographic translation function solution is three-dimensional (Table 4.6). The highest rotation function solution (α=22.58°, β=65.86°, γ=297.91°) yielded no translation function peak, when calculated in space group P2 (Table 4.7) providing unequivocal proof that D303E-MDH had crystallised in the monoclinic space group $P2_1$.

Peak Number	X	Y	Z	Dens/σ
1	0.773	0.725	0.868	27.36
2	0.065	0.860	0.859	13.15
3	0.764	0.535	0.860	12.75
4	0.765	0.801	0.860	11.77
5	0.767	0.874	0.870	10.86

Table 4.6: The 5 highest non-crystallographic translation function peaks after fixing the position of the first tetramer.

Peak Number	X	Y	Z	Dens/σ
1	0.380	0.000	0.375	4.66
2	0.492	0.000	0.020	4.47
3	0.098	0.000	0.171	3.83
4	0.463	0.000	0.116	3.70
5	0.089	0.000	0.161	3.69

Table 4.7: The 5 highest translation function peaks when calculated in space group P2, using the first orientation of the cross rotation (α=22.58°, β=65.86°, γ=297.91°).

4.5 Refinement of the D303E-MDH Structure

Following the placement of both $\alpha_2\beta_2$ tetramers in the asymmetric unit, the crystal packing of the solution was viewed using MOLPACK and seen to be sensible (Figure 4.7). Refinement of the model was performed with the program CNS using 48409 unique reflections within a resolution range of 20-3.0 Å. A test set of reflections (5 % of data) was used for R_{free} calculations. Initially, the model was subjected to 20 cycles of rigid body refinement, with each individual subunit being treated as a rigid body. Following rigid body refinement the R-factor decreased from 45.7 % (R_{free}=44.2 %) to 24.7 % (R_{free}=24.9 %). The resulting model was used to calculate electron density maps which were subjected to four fold NCS averaging.

Examination of the averaged maps using the program QUANTA revealed *Fo-Fc* density for one PQQ molecule (Figure 4.8) and one Ca^{2+} ion (Figure 4.9) in each α subunit (the

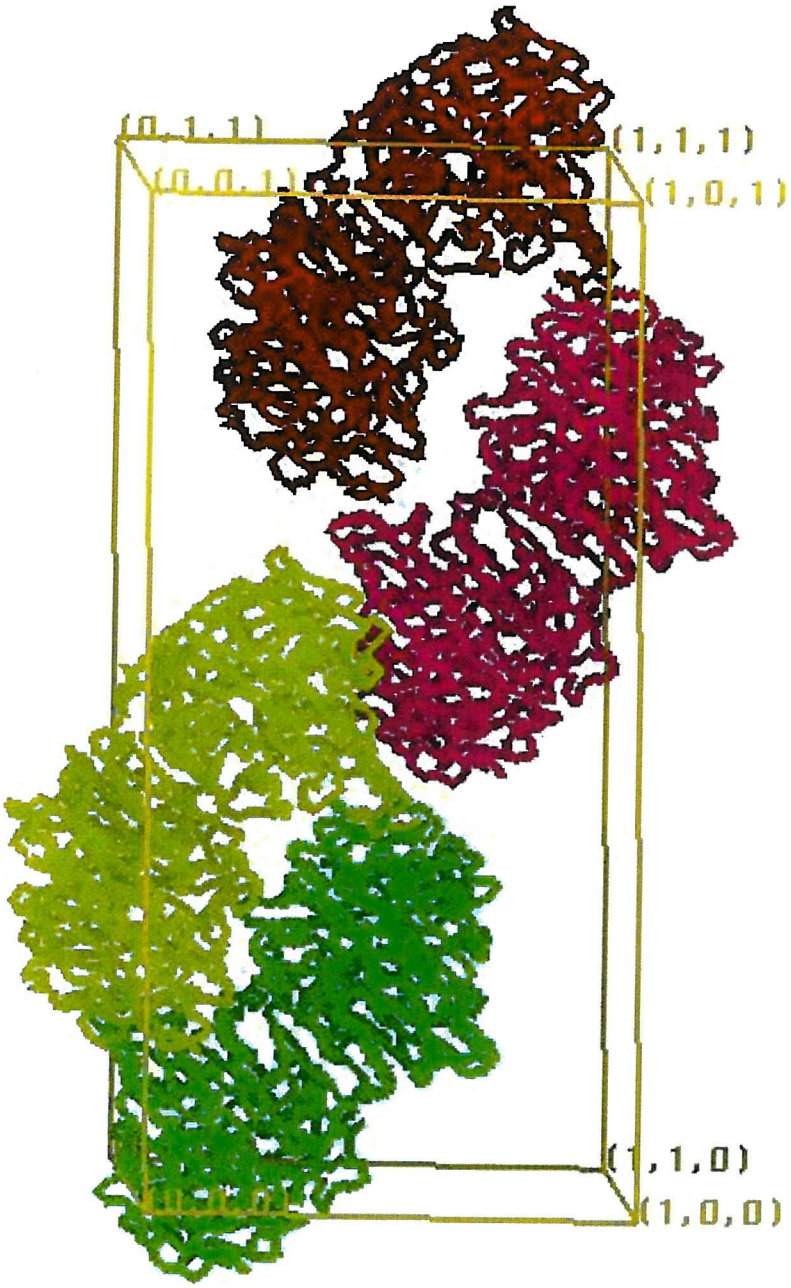


Figure 4.7: The crystal packing of D303E-MDH from *M. extorquens* (MOLPACK). There are four $\alpha_2\beta_2$ tetramers in the unit cell (two $\alpha_2\beta_2$ tetramers per asymmetric unit) and each is depicted in a different colour.

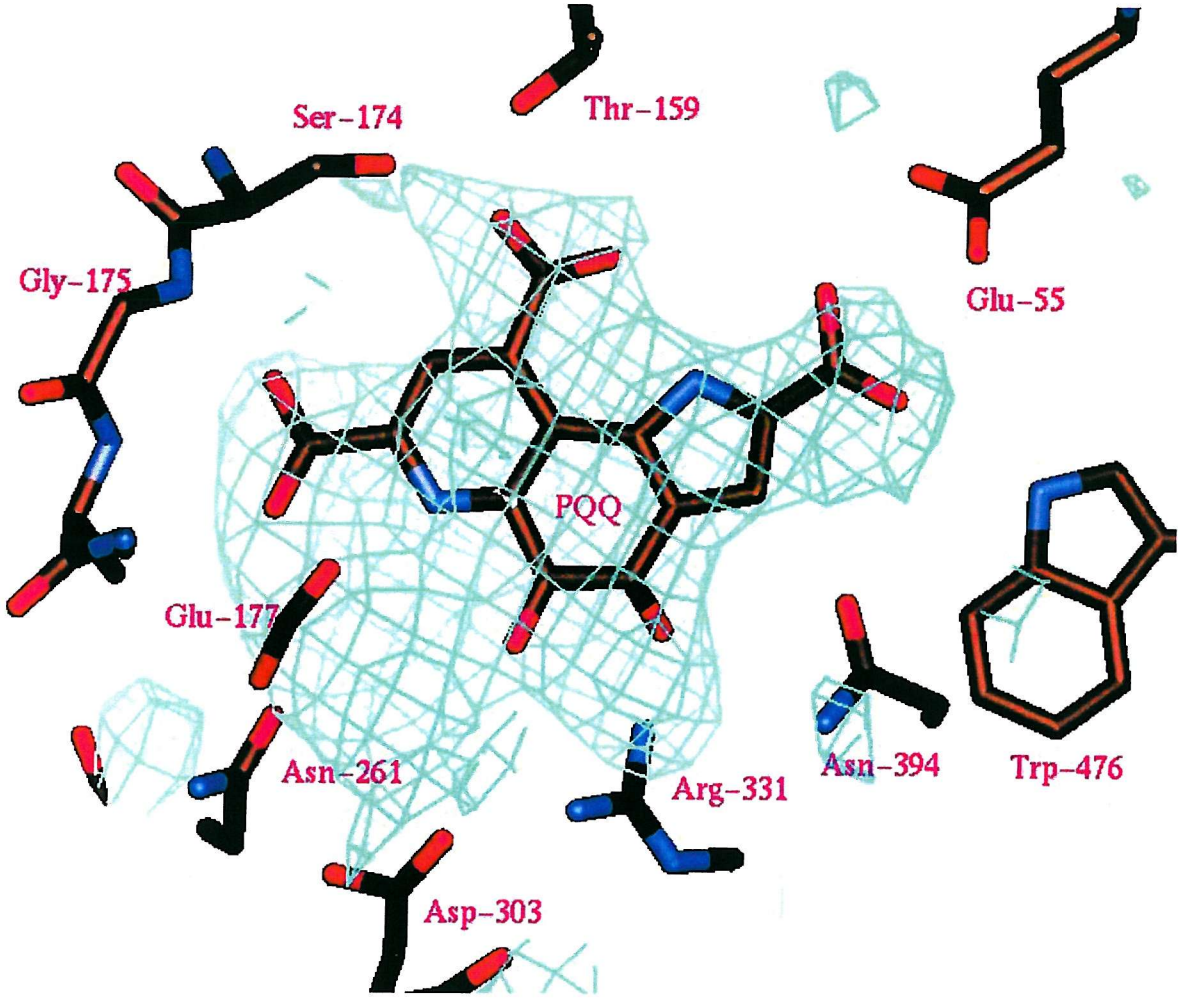


Figure 4.8: The active site of D303E-MDH from *M. extorquens* (QUANTA). Difference electron density (contoured at the 2 σ level) for the pyrrolo-quinoline quinone (PQQ) prosthetic group can be clearly observed. The active site side chains that form equatorial interactions with PQQ are also shown.

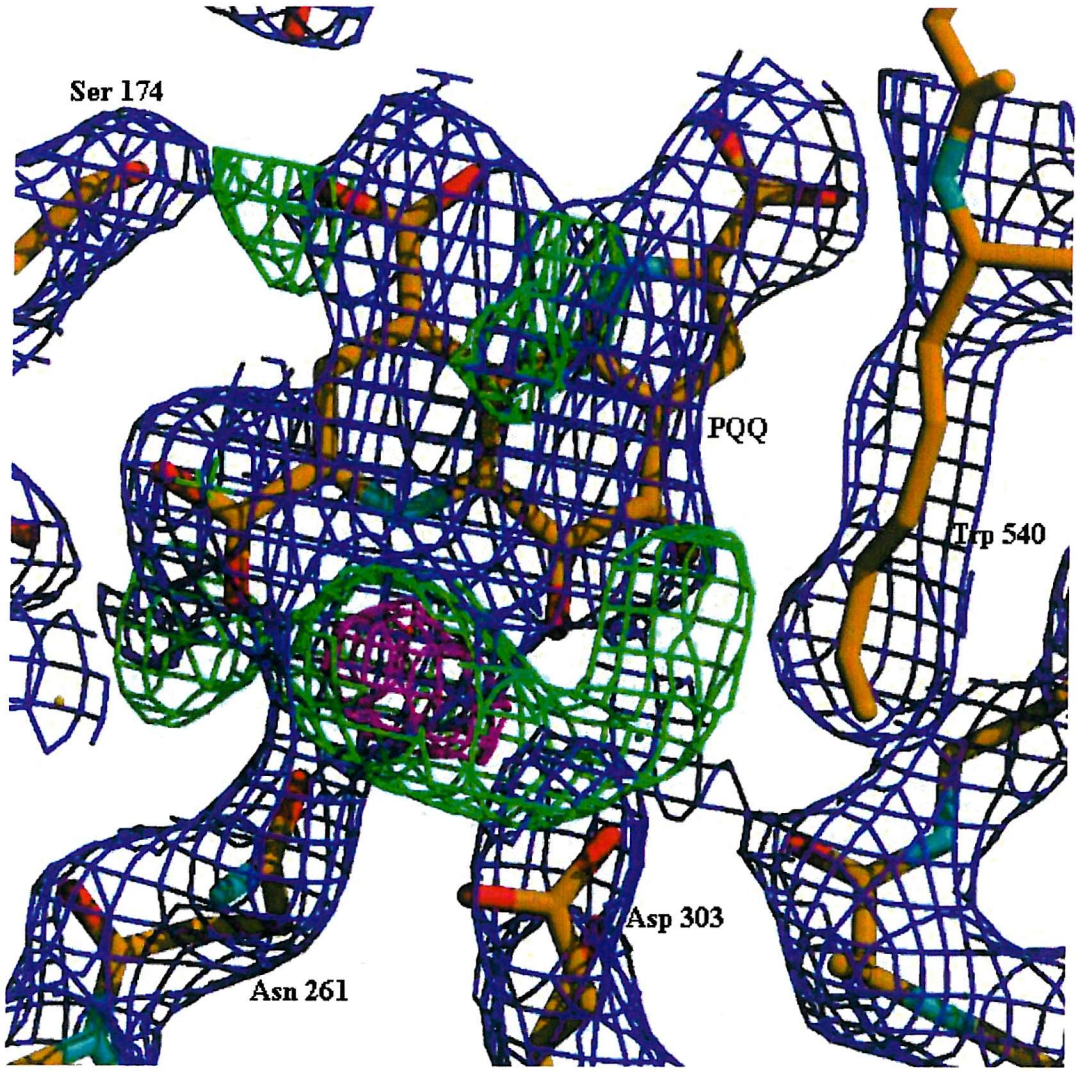


Figure 4.9: The active site of D303E-MDH (XtalView and rendered using Raster3D). Difference electron density for the Ca^{2+} ion is displayed in magenta (contoured at the 4 σ level). The Fo-Fc map contoured at the 2 σ level (displayed in green) reveals extra density at position 303 for the glutamate side chain. The 2Fo-Fc map (displayed in purple) is contoured at the 1 σ level.

PQQ cofactor and the metal ion had been omitted from the search model prior to molecular replacement). This confirmed that the calculated maps were not dominated by model bias and that the molecular replacement solution was correct. In addition, extended side-chain density was observed at residue 303, which was consistent with the D303E mutation (Figure 4.9). During rebuilding, regions of the model in poor agreement with the electron density were adjusted. Furthermore, side chains that lacked electron density were replaced with alanine. At this stage, the electron density for the PQQ was sufficiently interpretable for it to be built into the model. The model was then subjected to several alternating cycles of refinement and model building. The subsequent procedures involved rigid body refinement followed by torsion angle simulated annealing. Throughout refinement NCS restraints were maintained since their release did not lower the free R-factor. In the final stages of model building, the Ca^{2+} ion and the glutamate side chain at position 303 were added.

The final R-factor was 18.7 % (R_{free}=21.7 %) for a model with 20548 protein atoms and consists of 2668 amino acid residues, 4 molecules of PQQ and 4 Ca^{2+} ions. The refinement statistics are listed in Table 4.8. The model was analysed with PROCHECK and displayed acceptable geometry for main chain and side chain parameters. Inspection of the Ramachandran plot indicated that 82.6 % of the amino acid residues are in the 'most favoured' regions, 16.5 % are in 'additionally allowed' regions, 0.2 % are in 'generously allowed' regions and 0.7 % in 'disallowed' regions. The disallowed residues in each αβ subunit (Lys-19, Asn-52, Asp-105 and Lys-166) are well defined by electron density and reside in tight turn regions.

Resolution range (Å)	30-3
R-factor (%)	18.7
R-free (%)	21.7
Number of reflections in working set	45246
Number of reflections in test set	2385
Number of protein atoms	20548
Rms bond length deviation (Å)	0.007
Rms bond angle deviation (°)	1.4

Table 4.8: Refinement statistics for D303E-MDH crystal.

4.6 The Structure of D303E-MDH

4.6.1 Introduction

The superposition of the mutant enzyme structure on that of WT-MDH (for 667 C^{α} atoms) and W3A1-MDH (for 627 C^{α} atoms) yielded an rms deviation of 0.20 Å and 0.69 Å, respectively. The overall structure is identical to that of WT-MDH and W3A1-MDH, with significant differences being presented only in the active site region.

4.6.2 The PQQ Interactions in D303E-MDH

The D303 to E mutation caused no major perturbations within the active site but changes in the hydrogen-bonding network were observed (Table 4.9). The carboxyl group of Asp-303 (Asp-297 in W3A1-MDH) in the wild type enzyme does not interact with the C-5 carbonyl of PQQ (Figure 4.10a). However, the extended side chain of Glu-303 in the mutant structure brings the carboxylate oxygen (OE2) approximately 1.2 Å closer to the C-5 carbonyl of PQQ, implying the formation of a new hydrogen bond (Figure 4.10b). The question arises as to the source of the proton in the new bond.

