UNIVERSITY OF SOUTHAMPTON

THE STRUCTURE AND MECHANISM OF THE C-C BOND HYDROLASE MhpC

Mark Greer Montgomery

A thesis submitted for the degree of DOCTOR OF PHILOSOPHY

Division of Biochemistry and Molecular Biology

< July 2003 >

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE
DIVISION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Doctor of Philosophy

THE STRUCTURE AND MECHANISM OF THE C-C BOND HYDROLASE MhpC

By Mark Greer Montgomery

The woody tissues of trees and plants are composed primarily of macromolecules such as lignin, which can be decomposed initially by fungi and the resulting fragments then degraded by a plethora of microorganisms. Bacteria have developed catabolic pathways to degrade these phenolic compounds into primary metabolites utilised in the Krebs cycle. Many man-made, or xenobiotic, compounds can also be accepted by enzymes of these pathways. Bioremediation is a promising method for removing xenobiotics from the environment based on the capacity of microorganisms to transform pollutants to non-toxic products.

One such catabolic pathway is that which degrades phenylpropionic acid and has been found in a number of bacteria, including Escherichia coli. The third enzyme in this pathway is MhpC, a carbon-carbon bond hydrolase which catalyses the cleavage of 2-hydroxy-6-ketonona-2,4-diene-1,9-dioic acid to succinate and 2-hydroxypenta-2,4-dienoic acid. The mechanism of this enzyme has been studied extensively by biochemical methods and it is known that catalysis involves firstly the formation of a keto-intermediate followed by cleavage of the substrate utilising either an active site nucleophile, such as in the serine proteases, or by base-catalysed attack of water, such as in the aspartic proteases.

The principle aim of this research was to use crystallographic methods to obtain structures of MhpC in complex with ligands which may shed light on the mechanism. The native structure was solved to 2.1Å resolution and revealed that MhpC belonged to the α/β-hydrolase fold family with a catalytic triad of Ser110-Asp235-His263. Although MhpC exists as a dimer in solution it was found to adopt a tetrameric quaternary structure in the crystal. It was also observed that the active site serine may have been covalently modified, but attempts to identify the source of this modification have at present proved futile. The structures of MhpC complexed with two substrate analogues, azelaic acid and 2,6-diketo-nona-1,9-dioic acid (DKNDA), and two product analogues, laevulinic acid and 4,6-dioxoheptanoic acid (DOHA), were obtained. The structures with azelaic acid and laevulinic acid revealed many residues within the active site cavity that could be involved in substrate binding. The structure with DOHA revealed a disparity between subunits of the MhpC dimer, which was confirmed in the DKNDA structure and indicates half-site reactivity may be a feature of the MhpC-catalysed reaction. The structure with DKNDA also revealed that this compound forms a hemi-ketal adduct with the enzyme through Ser110 and hence that the formation of an acyl-enzyme intermediate is feasible. The presence of a tripartite oxyanion binding site was also revealed, involving Ser40, which could indicate maintenance of a catalytic tetrad. Thus, a further link between this family of C-C hydrolases and clan SC of the serine proteases is made.