There is no evidence that Glu-303 is protonated and thus supplies the proton for the new hydrogen bond. However, the hydrophobic environment induced by the neighbouring Trp-265 may increase the pKa of the carboxylic group resulting in the protonation of Glu-303 at higher pH. By contrast, Asp-303 is deprotonated in WT-MDH because both carboxylate oxygen atoms serve as hydrogen bond acceptors (Asp-303 OD2 — Asn-261 ND2 3.2 Å; Asp-303 OD1 — Arg-331 NH2 3.1 Å) (Figure 4.10a). These hydrogen bonding interactions are no longer possible in the mutant structure, which increases the likelihood of Glu-303 being protonated (Figure 4.10b). However, protonation of Glu-303 would be expected to render the enzyme inactive, as it would no longer be able to initiate the reaction by abstracting a proton from methanol.

The mutant enzyme displays some activity although the affinity for methanol has been reduced, leading to further doubt whether Glu-303 is protonated. A more likely explanation is that there is a hydroxyl group on the C-5 of PQQ, and thus the isolated

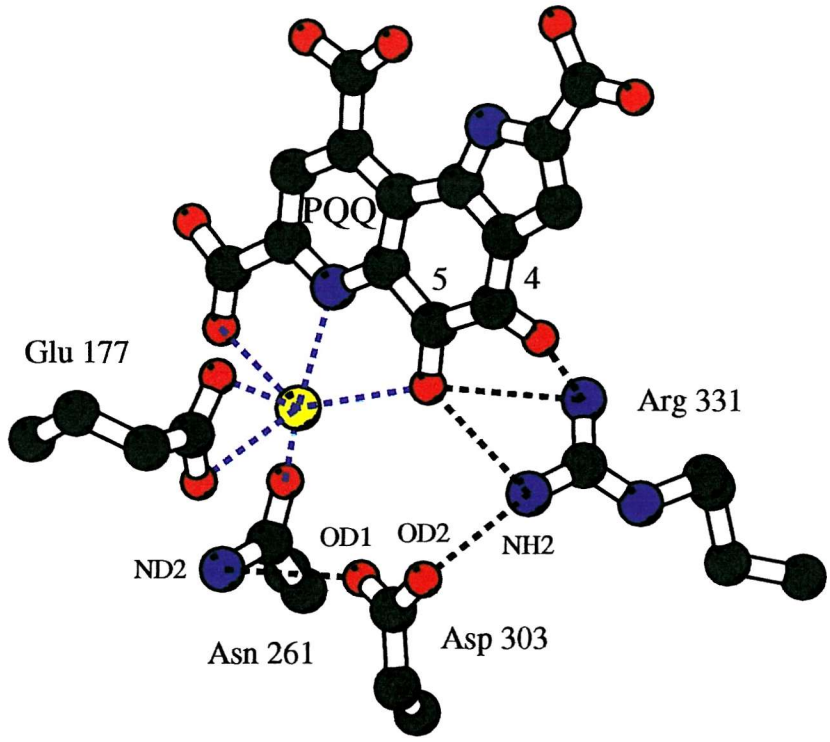


Figure 4.10a: The active site interactions of WT-MDH (BOBSCRIPT). The co-ordination sphere of Ca^{2+} is hexa-coordinate with ligands provided by both PQQ and active site residue side-chains. The proposed nucleophile (Asp-303) does not interact with the Ca^{2+} or the PQQ moiety. Also note that Asp-303 is unprotonated since both carboxylate oxygen atoms serve as hydrogen bond acceptors (forming hydrogen bonds with Asn-261 and Arg-331).

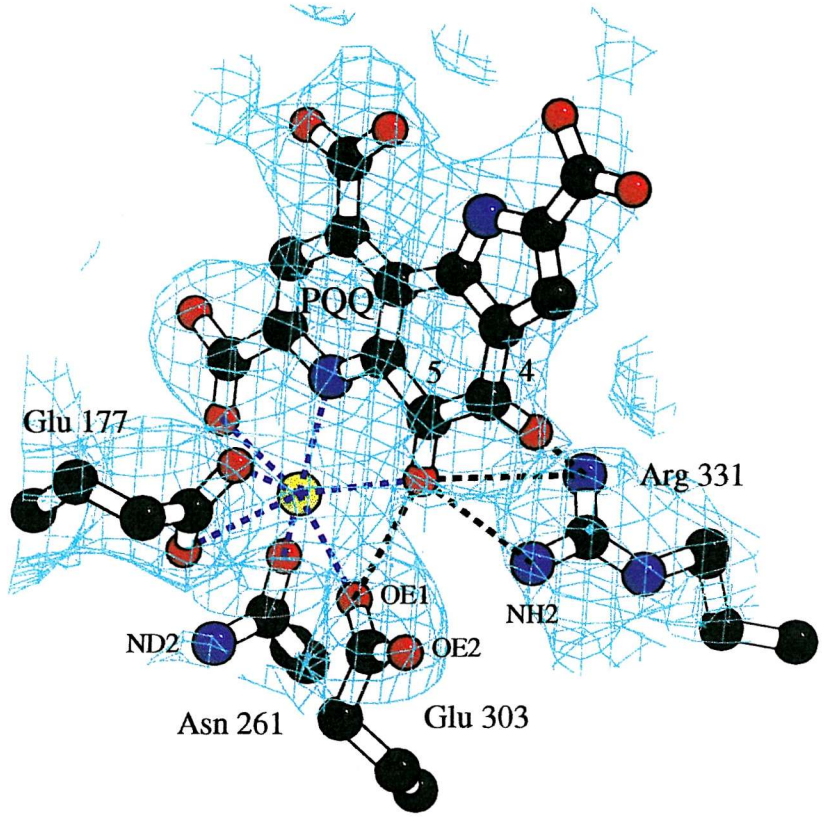


Figure 4.10b: The active site interactions of D303E-MDH (BOBSCRIPT). The co-ordination sphere of Ca^{2+} is hepta-coordinate (the extra ligand being provided by Glu-303). There is also a new hydrogen-bond between Glu-303 and the C-5 carbonyl oxygen of PQQ. The 2Fo-Fc electron density map is contoured at the 1 σ level.

mutant enzyme is in the reduced (or semiquinone) form. This would then coincide with the absorption spectrum obtained for the PQQ prosthetic group in D303E-MDH (Afolabi, 1999). Unfortunately, the resolution of the present structure is not high enough to ascertain the true configuration of the PQQ. The reduced (or semiquinone) form of the cofactor has previously been observed in the high resolution structures of MDH (Ghosh *et al.*, 1995; Xia *et al.*, 1999). In each of these structures the tetrahedral nature of C-4 and C-5 of PQQ is observed; the C-4 carbonyl oxygen atom projects out the plane of the PQQ ring in WT-MDH (Figure 4.10a), whereas it is the C-5 carbonyl oxygen in W3A1-MDH (Figure 4.11a). The reason for this difference is not known.

Atom 1	Atom 2	Distances (Å)		
		WT-MDH	D303E-MDH	W3A1-MDH
PQQ O4	Arg 331 NH1	2.71	3.1	3.11
PQQ O4	Asn 394 ND2	3.16	3.4	3.00
PQQ O5	Arg 331 NH2	2.94	2.9	2.99
PQQ O5	Asp 303 OD2	4.05	-	4.32
PQQ O5	Asp 303 OD1	3.80	-	4.00
PQQ O5	Glu 303 OE2	-	2.8	-
PQQ O5	Glu 303 OE1	-	3.5	-
Asp 303 OD1	Arg 331 NH2	3.06	-	3.11
Glu 303 OE1	Arg 331 NH2	-	3.5	-
Asp 303 OD2	Asn 261 ND2	3.21	-	3.21
Glu 303 OE2	Asn 261 ND2	-	3.8	-

Table 4.9: Comparison of the major active site interactions for WT-MDH, D303E-MDH and W3A1-MDH.

4.6.3 Calcium Co-ordination in D303E-MDH

The identification of the Ca^{2+} binding site in each of the α subunits of D303E-MDH was based on the strong *Fo-Fc* density peak of 5 σ and the oxygen rich environment. The co-ordination sphere of Ca^{2+} in D303E-MDH is similar to that in WT-MDH, with the PQQ moiety (O-5, O-7 and N6 ring atoms) and the active site protein atoms (OD1 of Asn-261 and both carboxylate oxygens of Glu-177) serving as metal ligands. The metal ligand bonds are listed in Table 4.10; none of them deviate significantly from those in WT-MDH (2.4-2.8 Å) and W3A1-MDH (2.5-2.9 Å). The co-ordination sphere of Ca^{2+} within

D303E-MDH has changed from hexacoordinate to the more usual heptacoordinate, although the conventional pentagonal bipyramid geometry is not observed (Figure 4.10b). The extra Ca^{2+} ligand is provided by Glu-303, since the introduction of the longer side chain shifts the carboxylate oxygen (OE2) to within 2.4 Å from the Ca^{2+} ion. In contrast, the carboxylate oxygen (OD2) of Asp-303 in WT-MDH does not interact with the Ca^{2+}, since it is at a distance of 3.5 Å (Figure 4.10a). Accompanying the change in co-ordination is the shifting of the Ca^{2+} ion ~0.5 Å in a direction perpendicular to the plane of the PQQ and towards the disulphide bridge.

The proposed structural role of the Ca^{2+} atom in WT-MDH is in maintaining the active configuration of the PQQ molecule (Anthony, 1996). It has also been proposed that Ca^{2+} plays a catalytic role by serving as a Lewis acid by co-ordinating to the C-5 carbonyl oxygen and facilitating the polarisation of the electrophilic C-5 for attack on the C-5 atom by an oxyanion or hydride (Anthony, 1996). It is conceivable that the positive charge of Ca^{2+} in D303E-MDH is dissipated through its interaction with Glu-303 and its electron withdrawing effect diminished. Consequently, the C-5 atom of PQQ would be less electrophilic, hence less likely to be attacked by an oxyanion or hydride.

Ca^{2+} ligand interactions	Distances (Å)		
	WT-MDH	D303E-MDH	W3A1-MDH
PQQ O5	2.77	2.6	2.76
PQQ N6	2.45	2.4	2.89
PQQ O7A	2.52	2.6	2.72
Asn 261 OD1	2.80	2.7	2.50
Glu 177 OE1	2.58	2.7	2.94
Glu 177 OE2	2.36	2.4	2.68
Asp 303 OD2	3.55	-	3.11
Glu 303 OE2	-	2.4	-

Table 4.10: Summary of Ca^{2+} ligand interactions in WT-MDH, D303E-MDH and W3A1-MDH.

4.6.4 The Modelling of Methanol in D303E-MDH

The decreased affinity of D303E-MDH for methanol indicates that differences exist between D303E-MDH and other MDH structures, concerning the relative position of Glu-303 with respect to the putative methanol binding site. In order to quantify these differences, the D303E-MDH was also compared to that of W3A1-MDH, for which a high resolution structure is available with bound methanol. The methanol binding site is proximal to both the carboxylate group of Asp-297 (equivalent to Asp-303 in *M. extorquens*) and the C-5 carbonyl oxygen of PQQ (Xia *et al.*, 1999). Based on this structure, the authors favoured the hemiketal intermediate mechanism since the methyl group of methanol is 3.9 Å from the C-5 of PQQ, making a direct hydride transfer unlikely (Figure 4.11a).

By superimposing the structurally conserved active site residues from D303E-MDH, WT-MDH and W3A1-MDH, the methanol binding site in the latter structure was compared (Table 4.11). These studies indicated no obvious steric reasons that would occlude methanol from the equivalent sites of the D303E mutant and WT-MDH; the hydroxyl oxygen of methanol is within hydrogen bonding distance (2.9 Å) of one of the carboxylate oxygens (OE1) of Glu-303 (3.1 Å in WT-MDH) (Figure 4.11b and c). These observations are consistent with the previously reported W3A1-MDH structure (Xia *et al.*, 1999), since the methanol hydroxyl donates a proton for hydrogen bond formation with the carboxylate oxygen atom of Asp-297 (at 3.0 Å) (Figure 4.11a). Consequently, the explanation for the altered activity of D303E-MDH with respect to methanol is less straightforward. One possibility is that Glu-303, through its interaction with the C-5 carbonyl oxygen and the Ca^{2+} ion, would be an unsuitable candidate for abstracting the initial proton from methanol. This interaction, in part, may explain why the D303E mutant exhibits an 80,000-fold decrease in affinity towards methanol, with respect to WT-MDH (Afolabi, 1999).

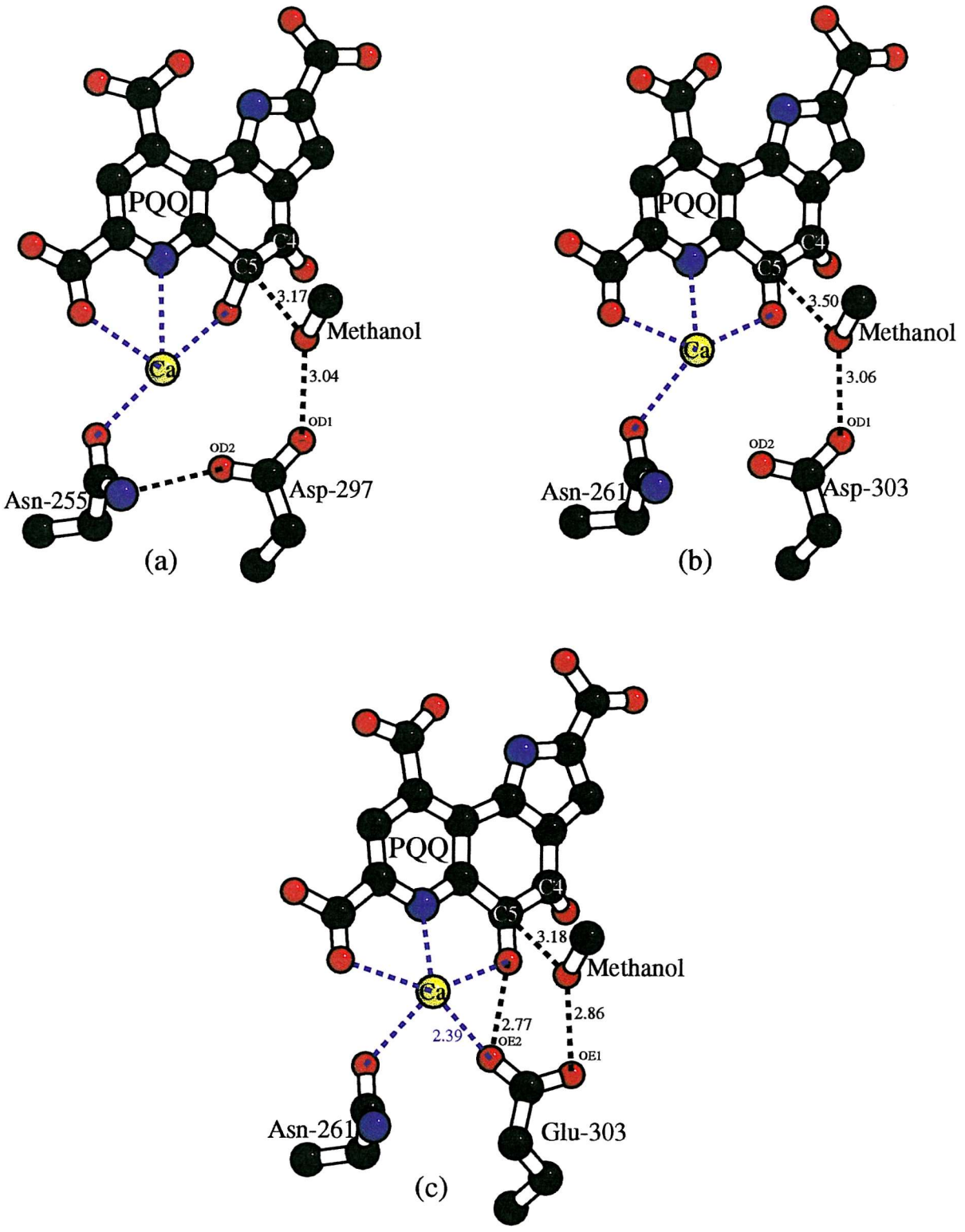


Figure 4.11: The active site interactions of W3A1-MDH (a) and the modelling of methanol in WT-MDH (b) and D303E-MDH (c) (BOBSCRIPT). In the active site of W3A1-MDH the C-5 of PQQ is closely associated with the hydroxyl oxygen (3.2 Å) but not the methyl group (3.9 Å) of methanol. This favours the hemiketal intermediate mechanism rather than a direct hydride ion transfer. The modelling of methanol in WT-MDH yields similar observations to W3A1-MDH. The modelling of methanol in D303E-MDH suggests that the carboxylate oxygen (OE1) of the Glu-303 side-chain forms a hydrogen bond to methanol. There is no steric hindrance to methanol binding at the active site of D303E-MDH.

Atom 1	Atom 2	Distances (Å)		
		WT-MDH	D303E-MDH	W3A1-MDH
Asp 303 OD2	MeOH (O)	4.08	-	4.32
Asp 303 OD1	MeOH (O)	3.06	-	3.04
Glu 303 OE2	MeOH (O)	-	3.2	-
Glu 303 OE1	MeOH (O)	-	2.9	-
PQQ C5	MeOH (CH_3)	4.14	3.9	3.75
PQQ C5	MeOH (O)	3.50	3.2	3.17

Table 4.11: Comparison of methanol interactions for WT-MDH, D303E-MDH and W3A1-MDH.

4.6.5 Conclusions

The kinetic properties of the mutant enzyme (reduced affinity for methanol and the decreased stability of the transition state complex) are consistent with the unique features observed within the active site of D303E-MDH including the slight movement of the Ca^{2+} ion, the change in the co-ordination of Ca^{2+} via Glu-303 (from hexacoordinate to heptacoordinate) and the new hydrogen bond formed between Glu-303 and the C-5 carbonyl oxygen of PQQ. It is plausible that the new interactions involving Glu-303, reduces its likelihood of abstracting a proton from methanol.

The correct position of the methanol binding site is now in doubt following a recent reinterpretation of the refined W3A1-MDH structure, which indicated that the methanol was absent (Zheng *et al.*, 2001). It is now conceivable that the methanol binding site is indeed closely associated with the active site base (Asp-303) and possibly the Ca^{2+} ion (increasing the number of ligands from 6 to 7) in the wild type enzyme. In which case, this would account for the decreased affinity for the substrate in the mutant structure, since methanol finds it more difficult to bind to the Ca^{2+} because the co-ordination geometry of Ca^{2+} has changed (through its interaction with Glu-303). The poorer binding of methanol to the Ca^{2+} in the mutant enzyme would also explain the decrease in binding energy of methanol. Alternatively, it could be that the introduction of the larger side chain sterically hinders methanol binding.

The reinterpretation of the W3A1-MDH structure has also demonstrated that the C-5 of PQQ is present in the tetrahedral conformation (carrying a hydroxyl group and a proton), which is not an intermediate in the hemiketal mechanism (Zheng *et al.*, 2001). This is consistent with the observation that Glu-303 forms a hydrogen bond with the C-5 carbonyl oxygen in D303E-MDH, which would only be possible if a hydroxyl group is present in this position. Moreover, it would also explain why the substrate modelling studies indicated that methanol could bind at the active site, although a longer side chain had been introduced. Based on their new observations, Zheng *et al* propose that MDH follows the alternative direct hydride transfer mechanism, which is similar to the mechanism for the related soluble glucose dehydrogenase published recently by Dijkstra and co-workers (Oubrie and Dijkstra, 2000). This revised mechanistic scheme may explain why no spectroscopic intermediates were detected during the course of the reaction catalysed by D303E-MDH (Afolabi, 1999).

The tautomerisation process in the direct hydride transfer is likely to involve the same base that catalyses the initial proton abstraction (Anthony *et al.*, 1994; Anthony, 1996). In such circumstances, the carboxylate oxygen of Asp-303 (or Glu-303) must be within 3 Å of the donor (C-5 of PQQ) and acceptor (O-4 of PQQ) atoms. However, the distances recorded indicate that this is not the case. The nearby Arg-331 is an alternative contender for this role, with its NH2 being within hydrogen bonding distance of O-4 of PQQ, although this is improbable because the pKa for the arginine side chain is 12.48. One can also postulate that the arm of the Asp-303 (or Glu-303) side chain is sufficiently flexible for it to carry out both functions; the initial proton abstraction and facilitation of the subsequent tautomerisation of the reduced PQQ.

4.7 Further Work

Further insight into the reaction catalysed by MDH can only be achieved following the identification of the substrate binding site. Currently, crystals of WT-MDH and D303E-MDH are being soaked and co-crystallised with methanol and cinnamyl alcohol. Data from these crystals will hopefully provide an ES complex allowing us to distinguish between the two mechanisms proposed for MDH.

4.8 References

Afolabi, P (1999) Ph. D. Thesis, University of Southampton

Anderson, D.J., Morris, C.J., Nunn, D.N., Anthony, C. and Lidstrom, M.E. (1990) *Gene* **90 (1)** 173-176

Anthony, C. and Zatman, L.J. (1964) *Biochem. J.* **92** 614-621

Anthony, C. and Zatman, L.J. (1965) *Biochem. J.* **96** 808-812

Anthony, C. (1975) *Biochem. J.* **146** 189-198

Anthony, C. (1982) In *The Biochemistry of Methylotrophs*, Academic Press, London

Anthony, C. (1986) *Adv. Microbial. Physiol.* **27** 113-210

Anthony, C. (1992a) *Int. J. Biochem.* **24** 29-39

Anthony, C. (1992b) *Biochim. Biophys. Acta* **1099** 1-15

Anthony, C., Ghosh, M. and Blake, C.C.F. (1994) *Biochem. J.* **304** 665-674

Anthony, C. (1996) *Biochem. J.* **320** 697-711

Anthony, C. (1997) In *Comprehensive biological catalysis; a mechanistic reference.* Academic Press, London

Anthony, C. and Ghosh, M. (1998) *Progs. Biophys. Mol. Biol.* **69** 1-21

Bolbot, J.A. and Anthony, C. (1980) J. *Gen. Microbial.* **120** 245-254

Cozier, G.E., Giles, I.G. and Anthony, C. (1995) *Biochem. J.* **307** 375-379

Cozier, G.E. and Anthony, C. (1995) *Biochem. J.* **312** 679-685

Dales, S.L. and Anthony, C. (1995) *Biochem. J.* **312** 261-265

Day , D.J. and Anthony, C. (1990) *Methods Enzymol.* **188** 210-216

Duine, J.A., Frank, J. and van Zeeland, J.K. (1979) *FEBS Lett.* **108** 443-446

Duine, J.A., Frank, J. and Verwiel, P.E.J. (1980) *Eur. J. Biochem.* **108** 187-192

Faber, H.R., Groom, C.R., Baker, H.M., Morgan, W.T., Smith, A. and Baker, E.N. (1995) *Structure* **3** 551-559

Frank, J., Dijskstra, M., Duine, A.J. and Balny, C. (1988) *Eur. J. Biochem.* **174** 331-338

Frank, J., van Krimpen, S.H., Verwiel, P.E.J., Jongejan, J.A., Mulder, A.C. and Duine, J.A. (1989) *Eur. J. Biochem.* **184** 187-195

Fulop, V., Moir, J.W.B., Ferguson, S.J. and Hajdu, J. (1995) *Cell* **81** 369-377

Ghosh, M., Anthony, C., Harlos, K., Goodwin, M.G. and Blake, C.C.F. (1995) *Structure* **3** 177-187

Goodwin, M.G. and Anthony, C. (1996) *Biochem. J.* **318** 673-679

Harms, N., Devries, G.E., Maurer, K., Hoogendijk, J. and Stouthamer, A.H. (1987) *J. Bacteriol.* **169 (9)** 3969-3975

Harris, T.K. and Davidson, V.L. (1994) *Biochem. J.* **300** 175-182

Imanaga, Y. (1989) in *PQQ and Quinoproteins* p87-96, Eds. Jongejan, J.X. and Duine, J.A., Kluwer Academic Publishers, Dordrecht

Ito, N., Phillips, S.E.V., Yadav, K.D.S. and Knowles, P.F. (1994) *J. Mol. Biol.* **238** 794-814

Janes, S.M., Mu, D., Wemmer, D., Smith, A.J., Kaur, S., Maltby, D., Buringame, A.L. and Klinman, J.P. (1990) *Science* **248** 981-987

Long, A.R. and Anthony, C. (1991) *J. Gen. Microbial.* **137** 2353-2360

Matsushita, K., Shinagawa, E., Inoue, T., Adachi, O. and Ameyama, M. (1986) *FEMS. Microbial. Letts.* **37** 141-144

McIntire, W.S., Wemmer, D.E., Chistoserdov, A. and Lidstrom, M.E. (1991) *Science* **252** 817-824

Mutzel, A. and Görisch, H. (1991) *Agric. Biol. Chem.* **55** 1721-1726

Nunn, D.N., Day, D.J. and Anthony, C. (1989) *Biochem. J.* **260** 857-862

Ohshiro, Y. and Itoh, S. (1993) in *Principles and application of quinoproteins* p309-339, Eds. Davidson, V.L., Marcel Dekker Inc, New York

Oubrie, A. and Dijkstra, B.W. (2000) *Protein Science* **9** 1265-1273

Page, M.D. and Anthony, C. (1986) *J. Gen. Microbial.* **132** 1553-1563

Richardson, I.W. and Anthony, C. (1992) *Biochem. J.* **287** 709-715

Varghese, J.N., Laver, W.G. and Colman, P.M. (1983) *Nature* **303** 35-40

Vellieux, F.M.D., Huitema, F., Groendijk, H., Kalk, K.H., Frank, J., Jongejan, J.A., Duine, J.A., Petratos, K., Drenth, J. and Hol, W.G.J. (1989) *EMBO. J.* **8** 2171-2178

Xia, Z.X., Dai, W.E., Xiong, J., Hao, Z., Davidson, V.L., White, S.A. and Mathews, F.S. (1992) *J. Biol. Chem.* **267** 22289-22297

Xia, Z.X., Dai, W.E., Zhang, Y.F., White, S.A., Boyd, G.D. and Mathews, F.S. (1996) *J. Mol. Biol.* **259** 480-501

Xia, Z.X., Dai, W.E., White, S.A., Boyd, G.D. and Mathews, F.S. (1999) *Biochemistry* **38** 1214-1220

Zheng, Y.J., Xia, Z.K., Chen, Z.W., Mathews, F.S. and Bruice, T.C. (2001) *Proc. Natl. Acad. Sci.* In Press

Chapter 5

Structural Studies of MhpC from *Escherichia coli*

5.1 Introduction to MhpC

5.1.1 The α/β-hydrolase Fold Family

The α/β-hydrolase fold was identified in the early 1990s in five seemingly distinct hydrolases (Ollis *et al.*, 1992), and since then over fifty α/β-hydrolase fold structures have been determined. Despite significant sequence similarity existing between them (Heikinheimo *et al.*, 1999), members of the α/β-hydrolase fold family share a core structure (an α/β sheet, not barrel, of eight β-strands connected by α-helices). Although their catalytic specificities are markedly diverse, members of the family are thought to have similar enzyme mechanisms.

The canonical α/β-hydrolase fold consists of an eight stranded α/β structure, with strand β2 anti-parallel to the other strands (Figure 5.1). The strand connectivity conforms to a +1, +2, -1x, +2x, +1x, +1x +1x rule (strand order 1 2 3 4 5 6 7 8). All members to date possess a catalytic triad; a nucleophile, an acid and a histidine, in the same order in the amino acid sequence. In contrast to other protease families, the nucleophile is not always the same in each enzyme; it can be a serine (pancreatic lipase), cysteine (dienelactone hydrolase) or aspartate (haloperoxidase). The nucleophile is located in a tight turn following strand β5 called the 'nucleophile elbow', which is highly conserved and used as the motif for superimposing the α/β-hydrolase structures.

Members of the α/β hydrolase superfamily can be classified into three distinct subgroups (siblings, cousins and distant relatives) depending on their sequence identity with one another. The α/β hydrolase fold proteins can be distinguished from others on the basis of four unique characteristics. Firstly, the catalytic triad has the sequence order nucleophile-acid-histidine, with the nuleophile on the tight turn (nucleophile elbow) following strand β5. Secondly, the nucleophile elbow possesses a conserved sequence motif GlyXaaNucXaaGly (where Xaa is a small residue and Nuc is the nucleophile). Thirdly, the structure comprises at least five strands, the first of which is the canonical strand β3. Finally, the extended loops following strands β7 and β8 contain the catalytic acid and histidine residues (Figure 5.1), which form hydrogen-bonding interactions at the active site (Heikinheimo *et al.*, 1999).

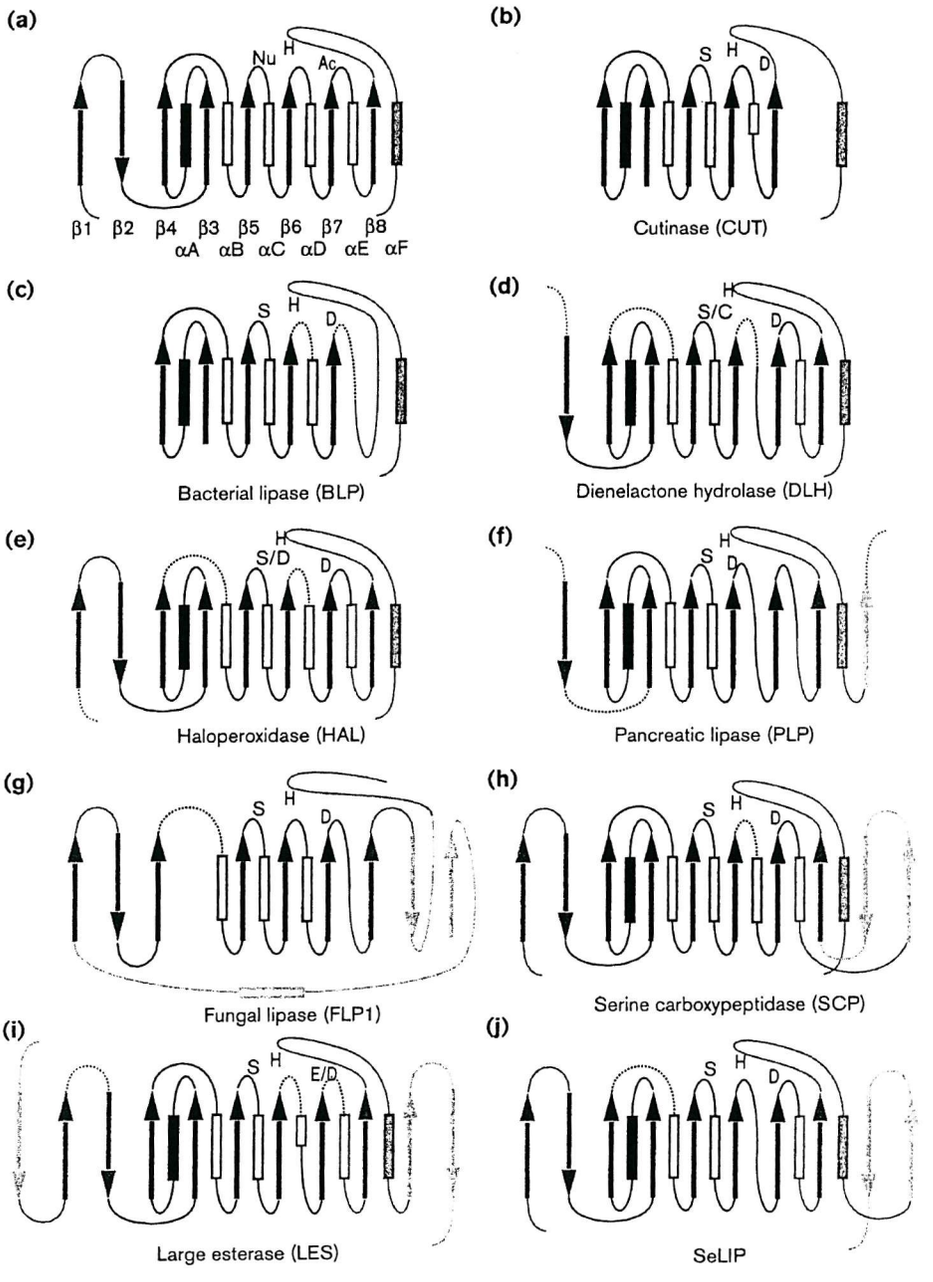


Figure 5.1: Topological diagrams of the α/β-hydrolase fold for various proteins. Strands are shown by arrows and helices by rectangles. The catalytic triad is shown in single letter code for each member (Heikinheimo *et al.*, 1999).

5.1.2 Aromatic Compounds

An increasing number of man-made aromatic compounds (such as pesticides, detergents, oils and solvents) have built up in the environment since the beginning of the industrial period. The removal of these compounds is dependent upon microbial enzymes that can degrade aromatic compounds; for example, the microbial BphD, XylF and TodF pathways are responsible for the degradation of biphenyl, xylene and toluene, respectively (Dagley, 1986). However, a considerable number of these compounds are either degraded slowly or not at all, resulting in their continuous presence in the environment.