CONTENTS

LIST OF FIGURES AND SCHEMES

CHAPTER 5 – STRUCTURES OF MhpC COMPLEXED WITH PRODUCT ANALOGUES 159

CHAPTER 6 – STRUCTURES OF MhpC COMPLEXED WITH SUBSTRATE ANALOGUES 184

LIST OF TABLES AND BOXES

ABBREVIATIONS

A_{260}	absorbance at 260nm
ACE	acetylcholinesterase
AEBSF	4-(2-aminoethyl)benzenesulphonyl fluoride
ArEHY	*Agrobacterium radiobacter* EHY
ATP	adenosine-5'-triphosphate
bBACE	bovine bile-salt-activated cholesterol esterases
BphC	2,3-dihydroxybiphenyl-1,2-dioxygenase
BphD	2-hydroxy-6-keto-6-phenylhexa-2,4-dienoic acid hydrolase
BsBES	*Bacillus subtilis* Brefeldin A esterase
BTX	benzene/toluene/xylene
°C	degrees Celsius
CaFLP3	*Candida antarctica* lipase
CBPA	carboxypeptidase A
CCD	charge-coupled device
CCP4	Collaborative Computing Project Number 4
CmtE	2-hydroxy-6-keto-7-methylocta-2,4-dienoate hydrolase
CNS	Crystallography and NMR System
CoA	coenzyme A
CPW	wheat carboxypeptidase II
1,2-CTD	catechol-1,2-dioxygenase
CUT	cutinase
Da.	Daltons
DHP	2,3-dihydroxyphenylpropionic acid
DIFP	diisopropylfluorophosphate (aka DIPF)
DIPF	diisopropylphosphofluoridate (aka DIFP)
DKNDA	2,6-diketonona-1,9-dioic acid
DLH	dienelactone hydrolase
DmpD	2-hydroxymuconic semialdehyde hydrolase
DNA	deoxyribonucleic acid
DOHA	4,6-dioxoheptanoic acid
DOSA	4,7-dioxosebacic acid

DDT	dichlorodiphenyltrichloroethane
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EHY	epoxide hydrolase
ES+	positive electrospray (mass spectrometry)
ESRF	European Synchrotron Radiation Facility
EtHPD	ethyl ester of HPD
FAD	flavin adenine dinucleotide
FAH	fumarylacetoacetate hydrolase
FAV	field application vector
FLP2	fungal type B lipase
FPLC	fast protein liquid chromatography
FsCUT	*Fusarium solani* cutinase
GcFLP2	*Geotrichum candidum* FLP2
GLP	*Geotrichum candidum* lipase
HAL	haloalkane dehalogenase
HCA	hydrocinnamic acid
HcaA	3-phenylpropionic acid 2,3-dioxygenase
HcaB	3-phenylpropionic acid 2,3-dihydrodiol dehydrogenase
3HCI	3-hydroxycinnamic acid
2,4-HDA	*trans,trans*-2,4-hexadienoic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
hHSL	human hormone-sensitive lipase
HKP	4-hydroxy-2-ketopentanoic acid
hLIP	human pancreatic lipase
HMPOBA	4-(hydroxymethylphosphinoyl)-3-oxo-butanoic acid
HPA	4-hydroxyphenylacetate
HPD	2-hydroxypenta-2,4-dienoic acid
HPLC	high performance liquid chromatography
3HPP	3-(3-hydroxyphenyl)propionic acid (aka MHP)
HT1	hereditary tyrosinemia type I
Hz	hertz
IPTG	*iso*-propyl-β-D-galactopyranoside

KNDA	4-ketonona-1,9-dioic acid
LB	Luria broth
LB (Amp)	LB containing ampicillin
LC-MS	liquid chromatography-mass spectrometry
LES	large esterase
mACE	mouse ACE
MAD	multiwavelength anomalous dispersion (or diffraction)
MD1-01	Molecular Dimensions Ltd. Structure Screen 1
MD1-02	Molecular Dimensions Ltd. Structure Screen 2
β-ME	β-mercaptoethanol
mEHY	mouse EHY
MeRFP	methyl ester of RFP
MES	2-(*N*-Morpholino)ethanesulphonic acid
MHP	*m*-hydroxyphenylpropionic acid
MhpA	*m*-hydroxyphenylpropionic acid 2-hydroxylase
MhpB	2,3-dihydroxyphenylpropionic acid 1,2-dioxygenase
MhpC	2-hydroxy-6-ketonona-2,4-diene-1,9-dioic acid 5,6-hydrolase
MhpD	2-hydroxypenta-2,4-dienoic acid hydratase
MhpE	4-hydroxy-2-ketopentanoic acid aldolase
MhpF	acetaldehyde dehydrogenase (acylating)
mLIP	*Rhizomucor miehei* lipase
MOPS	3-(*N*-Morpholino)propanesulphonic acid
MPD	2-methyl-2,4-pentanediol
m/z	mass-to-charge ratio
NAD(H)	nicotinamide adenine dinucleotide (reduced form)
NMR	nuclear magnetic resonance (spectroscopy)
OD_{600}	optical density at 600nm
OSA	4-oxosebacic acid
P13DLH	*Pseudomonas* sp. B13 DLH
PCB	polychlorinated biphenyl
2,3-PCD	protocatechuate-2,3-dioxygenase
3,4-PCD	protocatechuate-3,4-dioxygenase
4,5-PCD	protocatechuate-4,5-dioxygenase