Presently, bioremediation is emerging as the most promising technology for the removal of these man-made or xenobiotic compounds from the environment. The technique involves treating the polluted environment with microbes that have the ability to transform pollutants to non-toxic products. Intensive characterisation of enzymes involved in microbial catabolic pathways is essential to our understanding of the biotransformation process. This will provide a basis for improving existing pathways and for engineering novel degradative capabilities, contributing to the removal of recalcitrant and toxic compounds from the environment (Seah *et al.*, 1998).

5.1.3 Degradation of 3-hydroxy phenylpropionic acid

Bacteria utilise a diverse range of enzymes in the initial stages in the degradation of aromatic compounds. However, the aerobic microbial catabolic pathways are inclined to converge on a small number of critical dihydroxylated intermediates, which are subsequently cleaved and ultimately broken down to Krebs cycle intermediates by enzymes from either ortho or meta cleavage pathways (van de Meer *et al.*, 1992). Many pathways have been shown to exist in *E. coli* for the degradation of aromatic compounds. In particular, the meta cleavage pathway for the degradation of 3-hydroxy phenylpropionic acid (Figure 5.2) has been especially well characterised (Burlingame, 1986).

MhpA catalyses the hydroxylation of 3-hydroxy phenylpropionic acid to produce 2,3-dihydroxyphenylpropionic acid (DHP), which is then cleaved by non-haem-iron dependent DHP 1,2 dioxygenase (MhpB) to give the ring fission product (RFP), 2-hydroxy-6-keto-nona-2,4 diene-1,9-dicarboxylate. The RFP is subsequently hydrolysed by 2-hydroxy-6-

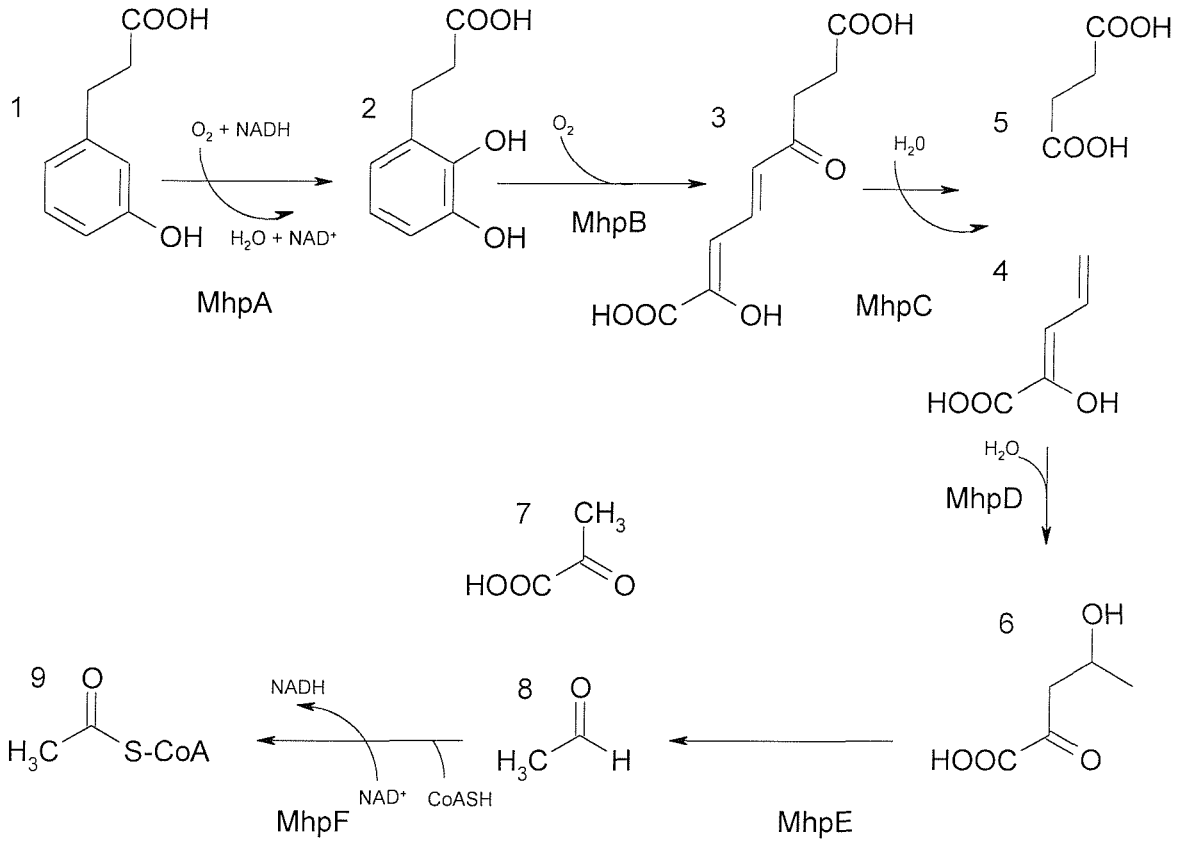


Figure 5.2: The *E. coli* meta cleavage pathway for the degradation of the aromatic compound 3-hydroxy phenylpropionic acid (Ferrández *et al.*, 1997). MhpA catalyses the hydroxylation of 3-hydroxy phenylpropionic acid (1) to produce 2,3-dihydoxy-phenylpropionic acid (DHP, 2). DHP is cleaved by a non-heme-iron dependent DHP 1,2 dioxygenase (MhpB) to give the ring fission product (RFP), 2-hydoxy-6-keto-nona-2,4 diene-1,9-dicarboxylic acid (3). The subsequent hydrolysis of the RFP by 2-hydoxy-6-keto-nona-2,4 diene-1,9-dioic acid-5,6-hydrolase (MhpC) yields 2-hydroxy-penta-2,4-dienoic acid (HPD, 4) and succinic acid (5). The former is hydrated to produce 4-hydroxy-2-ketopentanoic acid (HKP, 6), in the presence of MhpD. In the penultimate step of the pathway, MhpE catalyses the conversion of HKP to pyruvic acid (7) and acetaldehyde (8). Finally, acetaldehyde is converted to acetyl-CoA (9) in the presence of acetaldehyde dehydrogenase (MhpF).

keto-nona-2,4 diene-1,9-dioic acid-5,6-hydrolase (MhpC), a C-C hydrolyse enzyme, yielding succinic acid and 2-hydroxypenta-2,4-dienoic acid (HPD). In the presence of MhpD, HPD is subsequently hydrated to form 4-hydroxy-2-ketopentanoic acid (HKP). The penultimate step of the pathway involves a reverse aldol cleavage reaction catalysed by MhpE, resulting in the conversion of HKP to pyruvic acid and acetaldehyde. In the final step of the pathway acetaldehyde dehydrogenase (MhpF) catalyses the conversion of acetaldehyde to acetyl-CoA.

5.1.4 Functional Studies of *E. coli* MhpC

E. coli MhpC is a C-C hydrolase involved in a key step in the extradiol cleavage pathway for bacterial degradation of 3-hydroxy phenylpropionic acid. MhpC catalyses the hydrolytic cleavage of a C-C bond in the RFP, 2-hydroxy-6-keto-nona-2,4-diene-1,9 dioic acid, to give HPD and succinic acid (Figure 5.3). MhpC is a homodimeric enzyme (M_r 58 kDa) and its activity can be monitored either by following substrate degradation at 394 nm or by measuring the release of the succinate product (Lam and Bugg, 1997). The enzyme has a K_m of 2.1 μm for its natural substrate, displays high specificity towards the propionate side chain and requires no cofactors or metal ions for activity. Although group targeting reagents such as sodium iodoacetate and succinic anhydride have no affect on MhpC activity, its hydrolytic activity is reduced 68 % by *p*-hydroxymercuribenzoate (PHMB), implying the presence of a modifiable cysteine residue (Lam and Bugg, 1997).

5.1.5 Mechanism of MhpC

The reaction catalysed by MhpC has been well studied, providing much insight into its catalytic mechanism. The observation of an MhpC catalysed exchange of ^2H, at C-5 of the substrate (Lam and Bugg, 1997) and pre-steady state kinetic measurements (Henderson and Bugg, 1997) indicate the existence of a keto intermediate. It is conceivable that the α,β-unsaturated ketone functional group present in the keto intermediate serves as an electron sink, allowing C-C fragmentation to take place following an enzyme activated nucleophilic attack on the C-6 carbonyl.

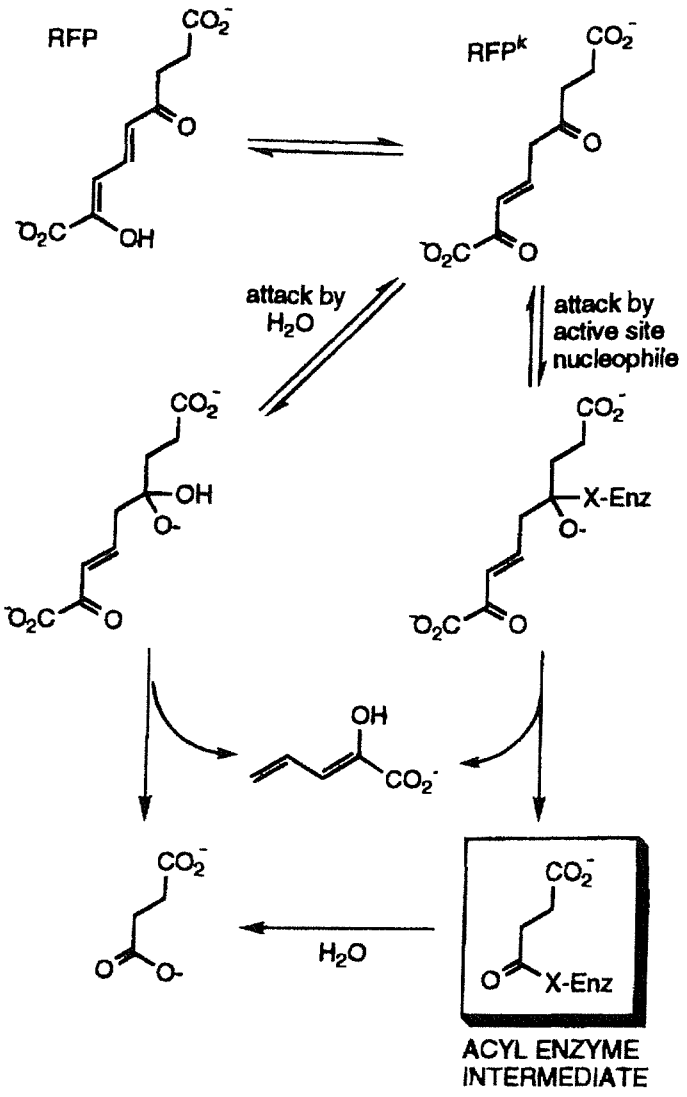


Figure 5.3: The proposed mechanism for the MhpC-catalysed reaction. The initial step involves ketonisation of the substrate (RFP) to give a keto intermediate (RFP^k). The subsequent C-C fragmentation step occurs upon attack of either water or an active site nucleophile at C-6 of RFP^k (Henderson and Bugg, 1997).

The identity of the nucleophile is uncertain since it could be either a water molecule or an active site residue (Figure 5.3) (Lam and Bugg, 1994; Lam and Bugg, 1997; Henderson and Bugg, 1997). In the latter case, formation of an acyl enzyme intermediate would be anticipated, similar to that seen in the mechanism of the serine proteases (Chapus *et al.*, 1976). Based on sequence alignments and chemical modification studies on other members of the C-C hydrolase family, a conserved serine residue has been implicated as the nucleophile, functioning in a similar manner as in the serine proteases. However, in contrast to serine proteases which accumulate an acyl enzyme intermediate at pH<5, MhpC is catalytically active at pH 4.0 (k_{cat} $1.0s^{-1}$).

Attempts to trap the putative acyl enzyme intermediate have been unsuccessful with *p*-nitrophenyl succinate (Lam and Bugg, 1997). More recently, radiochemical trapping methods using ^{14}C-labelled substrate produced no evidence to corroborate the existence of an acyl enzyme intermediate (Fleming *et al.*, 2000). In contrast, experiments with ^{18}O-labelled water gave 100 % incorporation of one ^{18}O atom into succinic acid, and 5-6 % incorporation of two ^{18}O atoms, consistent with the reversible formation of a gem-diol intermediate by attack of water (Fleming *et al.*, 2000). Furthermore, incorporation of ^{18}O atom into the carbonyl group of a non-cleavable substrate analogue 4-keto-nona-1,9-dioic acid (KNDA) has been observed in the presence of MhpC, implying that the enzyme-catalysed attack of water occurs at the C-4 ketone. Together, these results are consistent with a catalytic mechanism involving base catalysed attack of water, rather than a nucleophilic attack by a serine residue.

5.1.6 Evidence that MhpC is a Serine Hydrolase

Sequence alignment of C-C hydrolases from aromatic catabolic pathways reveals the existence of a conserved sequence motif (GlyXaaSerXaaGly), suggesting that the conserved serine residue may serve as a nucleophile in a manner similar to members of the serine hydrolase family. Furthermore, several invariant residues including two histidines, two aspartates and a glutamate have also been identified. Site directed mutagenesis studies of 2-hydroxy-6-oxo-(phenyl/chlorophenyl)hexa-2,4-dienoic acid (HOPD), a C-C hydrolase enzyme from the biphenyl/polychlorinated degradation pathway, indicate that HOPD is indeed a serine hydrolase; replacement of the conserved active site serine to

alanine (S112A) results in a total loss of enzyme activity (Ahmad *et al.*, 1995). Chemical modification studies of various C-C hydrolases add further support to the hypothesis that MhpC is a serine hydrolase. The activity of HOPD was reduced in the presence of phenylmethylsulfonyl fluoride (PMSF), a classic inhibitor of serine hydrolases (Ahmad *et al.*, 1995). XylF was significantly inactivated in the presence of serine directed reagents (diisopropyl fluorophosphate (DPF) and PMSF). The activity of XylF was also reduced by histidine targeting reagents such as diethylpyrocarbonate (DEPC) and tosyl-phenylalanine chloromethyl ketone (TPCK) (Diaz and Timmis, 1995). In contrast, no significant inactivation was observed for MhpC in the presence of 5 mM DEPC, although total activity was abolished at higher concentrations (200 mM) (Lam and Bugg, 1997).

The three-dimensional structure of biphenyl hydrolase (BphD; 2-hydroxyl-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase) from *Rhodococcus sp* RHA1 has recently been determined by MIR methods (Nandhagopal *et al.*, 1997). This is the first X-ray structure to be determined for a member of the C-C hydrolase family. The secondary structural topology of BphD (Figure 5.4) is consistent with that found in proteins of the α/β-hydrolase fold family. BphD contains a conserved serine residue and is inactivated by the serine protease inhibitor, PMSF, making it highly plausible that all the C-C hydrolase enzymes are serine hydrolases (Ollis *et al.*, 1992).

5.1.7 Research Aims

The aim of this work was to provide a greater insight into the catalytic mechanism of MhpC by determining its X-ray structure. The structure will be useful for determining the precise role of the putative active site serine residue. Furthermore, the structure of MhpC complexed with the time dependent inhibitor KNDA was determined to reveal the precise arrangement of active site residues and enzyme-substrate binding interactions.

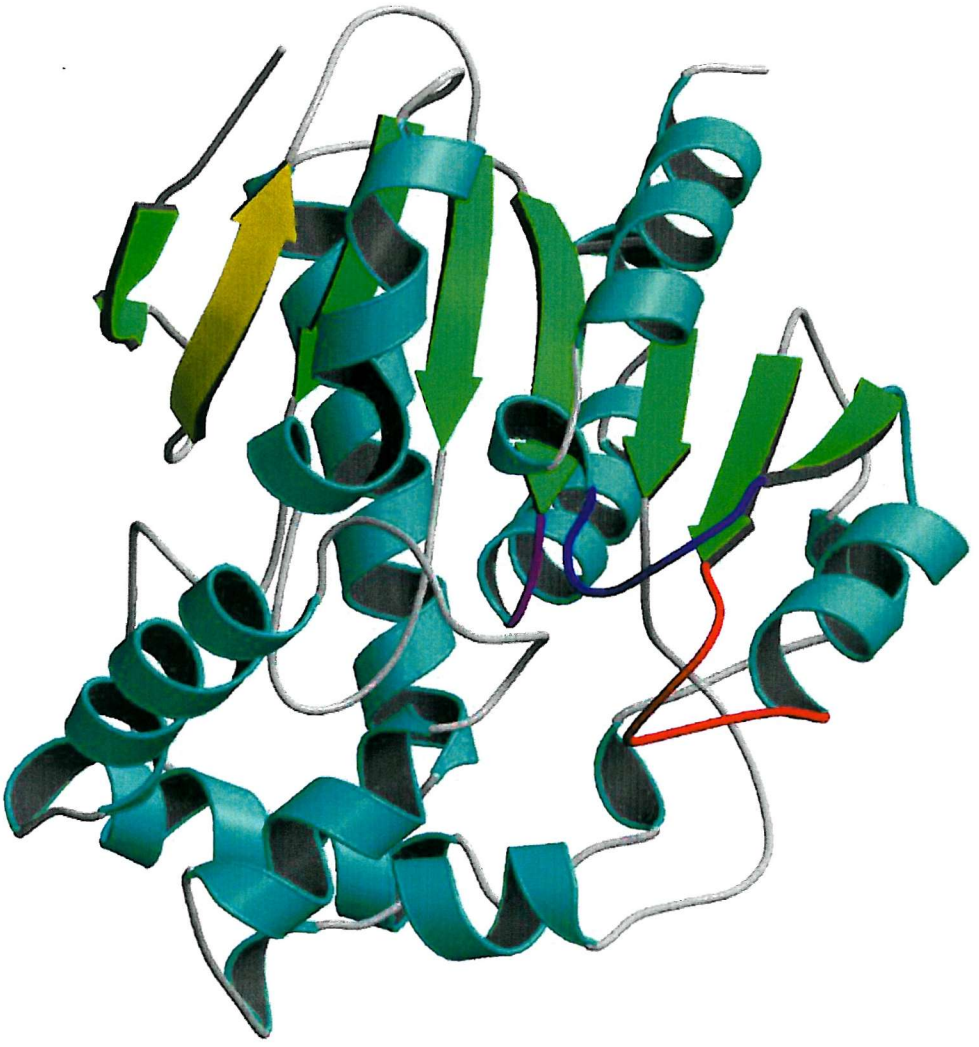


Figure 5.4: The structure of biphenyl hydrolase (BphD) from *Rhodococcus sp* RHA1 (Nandhagopal *et al.*, 1997) (MOLSCRIPT and rendered using Raster3D). The secondary structural topology is consistent with BphD being a member of the α/β-hydrolase fold family. Note the major features of the α/β-hydrolase fold; seven of the 8 strands run parallel (depicted in green) with strand β2 anti-parallel to the others (depicted in yellow). The nucleophilic elbow structure (containing the nucleophile) is depicted in purple, the loop following strand β7 (containing the acid) is depicted in red and the loop following strand β8 (containing the histidine) is depicted in blue.

5.2 Expression, Purification and Crystallisation of MhpC

E.coli MhpC was expressed and purified as described previously by Tom Robertson in the group of Professor Tim Bugg (Lam and Bugg 1997). Crystallisation conditions were determined using the hanging drop vapour diffusion method with Hampton crystal screens. Small crystals were obtained with 12 different conditions (Table 5.1). The most promising condition for crystal growth (hit 12 in Table 5.1) was optimised and refined; equilibrating against 0.2 M ammonium phosphate, 47 % MPD and 0.1 M Tris-HCl pH 8.5 at room temperature produced larger crystals. However, several problems were associated with this crystal form. Firstly, although a shower of crystals was present in each hanging drop, they were too small. Secondly, the presence of a skin probably due to denatured protein, made crystal mounting difficult. Finally, the crystals appeared to be twinned.

Hit	Salt	Buffer (pH)	Precipitant
1	0.02 M Calcium chloride	0.1 M Sodium acetate(4.6)	30 % MPD
2	None	None	0.4 M Ammonium dihydrogen phosphate
3	None	0.1 M Sodium citrate (5.6)	1.0 M Ammonium dihydrogen phosphate
4	0.02 M Calcium chloride	0.1 M Sodium acetate(4.6)	20 % 2-propanol
5	0.2 M Calcium chloride	0.1 M HEPES (7.5)	28 % PEG 400
6	None	0.1 M Sodium HEPES (7.5)	0.8 M Sodium phosphate 0.8 M Potassium phosphate
7	0.05 M Potassium phosphate	None	20 % PEG 8000
8	0.2 M Calcium acetate	0.1 M Sodium Cacodylate (6.5)	18 % PEG 8000
9	None	0.1 M Tris-HCl (8.5)	2.0 M Ammonium dihydrogen phosphate
10	0.1 M Sodium phosphate 0.1 M Potassium phosphate	0.1 M MES (6.5)	2.0 M Sodium chloride
11	0.1 M Calcium chloride	0.1 M Tris-HCl (8.5)	25 % tert-butanol
12	0.2 M Ammonium phosphate	0.1 M Tris-HCl (8.5)	50 % MPD

Table 5.1: The 12 original hits obtained from Hampton crystal screens using the hanging drop vapour diffusion method.

The size of the crystals was improved (Figure 5.5a) in subsequent crystallisation trials by decreasing the precipitant concentration from 47 to 44 % and, hence, reducing the number of nucleation sites in each hanging drop. Attempts to overcome the twinning problem included adding KNDA, a time dependent inhibitor for MhpC, to the crystallisation experiments. It was presumed that the non-cleavable substrate analogue would tightly bind to the enzyme and stabilise the complex. Unfortunately, data collection from one of these crystals at room temperature had to be halted after 10 images due to excessive radiation damage to the crystal. Consequently, subsequent data collection was carried out at cryogenic temperatures (100 K), with MPD in the mother liquor (44 %) serving as a cryo-protectant. The presence of two crystal lattices in the diffraction pattern together with large differences in the unit cell dimensions caused great difficulties in processing data and confirmed that the crystals were twinned.

To investigate the reason behind the twinning phenomenon, the protein was analysed by mass spectroscopy to detect any microheterogeneity. The mass spectrogram indicated the existence of three MhpC species (Figure 5.6), possibly arising from heterogeneity in N-terminal methionine cleavage during the enzyme preparation. Further trials for growing higher quality crystals involved using factorial screens to incorporate elements of all 12 original hits. In particular, different metal ions, salts and buffers were varied with hit 12 (see Table 5.1). One of the factorial screens provided large single crystals of high quality (Figure 5.5b). These crystals, which grew after 72 h at room temperature, were obtained by mixing MhpC (25mg/ml + 1mM KNDA in 10mM Tris-HCl pH 7.5) in a 1:1 ratio with reservoir solution (44 % MPD, 0.2 M ammonium chloride, 0.1 M sodium cacodylate pH 6.5 and 0.2 M calcium acetate).

Figure 5.5a: MhpC crystals grown by reducing the precipitant concentration following the optimisation of hit 12. The crystals appeared after 2 days at room temperature using the hanging drop method. The drop was set up by mixing MhpC (25 mg/ml + 1 mM 4-keto-nona-1,9-dioic acid (KNDA) in 10 mM Tris-HCl pH 7.5) with equal volumes of reservoir solution (44 % MPD, 0.2 M ammonium phosphate and 0.1 M Tris-HCl pH 8.5).

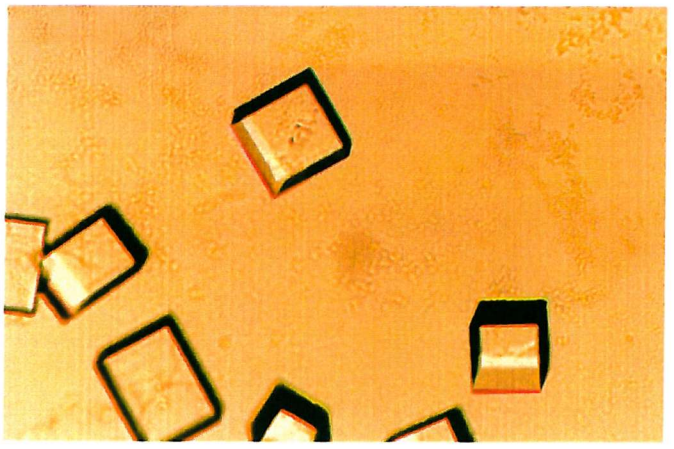


Figure 5.5b: Large single MhpC crystals grown by incorporating elements of all 12 original hits obtained from factorial screens. The crystals appeared after 4 days at room temperature. The drop was set up by mixing MhpC (25 mg/ml + 1 mM 4-keto-nona-1,9-dioic acid (KNDA) in 10 mM Tris-HCl pH 7.5) with an equal volume of reservoir solution (44 % MPD, 0.2 M ammonium chloride, 0.1 M sodium cacodylate pH 6.5 and 0.2 M calcium acetate).

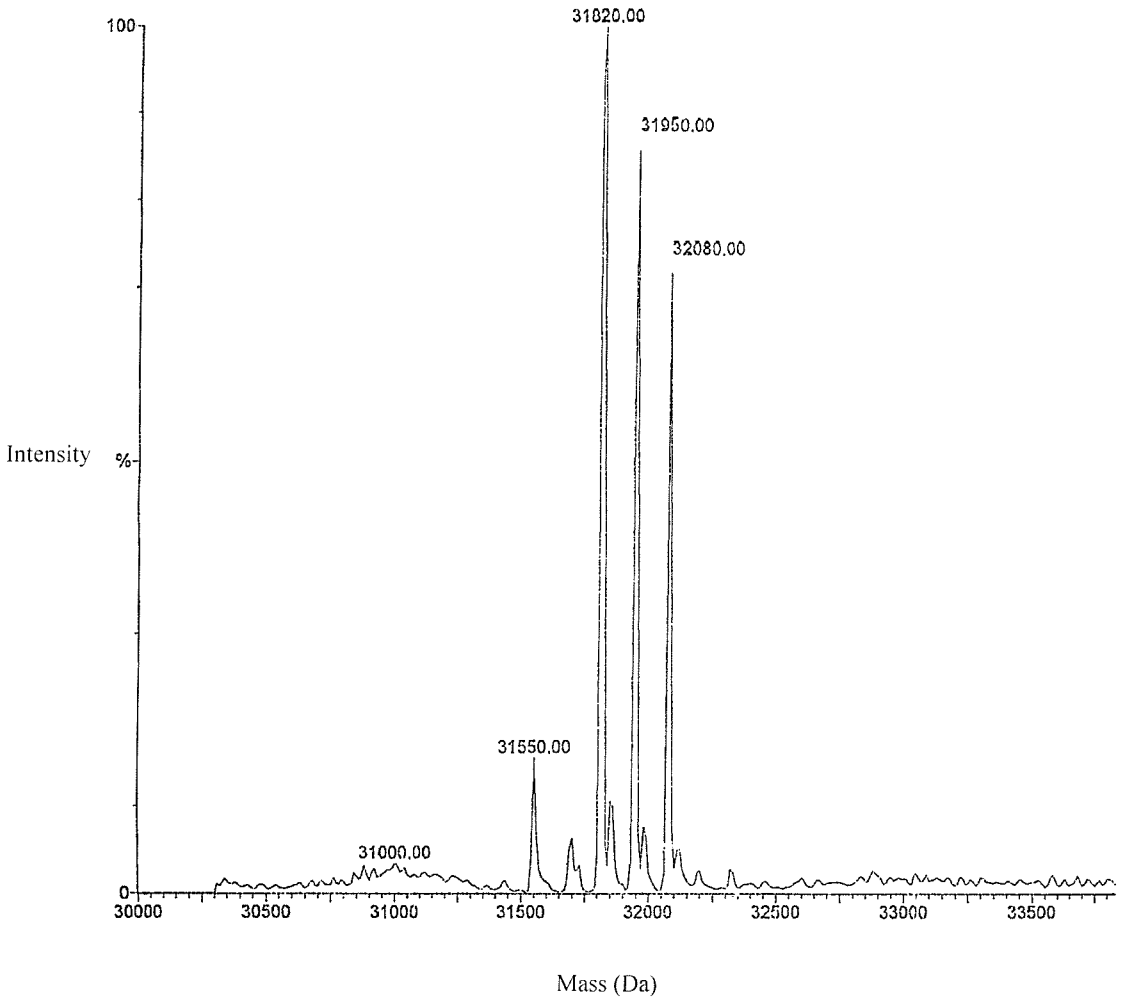


Figure 5.6: Mass spectrogram of purified MhpC used in the crystallisation trials. The 3 peaks (31820, 31950 and 32080 Da) separated by 150 Da, correspond to 3 MhpC species in the sample. This is most likely due to heterogenity in N-terminal methionine cleavage during enzyme preparation.

5.3 Data Collection and Processing of MhpC and MhpC-KNDA Complex

A native dataset were collected on beamline BW7B at DESY (Hamburg), from a single MhpC crystal using a 30 cm MAR-Research image plate. The native crystal form was vitrified in a stream of cold nitrogen gas at 100 K and diffracted X-rays to a resolution of 2.3 Å. The crystal was found to belong to the orthorhombic space group $P2_12_12$, as judged by systematically absent reflections (Figure 5.7), with unit cell dimensions of a=144.99 Å, b=144.43 Å and c=62.35 Å. A total of 90 1°-oscillation frames, were measured with an exposure time of 60 s per frame and a crystal to detector distance of 390 mm.

Data from a MhpC crystal co-crystallised with KNDA were collected at 100 K on ESRF beamline ID14-EH3 (Grenoble, France) using a Mar CCD detector. The crystal diffracted to a resolution of 2.5 Å and crystallised in the same space group as the native crystal form, with unit cell dimensions a=145.45 Å, b=144.97 Å and c=62.99 Å. A total of 180 images, each covering an oscillation angle of 0.5°, were collected with an exposure time of 120 s per frame and a crystal to detector distance of 135 mm. Intensity data from both crystals were processed with the program MOSFLM. Using programs of the CCP4 suite, the sorted data were merged and scaled and intensities converted to structure factors. The relevant data processing statistics are listed in Table 5.2.

Data Processing Parameters	Native MhpC	MhpC-KNDA Complex
Wavelength (λ)	0.834	1.00
Total number of reflections	207600	167771
Number of unique reflections	56646	46119
Resolution (Å)	2.3	2.5
Completeness (%)	99.4	98.3
R_{merge} (%) (outer resolution shell)	6.5 (27.3)	9.9 (23.8)
Multiplicity	3.7	3.6
Average I/σ(I) (outer resolution shell)	5.6 (2.5)	4.3 (2.5)

Table 5.2: Data processing statistics for MhpC and MhpC-KNDA crystals.

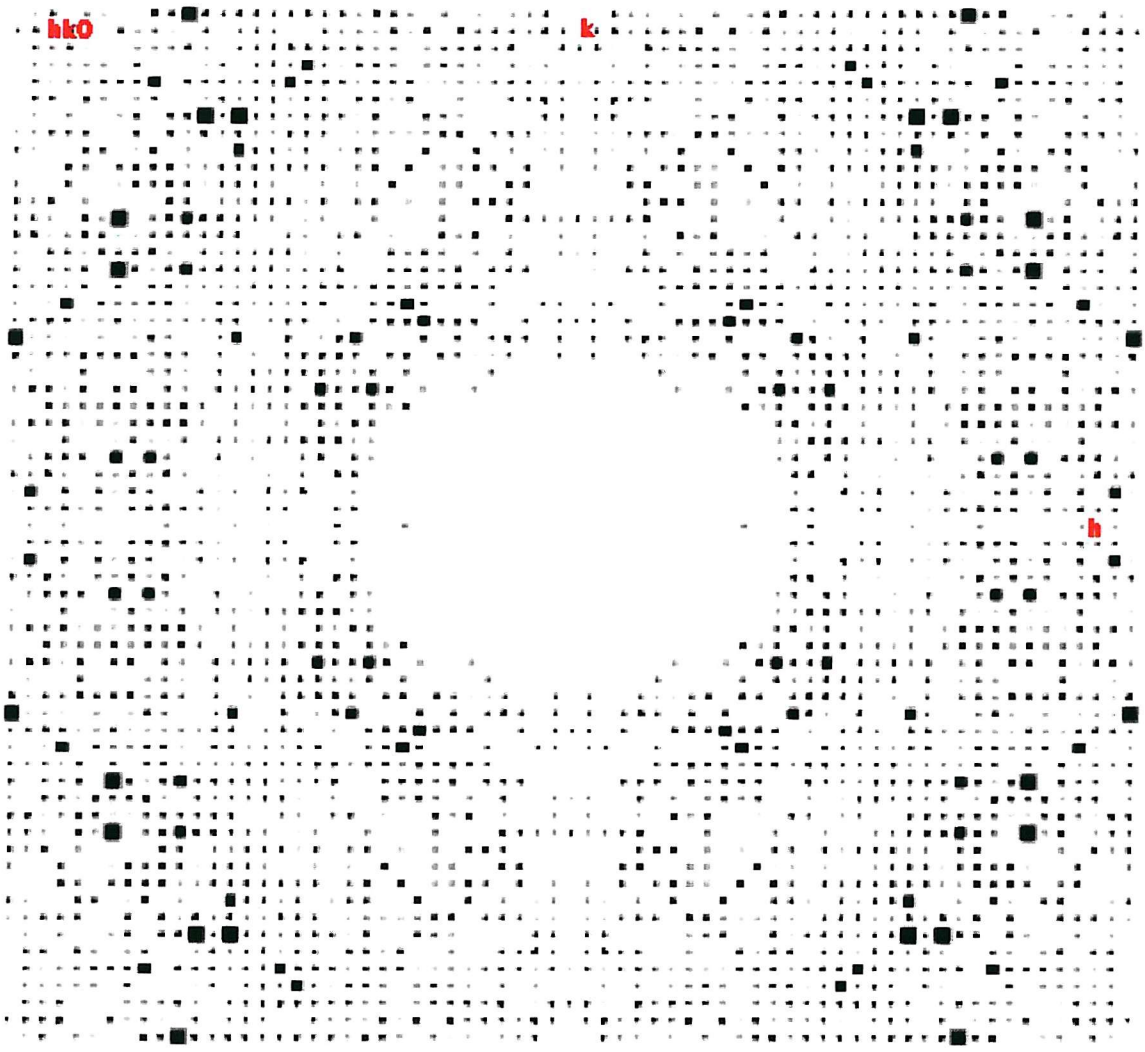


Figure 5.7: A pseudo precision picture displaying a 2_1 screw axis along zones H00 (h=2n) and 0K0 (k=2n) for native MhpC when processed in P222 (HKLVIEW).