PCDD	polychlorinated benzodioxin
PCE	tetrachloroethylene
PEG	polyethylene glycol
PfCES	*Pseudomonas fluorescens* carboxylesterase
PLP	pancreatic lipase
PMSF	phenylmethylsulphonylfluoride
PP	3-phenylpropionic acid (aka HCA)
pPOP	porcine proline oligopeptidase
PVC	polyvinylchloride
RFP	ring fission product (eg. substrate for MhpC)
RFP^k	ketonised ring fission product
RNase A	ribonuclease A
rpm	revolutions per minute
SCP	serine carboxypeptidase
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
t3N2O	*trans*-3-nonene-2-one
TAE	tris-acetate-EDTA
TCA	tricarboxylic acid
TcACE	*Torpedo californica* ACE
TCE	trichloroethylene
TFA	trifluoroacetic acid
TodF	2-hydroxy-6-keto-2,4-heptadienoate hydrolase
TPCK	tosyl-L-phenylalanine chloromethylketone
Tris (HCl)	tris(hydroxymethyl)aminomethane (hydrochloride)
UV	ultraviolet
wSCP	wheat SCP
XaHAL	*Xanthobacter autotrophicus* HAL
XylF	2-hydroxymuconic semialdehyde hydrolase

ACKNOWLEDGMENTS

First and foremost, deepest gratitude goes to the guy who gave me the opportunity in the first place, Prof. Steve Wood. Top Man! There aren't superlatives enough in the World to do him justice. What an absolutely Diamond Geezer!

Next up, Dr. Jon Cooper, who almost single-handedly sustained me with his exhaustive supply of biscuits and frequent trips to the Garage (soon to be novelised apparently). He also put up with a lot of inane questioning on Crystallography and always came up with an answer. A dab hand at bleeding brakes too. Quality!

Thanks also to Prof. Tim Bugg, at Warwick Uni, for not only introducing us all to MhpC but being around for the resultant questioning and giving a reason to escape Southampton for the exotic climes of Coventry.

Many thanks to present and past (ie. those that scarpered to Pompey/Reading) members of the Wood and Cooper groups. In alphabetical order because I know what they're like: Doug 'Tall, Ginger and for some reason he liked LINUX' Baker, Gordon 'Yes, he really has always been that tall' Beaven, Leighton 'He's a bit keen on his DVDs' Coates, Alun 'My wallpaper needs removing if you really need a job' Coker, Graham 'If I say Michael Owen will score, he will score' Dunn, Peter 'He's the TURBO-FRODO daddy, don't ya know' Erskine, Raj 'I could take the Uni to court for this, Is it cos I is Brown?' Gill, Michelle 'Blimey, there's a girl in the group?!' Jenvey, Simon 'There's bound to be a Gilson hidden in his cupboard somewhere' Kolstoe, Sanjay 'King of the Landlords and likes a pot of tea' Mall, Fiyaz 'He was here earlier, he's probably playing tennis' Mohammed, Alan 'rendering pictures is my middle name' Purvis, Terry 'Jim Bean, Lennons, say no more' Robinson, Darren 'Yes, he's Ginger but he knows his Crystallography' Thompson, and last but not least Paul 'I'm a Patrick Rafter me, one more for the road?' Williams. It's been a pleasure to know all of these people, they've helped me in so many different ways, and some of them especially have been known to share a quick drink or two with me. Cheers!

Hi to all the other people that I've bumped into in and around the Department during my time here, especially those in the Fox and Jordan groups. Special thanks to Stuart Findlow for his valuable help with NMR and also thanks to Joan Street in the Chemistry Dept for pressing the start button. Cheers to Neville Wright and Paul Skipp for help with all the Mass Spec work and to Bob Broadbridge for help with organic synthesis type stuff.

Now to those people outside of Boldrewood that I've had the pleasure to meet during my time in Southampton. And no not all of them were publicans or bar staff! Big up to Sarah 'n' James for putting up with me at their place many times and loaning me cash. Nice one! Thanks to those people who let me stay in America: Tom 'Crazy-legs Crane' Robertson, 'Bob' Bomassi & Tom 'Yankee Boy' Bernhardt. Cheers to all the ex-Bugg group members and others at Warwick Uni, especially Kirstin and Simon for letting me crash at their places. Ross and Big Nige for still being around for a few drinks and the annual Power in the Park pilgrimage. Ruth for managing to get a life outside of Chemistry, it can be done! All my housemates, from the Montefiore days through Brickfield Road to the current crop (who just happen to be Chemists). So, cheers to the butcherisers that are Ro, Dicko and Head. Oh and not forgetting Rach for all the fun and games in the final year!