5.4 Multiple Isomorphous Replacement Studies on MhpC

The lack of a phasing model with a high sequence similarity necessitated the use of MIR methods for the structure determination of MhpC. Initial attempts to produce isomorphous heavy atom derivatives involved using a range of thiol directed reagents (MhpC contains 3 cysteine residues), since previous chemical modification studies using PHMB had identified a modifiable cysteine residue in the protein (Lam and Bugg 1997). These trials demonstrated differential activity between the protein crystal and the heavy atom reagent; some eliminated diffraction or increased disorder, while others produced no discernible damage to the crystal. To improve the strategy for producing a heavy atom derivative, MhpC crystals were soaked in a diverse range of heavy atom reagents (Table 5.3). Unfortunately none of the heavy atom derivatives yielded a resolvable Patterson map.

Heavy atom	Concentration (mM)	Soak time (hours)	Comments
Thiomersal	0.5	11	Diffraction to 2.9 Å. No derivative
	0.5	83	Diffraction to 3.4 Å. No derivative
p-Chloromercuriphenyl sulphonic acid	1.0	11	Diffraction to 2.9 Å. No derivative
Mercury chloride	10.0	132	No diffraction
	5.0	132	No diffraction
	2.5	132	No diffraction
	0.1	3	Diffraction to 3.5 Å. No derivative
Ethylmercury phosphate	1.0	5	Diffraction to 2.9 Å. No derivative
Ethylmercury chloride	1.0	11	Diffraction to 2.9 Å. No derivative
Platinum chloride	5.0	13	No diffraction. Crystal disordered
Potassium tetracyanoplatinate	10.0	2	Crystal cracked
	5.0	10	Diffraction to 3.2 Å. No derivative
Sodium gold cyanide	25.0	18	Diffraction to 3.2 Å. No derivative
	1 0.0	2	Diffraction to 3.3 Å. No derivative
	5.0	40	Diffraction to 2.9 Å. No derivative
Uranyl acetate	3.0	10	Diffraction to 3.2 Å. No derivative
Uranyl sulphate	3.0	2	Diffraction to 3.4 Å. No derivative
Lead cyanide	5.0	1	Crystal cracked

Table 5.3: Heavy atoms used in trials to produce a derivative MhpC crystal for MIR phasing.

5.5 Molecular Replacement Studies of MhpC

5.5.1 Introduction

Concurrent with heavy atom soaking trials, attempts were also made to overcome the phase problem by using the molecular replacement technique. Based on secondary structure predictions using the program PredictProtein (Rost and Sander, 1993), an α/β hydrolase fold was identified for MhpC. Following the submission of the MhpC sequence to SWISS-MODEL, a putative structural homologue of MhpC (chloroperoxidase) was identified. The homologue shared 21 % sequence identity to MhpC and possessed the α/β hydrolase fold (Hofmann *et al.*, 1998). However, no significant rotation function solutions were found. Other unproductive phasing models included prolyl aminopeptidase from *Serratia marcescens* (Yoshimoto *et al.*, 1999) and epoxide hydrolase from *Agrobacterium radiobacter* (Nardini *et al.*, 1999). More recently, the structural co-ordinates for BphD from *Rhodococcus sp* RHA1 (28 % sequence identity to MhpC), a member of the C-C hydrolase family, became available (Nandhagopal *et al.*, 1997). Although BphD was the most promising phasing model to date, no significant molecular replacement solutions were obtained. The reason for the lack of success in the cross rotation searches were investigated in greater detail (see section 5.7).

5.5.2 Structure Determination of MhpC by MAD

The structure of MhpC was eventually solved by Dr Graham Dunn (at Southampton) using the MAD method. MhpC were expressed in a methionine auxotroph strain of *E. coli* in the presence of seleno-methionine (Dunn, 2000). The expressed protein was purified and crystallised in monoclinic space group $P2_1$ (cell dimensions a=62.22 Å, b=154.72 Å, c=78.94 Å, β=89.7°) with 4 MhpC molecules per asymmetric unit. MAD data were collected at the ESRF beamline BM14 (Grenoble, France) by Dr Jon Cooper. The seleno-methionine crystal diffracted X-rays to a resolution of 2.1 Å and phases were determined by locating 32 selenium atoms by direct methods using the 'Shake and Bake' program (Weeks and Miller, 1999).

5.6 Molecular Replacement and Refinement of MhpC-KNDA Complex

5.6.1 Introduction

The structure of MhpC co-crystallised with the time dependent inhibitor, 4-keto-nona-1,9-dioic acid (KNDA), was determined to provide insight into the precise arrangement of active site residues and enzyme-substrate binding interactions. The MhpC-KNDA complex structure was analysed to establish whether MhpC possesses a catalytic triad similar to the serine hydrolase family.

5.6.2 Cross Rotation Search of the MhpC-KNDA Complex

Assuming the MhpC-KNDA crystal contains 4 MhpC molecules (116 kDa) per asymmetric unit, a solvent content of 57 % was calculated (V_m=2.86 $Å^3$/Da), which is similar to that of the seleno-methionine crystal form. Initial phases for the MhpC-KNDA structure were obtained by molecular replacement using a partially refined monomer of MhpC solved by MAD methods. The cross rotation calculations were performed with the program MOLREP using reflections in the 10 to 3.0 Å resolution range. Different integration radii values (20-30 Å) were chosen during successive cross rotation searches; the rotation function yielded two significant peaks at 15.3 and 14.8 σ, with a radius of integration of 30 Å (Table 5.4).

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	27.39	90.00	2.61	15.34
2	20.40	90.00	2.71	14.78
3	156.20	62.32	122.52	5.17
4	157.80	72.52	146.08	5.14
5	7.55	90.00	3.04	4.45

Table 5.4: The 5 highest peaks following cross rotation calculations performed with MOLREP using data between 10 and 3 Å.

5.6.3 Translation Search for MhpC-KNDA Complex

Since only two rotation function solutions (Peaks 1 and 2 from Table 5.4) were determined, it was possible that the asymmetric unit of MhpC-KNDA crystal contained only two molecules as opposed to the four predicted by the solvent content calculations. However, translation searches calculated in space group $P2_12_12$ yielded 2 large translation function peaks for each rotation function solution (Tables 5.5 and 5.6), confirming that the asymmetric unit consisted of 4 MhpC molecules with two pairs being in the same orientation. The position of the first MhpC monomer was fixed (peak 1 of rotation function and peak 1 of translation function) and the translation search repeated in order to determine the position of each of the other three monomers with respect to the same origin (Tables 5.7, 5.8 and 5.9).

Peak Number	X	Y	Z	Dens/σ
1	0.204	0.302	0.069	11.93
2	0.187	0.569	0.047	10.89
3	0.032	0.302	0.069	7.86
4	0.189	0.327	0.052	7.64
5	0.487	0.303	0.068	7.01

Table 5.5: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the first rotation function solution (α=27.39°, β=90.00° and γ=2.61°).

Peak Number	X	Y	Z	Dens/σ
1	0.117	0.308	0.044	10.86
2	0.133	0.089	0.061	9.82
3	0.237	0.309	0.044	6.84
4	0.375	0.447	0.358	5.92
5	0.150	0.089	0.063	5.90

Table 5.6: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the second rotation function solution (α=20.40°, β=90.00° and γ=2.71°).

Peak Number	X	Y	Z	Dens/σ
1	0.117	0.808	0.043	22.49
2	0.663	0.089	0.063	22.33
3	0.633	0.537	0.063	9.44
4	0.633	0.573	0.048	8.20
5	0.668	0.087	0.067	7.96

Table 5.7: The 5 highest non-crystallographic translation function peaks calculated in space group $P2_12_12$ after fixing the position of the first MhpC molecule.

Peak Number	X	Y	Z	Dens/σ
1	0.687	0.570	0.048	36.99
2	0.688	0.084	0.066	15.07
3	0.687	0.029	0.047	13.10
4	0.720	0.034	0.091	9.75
5	0.688	0.528	0.069	9.74

Table 5.8: The 5 highest non-crystallographic translation function peaks calculated in space group $P2_12_12$ after fixing the position of two MhpC molecules.

Peak Number	X	Y	Z	Dens/σ
1	0.633	0.088	0.060	40.26
2	0.633	0.036	0.046	10.10
3	0.657	0.061	0.058	8.40
4	0.602	0.078	0.021	8.24
5	0.599	0.086	0.070	7.78

Table 5.9: The 5 highest non-crystallographic translation function peaks calculated in space group $P2_12_12$ after fixing the position of three MhpC molecules.

5.6.4 Refinement of the MhpC-KNDA Structure

Visualisation of the molecular replacement solution using MOLPACK confirmed that the crystal packing was feasible. Refinement of the molecular model was performed with CNS, using 45731 reflections within a resolution range of 20 to 2.3 Å. A randomly selected subset of reflections (5 % of data, 2265 reflections) were excluded and used for R_{free}

calculations. NCS restraints were imposed on the four monomers in the initial round of refinement; the protocol consisted of rigid body refinement, simulated annealing (torsion angle) and grouped B-factor refinement. At this stage the R-factor had been reduced to 27.1 % (R_{free}=28.8 %). The refined model was used to calculate sigmaA weighted electron density maps that were subjected to four-fold NCS averaging. Examination of the averaged maps in QUANTA revealed no extra density for the KNDA molecule.

5.6.5 Conclusions

The MhpC structure possesses the α/β-hydrolase fold (a central α/β-sheet composed of 8 strands and connected by α-helices), as do the serine hydrolase enzymes. Primary sequence alignment studies of MhpC have identified three highly conserved residues that together may form a catalytic triad (Ser-110, Asp-235 and His-263). Preliminary analysis of the active site reveals that Ser-110 is located on the nucleophilic elbow following strand 5, whereas Asp-235 and His-263 are located on loop regions following strands 7 and 8. If the enzyme utilises a catalytic triad then one would expect a hydrogen-bonding interaction between Ser-110 and His-263. However, this interaction is not observed since the hydroxyl group of Ser-110 swings away from His-263 (3.9 Å). Interestingly, Asp-235 forms a hydrogen bond with His-263 (2.6 Å).

Unfortunately, detailed examination of the active site reveals no difference electron density that resembles the KNDA moiety. There is however density present (contoured at the 3 σ level) close to Ser-110 that could presumably be a water molecule, although the shape is not spherical (Figure 5.8). It is conceivable that an enzyme-catalysed attack of water occurs following activation either by Ser-110 or Ser-40 (located opposite to Ser-110 at a distance of 3.9 Å). These observations would confirm the results obtained in Professor Tim Bugg's laboratory, whereby evidence for a water molecule participating in the reaction has been demonstrated using ^{18}O-labelled water (Fleming *et al.*, 2000). However, by refining the structure with the water molecule added, the subsequent *2Fo-Fc* maps provided inconclusive evidence for the presence of a water molecule; a water molecule positioned in between Ser-110 and Ser-40 becomes closely associated with either residue (2.0 Å). Alternatively, the density may arise from a Ca^{2+} ion, since the crystals were grown in the presence of 200 mM calcium acetate.

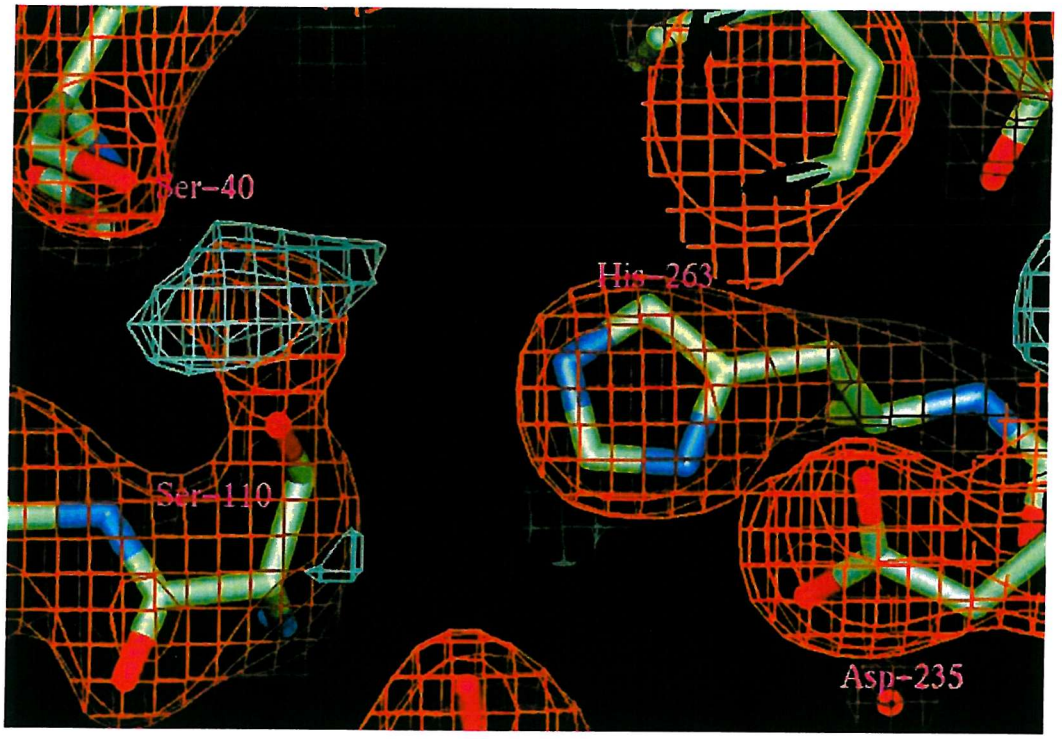


Figure 5.8: The active site of the MhpC-KNDA complex (QUANTA). There is no Fo-Fc electron density that resembles the 4-keto-nona-1,9-dioic acid (KNDA) molecule. The proposed catalytic triad (Ser-110, Asp-235 and His-263) is displayed; the hydroxyl group of Ser-110 swings away from His-263 (3.9 Å), whereas Asp-235 forms a hydrogen bond with His-263 (2.6 Å). The Fo-Fc electron density (contoured at the 3 σ level depicted in green) close to Ser-110 could presumably be a water molecule that serves as a nucleophile. Alternatively, the enzyme-catalysed attack of water occurs following activation by Ser-40 (located opposite to Ser-110 at a distance of 3.9 Å). The 2Fo-Fc map is contoured at the 1.3 σ level (depicted in orange).

5.6.6 Further Work

It is clear that more information regarding the mechanism of MhpC will only be available upon the determination of the structure of an enzyme-inhibitor complex. Initial crystallisation trials with KNDA (present at 1mM) have been unsuccessful and therefore need to be repeated at higher concentrations. Furthermore, several other substrate analogues (ie laevulinic acid) are available and may soak into the crystals with greater ease. Concomitant with the soaking studies, attempts will be made to co-crystallise with the substrate analogues.

5.7 Retrospective Analysis of MhpC Structure Determination

5.7.1 Introduction

The MhpC project was initiated when no structural information regarding the family of C-C hydrolase enzymes was available. Initial attempts at solving the phase problem involved a combination of MIR and molecular replacement techniques. Unfortunately, all attempts to obtain heavy atom derivatives failed, largely as result of damage caused to the crystals and lack of heavy atom binding. Search models for molecular replacement calculations were selected on the basis that they possessed the α/β-hydrolase fold (the same fold that had been predicted for MhpC) and had significant sequence identity to MhpC (ie 20 %). However, although high structural similarity between MhpC and search models existed, the use of the molecular replacement method in the structure determination of MhpC proved unsuccessful.