Finally, how could it be complete without acknowledging my family and friends back in Devon. Big hugs to all of you back at home. I've missed you loads.

Oh, and cheers Paul Sturrock.

CHAPTER 1

INTRODUCTION

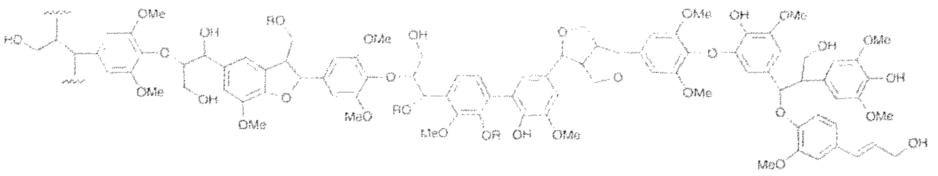
The whole biosphere exists in a dynamic state in which a great number of organic compounds are continually synthesised, transformed and decomposed. These segments together constitute the carbon cycle of our planet, and for smooth operation of this cycle, degradation is just as important as synthesis. The Plant Kingdom accumulates great quantities of polymers like lignin, and innumerable compounds like alkaloids, terpenes and flavanoids that are biochemically inert.

The release of their carbon for recycling depends almost entirely upon the action of microbial degradative enzymes. Aerobic bacteria employ most of the sequences and cycles which occupy a central position in metabolic maps and which are generally found in other living forms. Their unique biochemical assets lie in their ability to catalyse early steps in degradation that other organisms cannot accomplish, and thereby to form metabolites that can enter the common pathways of metabolism, such as the Krebs cycle or the fatty acid "spiral".

It has been estimated that about $1.5x10^{10}$ tons of carbon in the form of carbon dioxide is annually transformed into wood (Bolin, 1973), of which 18-35% by dry weight is lignin (Dagley, 1975). Since microorganisms are solely responsible for hydrolysing cellulose and degrading lignin, it is evident that they must make a tremendous contribution towards the continuous operation of the carbon cycle.

1.1. Microbial degradation of lignin

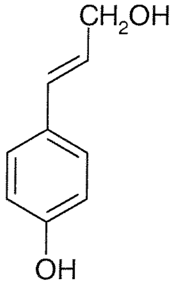
Lignin is the most abundant renewable aromatic material on Earth, but is degraded by a narrower array of microbes than the other major biopolymers. Lignin biodegradation is central to the Earth's carbon cycle because lignin is second only to cellulose in abundance and, perhaps more significantly, because lignin physically protects most of the world's cellulose and hemicelluloses from enzymatic hydrolysis (Kirk and Farrell, 1987).



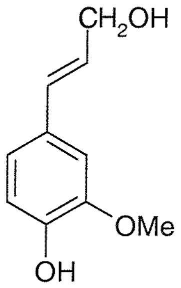
Diagrammatical representation of lignin

Lignin is found in higher plants and constitutes 20-30% of their wood and other vascular tissues. Most lignin is found within the cell walls, forming a matrix with the hemicelluloses that surround the orderly cellulose microfibrils.

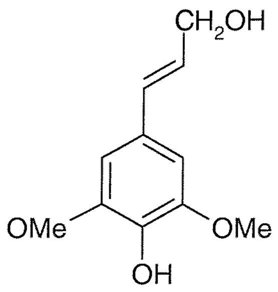
Biosynthetically, lignin arises from three precursor aromatic alcohols: *p*-hydroxycinnamyl (coumaryl) alcohol (**1**), which gives rise to *p*-hydroxyphenol units in the polymer; 4-hydroxy-3-methoxycinnamyl (coniferyl) alcohol (**2**), the guaiacyl units; and 3,5-dimethoxy-4-hydroxycinnamyl (sinapyl) alcohol (**3**), the syringyl units. Free radical copolymerisation of these alcohols produces the heterogenous, optically active, cross-linked, and highly polydisperse polymer.