The structure of MhpC was eventually solved by MAD methods with a different crystal form ($P2_1$). In retrospect, it would be very instructive to determine why the molecular replacement method failed to yield any solution. Before this analysis could take place, the molecular replacement solution for native MhpC in the orthorhombic crystal form was determined. These solutions were subsequently used as a control and compared with solutions obtained using different search models and molecular replacement packages.

5.7.2 Molecular Replacement for the Orthorhombic Crystal Form

Assuming there are 4 MhpC molecules in the asymmetric unit, a solvent content of 55 % was calculated for this crystal form. The molecular replacement solution for this crystal form was obtained with MOLREP using a single MhpC molecule solved by MAD methods as the search model. The rotation function yielded two significant peaks at 13.3 and 12.8 σ (with the next peak being at 4.7 σ), with a radius of integration of 25 Å (Table 5.10). Translation searches calculated in space group $P2_12_12$ yielded 2 large translation function peaks for each rotation function solution (Tables 5.11 and 5.12), verifying that the asymmetric unit does indeed consist of 4 MhpC molecules, with two pairs being in the same orientation.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	152.10	90.00	182.47	13.34
2	160.11	90.00	183.08	12.83
3	158.70	75.26	145.64	4.69
4	155.64	61.82	121.67	4.44
5	8.06	90.00	4.51	4.41

Table 5.10: The 5 highest cross rotation peaks calculated performed with MOLREP. Data between 14.5 and 3 Å were used with a radius of integration of 25 Å.

Peak Number	X	Y	Z	Dens/σ
1	0.291	0.300	0.425	12.37
2	0.308	0.069	0.448	10.72
3	0.291	0.300	0.004	6.93
4	0.291	0.269	0.427	6.80
5	0.291	0.392	0.424	6.39

Table 5.11: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the first rotation function solution (α=152.10°, β=90.00° and γ=182.47°).

Peak Number	X	Y	Z	Dens/σ
1	0.387	0.306	0.459	11.32
2	0.371	0.088	0.442	9.59
3	0.387	0.306	0.083	6.42
4	0.069	0.307	0.456	6.12
5	0.268	0.307	0.458	6.02

Table 5.12: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the second rotation function solution (α=160.11°, β=90.00° and γ=183.08°).

5.7.3 Structure Determination of MhpC using Selected Search Models

5.7.3.1 Introduction

After eliminating possible phasing models with low sequence homology to MhpC, there are three possible structural homologues remaining to be considered in greater detail. These are biphenyl hydrolase (BphD) from *Rhodococcus sp* RHA1 (Nandhagopal *et al.*, 1997), chloroperoxidase from *Streptomyces lividans* (Hofmann *et al.*, 1998) and epoxide hydrolase from *Agrobacterium radiobacter* (Nardini *et al.*, 1999). The sequence identity between MhpC and the search models used in the molecular replacement calculations is relatively low (Figure 5.9). The relevant characteristics for each search model used are listed in Table 5.13. The high structural homology between MhpC and the search models is readily apparent when superimposing the structures (C-alpha atoms) on MhpC (Figures 5.10, 5.11 and 5.12) using TURBO-FRODO (Bio-Graphics, Marseille).

Search Model	PDB code	Resolution (Å)	R-factor	Sequence identity	Sequence similarity	rms deviation
Biphenyl hydrolase	1C4X	2.4	17.8 %	28 %	47 %	0.63 Å
Chloroperoxidase	1A88	1.9	16.0 %	21 %	32 %	0.71 Å
Epoxide hydrolase	1EHY	2.1	19.0 %	19 %	33 %	0.81 Å

Table 5.13: Relevant characteristics of the search models used in molecular replacement.

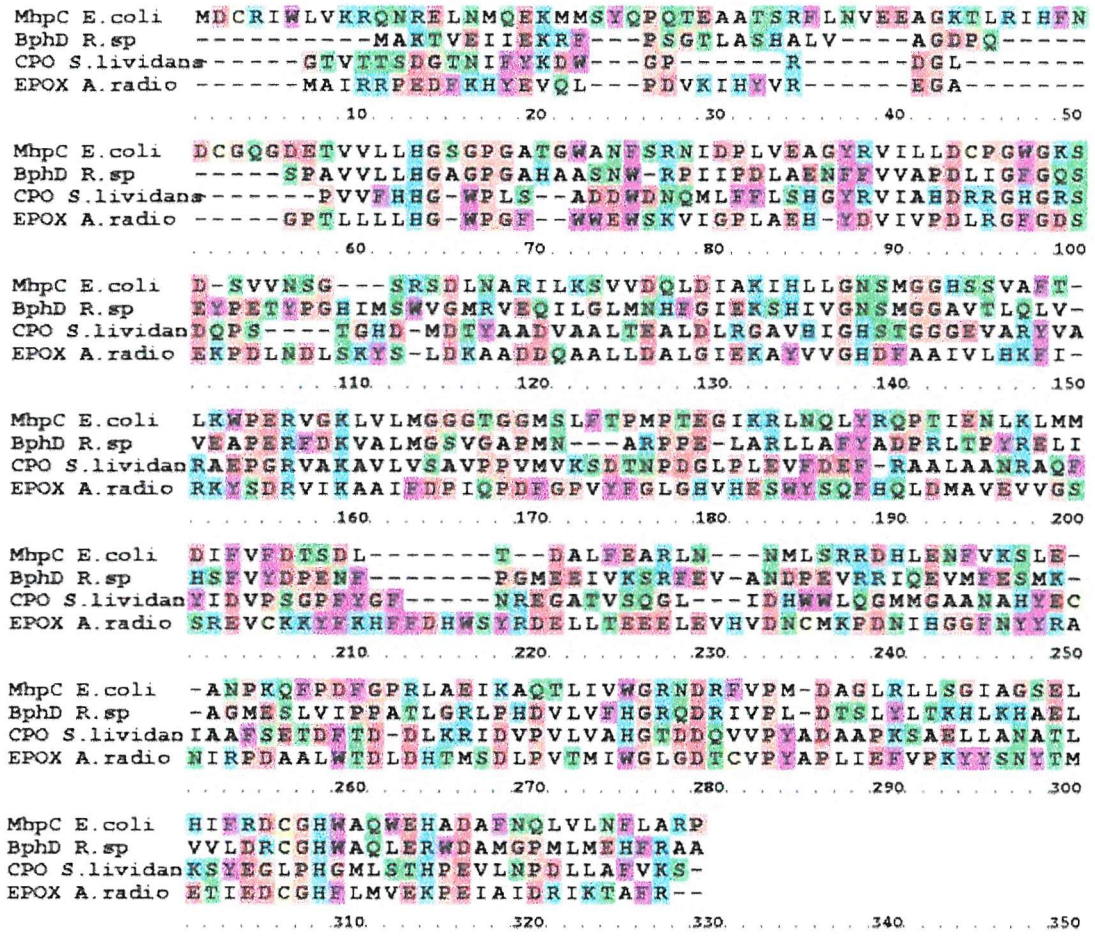


Figure 5.9: Primary sequence alignment of *E. coli* MhpC with biphenyl hydrolase (BphD) from *Rhodococcus sp* RHA1, chloroperoxidase (CPO) from *Streptomyces lividans* and epoxide hydrolase (EPOX) from *Agrobacterium radiobacter* (CINEMA 2.1). The alignment was performed using MALIGN. The sequence identity between MhpC and the putative homologues used in the molecular replacement are as follows: BphD; 29 %, CPO; 21 % and EPOX; 19 %. The colour key for residue types is the following: polar positive (blue), polar negative (red), polar neutral (green), non-polar aliphatic (white), non-polar aromatic (purple), glycine and proline (brown) and cysteine (yellow). The sequences were obtained from the SWISS-PROT/TrEMBL protein sequence database.

5.7.3.2 Using Complete Search Models

All search models were placed in the same orientation and position as the MhpC monomer. Initially the molecular replacement calculations were performed using the complete structures of the models. In each case the molecular replacement was performed using three different packages (X-PLOR, AMoRe and MOLREP) and potential solutions were compared with the control (Table 5.10). Unfortunately, no rotation function solutions consistent with the control were obtained with any of the search models and molecular replacement packages. Similar results were obtained when modifying some of the parameters that affect the rotation function (e.g. radius of integration and resolution limits).

5.7.3.3 Using Incomplete Search Models

The searches with complete search models provided no significant rotation function solutions. The next step involved removing parts of the model structure that had no counterpart in the MhpC structure. Detailed analysis of each three dimensional superimposition revealed several regions of the model which deviated significantly from the MhpC structure (Figures 5.10, 11 and 12) and were subsequently removed from the model (Table 5.14). The regions removed from each search model were selected according to their rms deviation from the corresponding MhpC structure (Figure 5.13a, b and c).

Search Models	Residues Removed	% removed
Biphenyl hydrolase	1-19, 71-88 and 135-226.	45
Chloroperoxidase	1-7, 18-23, 63-72, 88-91 and 120-217.	45
Epoxide hydrolase	1-20, 72-83 and 131-234.	46

Table 5.14: The regions removed from each search model.

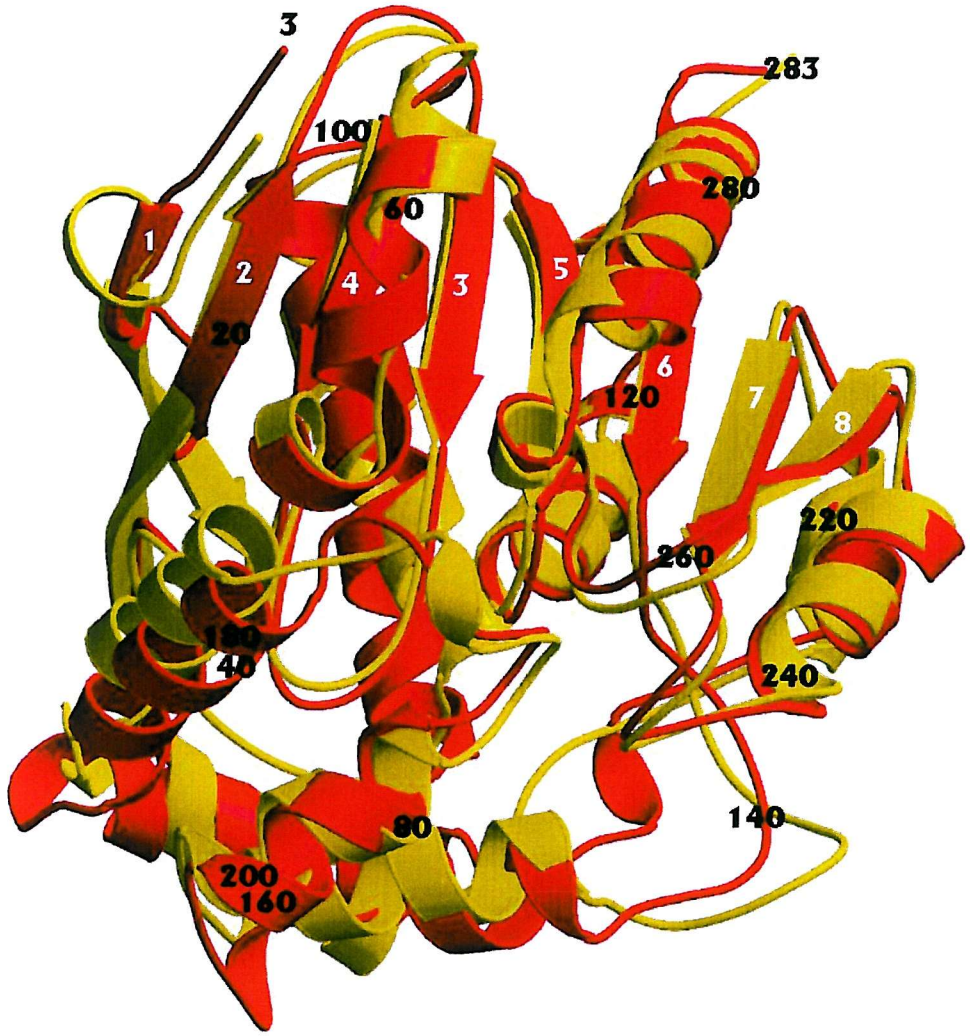


Figure 5.10: A superposition of the structures of MhpC from *E. coli* (yellow) and biphenyl hydrolase from *Rhodococcus sp* strain RHA1 (red) (MOLSCRIPT and rendered using Raster3D). Note that the highest structural similarity is observed for the central β-sheets (eight strands numbered 1 to 8 in white). The largest structural deviations from the MhpC structure (> 1.5 Å) occur at the N-terminal (residues 1-19), loop following strand 4 (residues 71-88) and excursion D following strand 6 (a series of helices and connecting loops that involve residues 135-226). These regions were removed from the biphenyl hydrolase structure in order to achieve a successful molecular replacement solution.

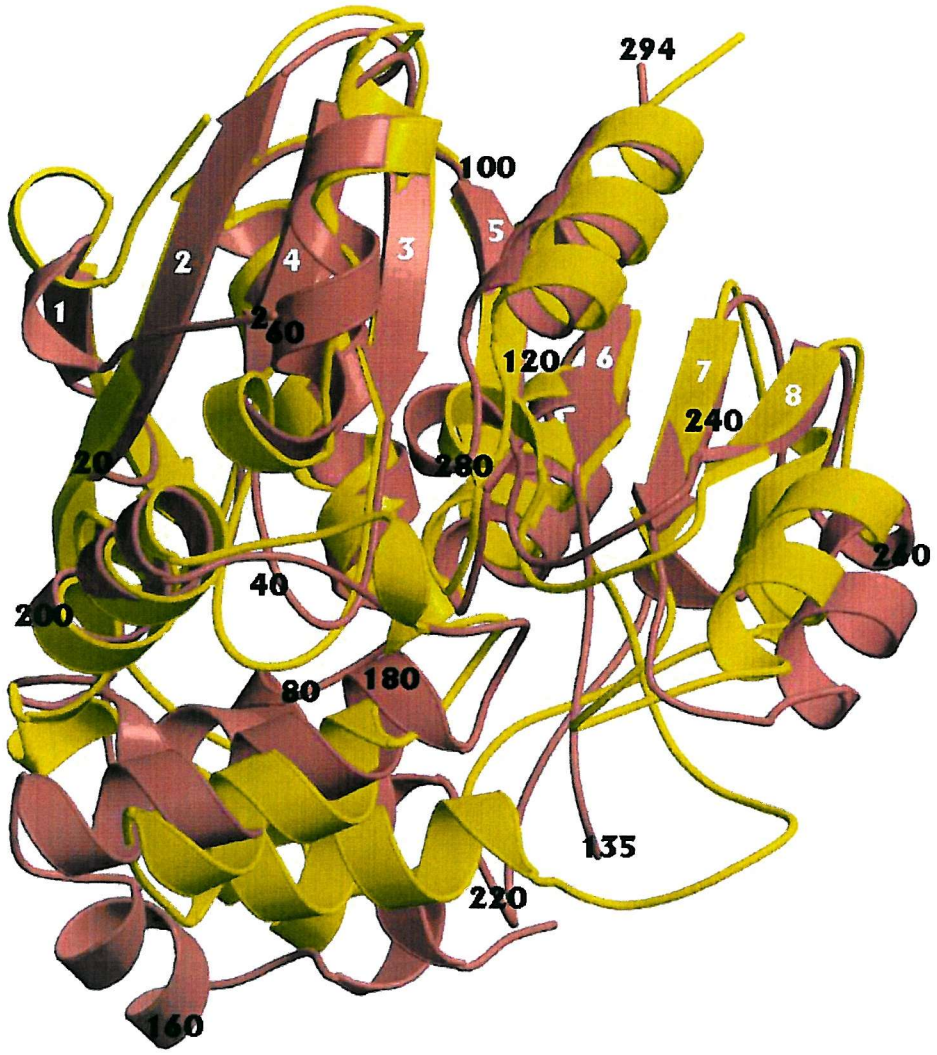


Figure 5.11: A superposition of the structures of MhpC from *E. coli* (yellow) and epoxide hydrolase from *Agrobacterium radiobacter* (pink) (MOLSCRIPT and rendered with Raster3D). Note that the central β-sheets (each strand is numbered 1 to 8 in white) are well superposed. The largest structural deviations from the MhpC structure (> 1.5 Å) occur at the N-terminal (residues 1-20), the loop following strand 4 (residues 72-83) and excursion D following strand 6 (a series of helices and connecting loops that involve residues 131-234).

Figure 5.12: A superposition of the structures of MhpC from *E. coli* (yellow) and chloroperoxidase from *Streptomyces lividans* (green) (MOLSCRIPT and rendered with Raster3D). Note that the central β-sheets (each strand is numbered 1 to 8 in white) are well superposed. The largest structural deviations from the MhpC structure (> 1.5 Å) occur at the N-terminal (residues 1-7), the loop following strand 2 (residues 18-23), the loop following strand 4 (residues 63-72), the loop preceding strand 5 (residues 88-91) and excursion D following strand 6 (a series of helices and connecting loops that involve residues 120-217).

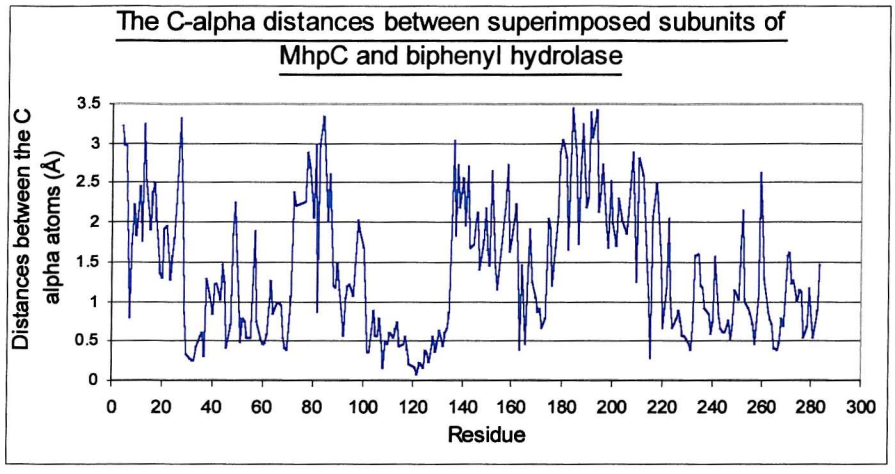


Figure 5.13a: The C-alpha distances between superimposed subunits of MhpC and biphenyl hydrolase (y-axis) as a function of residue number (x-axis). Large deviations (> 1.5 Å) are localised to 3 regions (residues 1-19, 71-88 and 135-226).

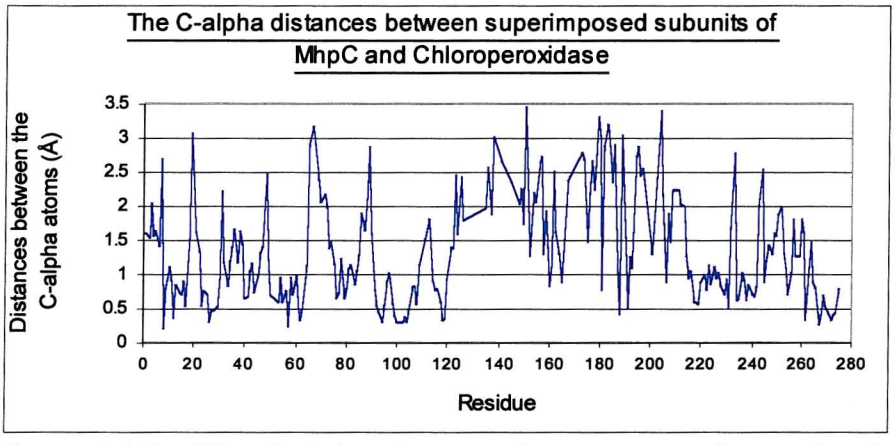


Figure 5.13b: The C-alpha distances between superimposed subunits of MhpC and chloroperoxidase (y-axis) as a function of residue number (x-axis). Large deviations (> 1.5 Å) are localised to 5 regions (residues 1-7, 18-23, 63-72, 88-91 and 120-217).

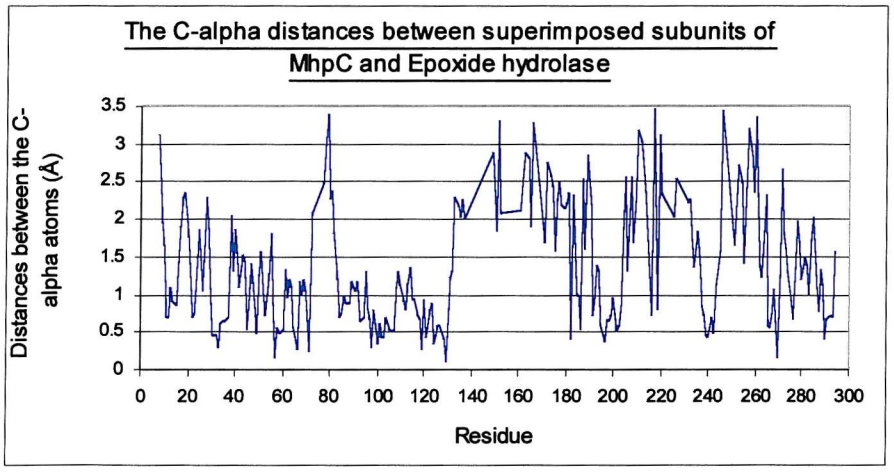


Figure 5.13c: The C-alpha distances between superimposed subunits of MhpC and epoxide hydrolase (y-axis) as a function of residue number (x-axis). Large deviations (> 1.5 Å) are localised to 3 regions (residues 1-20, 72-83 and 131-234).

The molecular replacement calculations were repeated using the trimmed model structures. Once again no rotation function solutions were obtained with the AMoRe and X-PLOR packages. However, MOLREP determined both rotation function solutions (solution 1; α=159.44°, β=90.00°, γ=184.31° and solution 2; α=151.39°, β=88.65°, γ=182.36°) when using the trimmed chloroperoxidase structure as the search model (Table 5.15), although the signal-to-noise ratio was considerably reduced compared to the control. Furthermore, one of the rotation function solutions was also determined (Tables 5.16 and 5.17) when using the trimmed epoxide hydrolase and biphenyl hydrolase structures as the search model. In all cases the best solutions were obtained using data between 14.5 and 3 Å with a radius of integration of 30 Å.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	159.44	90.00	184.31	5.19
2	151.39	88.65	182.36	5.16
3	38.98	50.99	169.57	4.41
4	162.83	44.32	97.58	3.90
5	95.89	19.19	44.80	3.72

Table 5.15: The 5 highest cross rotation peaks calculated with MOLREP using the trimmed chloroperoxidase structure as the search model.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	158.50	87.87	181.89	5.18
2	161.44	63.74	133.32	3.87
3	135.79	81.99	158.56	3.75
4	107.03	59.97	46.33	3.72
5	70.19	54.84	245.54	3.67

Table 5.16: The 5 highest cross rotation peaks calculated with MOLREP using the trimmed biphenyl hydrolase structure as the search model.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	151.52	88.15	182.13	5.14
2	156.08	70.15	148.04	4.23
3	42.43	41.35	262.27	3.79
4	149.72	58.34	126.98	3.76
5	18.26	61.80	61.81	3.75

Table 5.17: The 5 highest cross rotation peaks calculated with MOLREP using the trimmed epoxide hydrolase structure as the search model.

5.7.4 Conclusions

A closer inspection of the three model structures when superposed on the MhpC structure demonstrates why the molecular replacement failed. In accordance with general scheme of the α/β-hydrolase fold, the largest structural deviations (1.5-3.5 Å) were due to different excursions of varying length, which occur between strands 6 and 7 (excursion D) and strands 7 and 8 (excursion E). In particular, excursion D following strand 6 is the most variable since this region encompasses 90-100 residues which are the least homologous in sequence. It is no surprise that the complete search model structures provided no molecular replacement solutions, since there were at least 100 residues in the wrong positions which increased the background noise.

Performing the molecular replacement calculations with parts of the search model removed yielded differing results; of the three molecular replacement packages only MOLREP provided rotation function solutions resembling those obtained for native MhpC (control). Both rotation function solutions were obtained using trimmed chloroperoxidase structure as the search model, although the signal-to-noise ratio had been considerably reduced (peak height decreased by 60 % in each case). Although the trimmed search model provided the correct molecular replacement solution, the quantity of scattering matter placed in the unit cell was reduced, hence reducing the signal that is being searched for. Only one of the rotation function solutions was obtained when using trimmed epoxide hydrolase and biphenyl hydrolase structures. It is of interest that although biphenyl hydrolase has the highest sequence homology to MhpC and is a member of the C-C hydrolase family, it failed to provide both solutions.

All molecular replacement calculations performed with X-PLOR and AMoRe failed to provide either of the rotation function solutions. The success of MOLREP in determining the both solutions must be due to unique features that are not implemented in X-PLOR or AMoRe. These include new approaches to the way data is handled and the rotation and translation searches performed (Vagin and Teplyakov, 1997). The strategy employed by MOLREP to scale the observed and calculated amplitudes is based on scaling of origin Patterson maps (Rogers, 1965); scaling by Patterson synthesis is advantageous when only low resolution data is available and during the cross rotation function calculations where the cell differs for the search model and the target structure.

Low resolution reflections are often omitted from molecular replacement calculations because they contain information contributed by the bulk solvent. As a result, a series termination effect is introduced whereby systematic errors in electron density are observed. To overcome this problem MOLREP applies a 'soft' low resolution cut-off in all calculations by multiplying the structure factor amplitudes by a special coefficient. This reduces the weight of the low resolution terms rather than removing them and has been shown to increase the signal-to-noise ratio (Vagin and Teplyakov 1997). Another feature that may have increased the likelihood of obtaining a solution with MOLREP is that the rotation function is performed for three different orientations of the search model and averaged over them, thereby further increasing the signal-to-noise ratio.

In the case of MhpC, it is a combination of factors that colluded to prevent a successful molecular replacement solution, even though all the search models possessed the same fold. These factors include the low sequence homology between MhpC and the search models, the relatively large number of MhpC molecules being searched for in the target asymmetric unit and the high proportion of residues being in the wrong position (series of helices and connecting loops contributing to the background noise). The search models were improved by removing non-homologous parts of the structure resulting in the signal being searched being considerably reduced, making the molecular replacement more difficult; a single search model constitutes less and less of the scattering matter in the target asymmetric unit as the number of molecules in it are increased. However using the trimmed chloroperoxidase structure as the search model, the noise was reduced to a greater extent allowing a successful rotation function solution to be determined.

5.8 References

Ahmad, D., Fraser, J., Sylvestre, M., Larose, A., Khan, A., Bergeron, J., Juteau, J.M. and Sondossi, M. (1995) *Gene* **156** 69-74

Chapus, C., Semeriva, M., Bovier-Lapierre, C. and Desnuelle, P. (1976) *Biochemistry* **15** 4980-4987

Dagley, S. (1986) in *The Bacteria* eds Sokatch, J.R. and Ornston, L.N. 527-555, Academic Press, New York

Diaz, E. and Timmis, K.N. (1995) *J. Biol. Chem.* **270** 6403-6411.

Dunn, G. (2000) Ph. D. Thesis, University of Southampton

Ferrández, A, Garcia, J.L. and Díaz, E. (1997) *J. Bacteriol.* **179 (8)** 2573-2581

Fleming, S.M., Robertson, T.A., Langley, J.G. and Bugg, T.D.H. (2000) *Biochemistry* **39** 1522-1531

Heikinheimo, P., Goldman, A., Jeffries, C. and Ollis, D.L. (1999) *Structure* **7** R141-R146

Henderson, I.M.J. and Bugg, T.D.H. (1997) *Biochemistry* **36** 12252-12258

Hofmann, B., Tolzer, S., Pelletier, I., Altenbuchner, J., van Pee, K.H. and Hecht, H.J. (1998) *J. Mol. Biol.* **278** 889-900

Lam, W.W.Y. and Bugg, T.D.H. (1994) *J. Chem. Soc. Chem. Commun.* 1163-1164

Lam, W.W.Y. and Bugg, T.D.H. (1997) *Biochemistry* **36** 12242-12251.

Nandhagopal, N., Senda, T., Hatta, T., Yamada, A., Masai, E., Fukuda, M. and Mitsui, Y. (1997) *Proc. Japan. Acad* **73(B)** 154-157

Nardini, M., Ridder, S., Rozeboom, H.J., Kalk, K.H., Rink, R., Janssen, D.B. and Dijkstra, B.W. (1999) *J. Biol. Chem.* **274** 14579-14586

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frowlow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G. and Goldman, A. (1992) *Protein. Eng.* **5** 197-211

Rogers, D. (1965) in *Computing Methods in Crystallography* ed Rollett, J.S. 117-148, Oxford University Press

Rost, B. and Sander, C. (1993) *J. Mol. Biol.* **232** 584-599

van de Meer, J.R., de Vos, W.M., Harayama, S. and Zehnder, A.J.B. (1992) *Microbial. Rev.* **56** 677-694

Seah, S.Y.K., Terracina, G., Bolin, J.T., Riebel, P., Snieckus, V. and Eltis, L.D. (1998) *J Biol Chem.* **273** 22943 22949

Weeks, C.M. and Miller, R. (1999) *J. Appl. Cryst.* **32** 120-124

Yoshimoto, T., Kabashima, T., Uchikawa, K., Inoue, T., Tanaka, N., Nakamura, K.T., Tsuru, M. and Ito, K. (1999) *J. Biochem.* **126** 559-565
(http://www.expasy.ch/swissmod/swiss-MODEL.html)

General Summary

General Summary

The molecular replacement method has been used to determine the three-dimensional structures of four enzymes; human R167Q-uPBGD, bovine IMPase, D303E-MDH from *M. extorquens* and MhpC from *E. coli*. The relative ease of determination of the correct molecular replacement solution for each enzyme varied depending on the quality of the search model, the number of molecules in the asymmetric unit and the quality of X-ray data. In this work, the most important parameter to consider was the degree of sequence identity that existed between the search model and the target protein.

Of the four enzymes, the molecular replacement solutions for D303E-MDH were determined relatively easily even though the space group of the crystal was initially uncertain (either P2 or $P2_1$). The wild-type MDH structure was used as the search model (100 % sequence identity) and the calculations were performed using MOLREP. The high signal-to-noise ratio in the rotation and translation function calculations made the selection of a correct solution clear. This also provided the opportunity to use the translation function to overcome space group ambiguity. Similarly, a correct molecular replacement solution for bovine IMPase was determined without difficulty using AMoRe and cross validated with X-PLOR. Structural analysis of bovine IMPase was made easier owing to the quality of the search model (human IMPase shares 85 % sequence identity with the bovine enzyme) and the crystal quality (a complete dataset was collected to 1.65 Å).

There were several problems associated with the structure determination of human R167Q-uPBGD. Firstly, an incomplete dataset was collected (high resolution data ranges from 2.7 to 3.5 Å) as a result of the deleterious effect of radiation on the crystal. Secondly, following data processing a space group ambiguity for the R167Q-uPBGD crystal form existed ($P222_1$, $P2_12_12_1$ or $P2_12_12$). Thirdly, it was anticipated that the search model (*E. coli* PBGD) would be considerably different even though the overall fold was similar (46 % sequence identity and 70 % sequence similarity); the most significant differences were expected to arise from the relative movement of the three domains. Finally, it was possible that the presence of the large insert in R167Q-uPBGD would have an aberrant affect on the molecular replacement calculations. However, the structure of R167Q-uPBGD was solved using MOLREP although it would have been a perfect

candidate for the application of PC refinement in X-PLOR since it possesses three flexible domains. Doubts about the low signal-to-noise ratio for the second molecule in the non-crystallographic translation function were put to rest following the demonstration of sensible crystal packing and the drop in the R-factor and R_{free} during the initial round of refinement.

The limitations of the molecular replacement method were clearly established during the structural analysis of MhpC. The major problems associated with the particular crystal of MhpC were the twinning phenomenon and the presence of a relatively large number of molecules in the asymmetric unit (four based on solvent content calculations). Furthermore, although all the search models possessed the α/β-hydrolase fold (as predicted previously for MhpC), the sequence identity with MhpC was low (19-29 %). The absence of a molecular replacement solution was largely attributed to a combination of the above factors.

A retrospective analysis of the structure determination of MhpC highlighted the importance of the structural similarity between the search model and the target protein; removal of approximately 120 residues (differences of more than 1.5 Å) from the chloroperoxidase structure provided the correct solutions. Therefore, it is evident that the molecular replacement method often fails if differences between the search model and the target structure are more than ~1.5 Å. The principal cause may be due to the fact that when the degree of sequence identity between search model and the unknown structure is less than 50 %, the overall fold is probably maintained but the relative orientations and positions of the secondary structural elements or domains can change dramatically. This may explain why the search models with the same fold as MhpC failed to provide the correct molecular replacement solutions.

In conclusion, there is no doubt that molecular replacement is a powerful tool for protein structure determination, although its success will always depend on the quality of the search model employed. It is also apparent that through structural genomics, the molecular replacement method will become more powerful as the number of proteins deposited in the Brookhaven PDB increases, and that the more time consuming MIR analysis will be used less in future.