1



2



3

In the polymerisation process, secondary reactions lead to cross-linking between lignin and hemicelluloses. In addition, the lignins of grasses and certain woods contain aromatic or cinnamic acid esterified through side-chain hydroxyl groups of lignin.

Studies on the decomposition in anaerobic lake sediments of synthetic lignins, an alkali-degraded synthetic lignin, a dimeric lignin model compound and lignin-related phenols show that lignin is apparently not biodegraded anaerobically (Zeikus *et al.*, 1982).

Neither rapid nor extensive bacterial degradation, even under highly aerobic conditions, has been reported. As in the case of the anaerobes, a limiting factor might be the size of the lignin polymer. The most rapid and extensive degradation described is caused by certain fungi, particularly the white-rot fungi, in highly aerobic environments.

Soft-rot wood decay caused by various species of ascomycetes and fungi imperfecti involves lignin degradation, although wood polysaccharides are preferentially degraded (Kirk, 1984; Kirk and Cowling, 1984). Several ascomycetes, fungi imperfecti, and phycomycetes that are not associated with soft rot of wood, including 12 marine fungi (Sutherland *et al.*, 1987), 18 *Trichoderma* strains (Flegel *et al.*, 1982), and *Trichoderma harzianum* (Kern, 1983), failed to degrade lignin significantly.

Basidomycetes that cause brown-rot wood decay also partially decompose lignin. They are closely related to the white-rot fungi; and invade the lumens of wood cells, where they secrete enzymes that decompose and remove the polysaccharides, leaving behind a brown, modified lignin residue. Studies show that the lignin undergoes limited aromatic hydroxylations and ring cleavage, but that the major effect is demethylation of aromatic methoxyl groups (Kirk, 1984; Kirk and Adler, 1970). Ability of the brown-rot fungi to mineralise the backbone lignin polymer thus seems to be limited.

The white-rot basidomycetes degrade lignin more rapidly and extensively than the other studied microbial groups. Like the brown-rot fungi, they invade the lumens of wood cells, where they secrete enzymes that degrade lignin and the other wood components. During its mineralisation by white-rot fungi, lignin undergoes a number of oxidative changes, including aromatic ring cleavage (Chen and Chang, 1985). Also, a progressive depolymerisation occurs and releases a wide array of low-molecular weight fragments. Generally, fragments of <1kDa predominate (Faix *et al.*, 1985; Leisula *et al.*, 1983).

One species of white-rot fungus, *Phanerochaete chrysosporium*, has been studied widely and exhibits the highest reported rates of lignin degradation (Yang *et al.*, 1980). Several enzymes, including ligninases, Mn peroxidases, phenol-oxidising enzymes, and H_2O_2-producing enzymes, are thought to have roles in lignin biodegradation.

Ligninase (lignin peroxidase) activity, an extracellular H_2O_2-requiring enzyme activity, was discovered in *P. chrysosporium* in 1983 (Gold *et al.*; Kirk and Tien). Ligninase was shown to contain one mole of protoheme IX per mole of enzyme. Studies proved that the essential mechanism of ligninase is peroxidative. The Mn peroxidases are extracellular enzymes that function as phenol-oxidising enzymes and may also participate in H_2O_2 production.

Most white-rot fungi produce extracellular laccase (EC 1.10.3.2). This blue copper oxidase catalyses the one-electron oxidation of phenols to phenoxy radicals (Reinhammer, 1984). Among the consequences of the one-electron oxidation in lignin-related phenols are Cα-oxidation, limited demethoxylation, and aryl-Cα cleavage (Kirk and Shimada, 1985). Although *P. chrysosporium* belongs to a minority of white-rot fungi that produce no detectable laccase, all white-rot fungi secrete enzymes capable of oxidising phenols.

As wood is degraded by white-rot fungi, the lignin exposed on the interior microsurfaces within the wood is oxidised. Studies have revealed that fungal attack decreases the methoxyl, phenolic, and aliphatic hydroxyl contents, cleaves aromatic nuclei to aliphatic carboxyl-containing residues, creates new Cα-carbonyl and carboxyl groups, and forms alkoxyacetic acid, phenoxyacetic acid, and phenoxyethanol structures (Chen and Chang, 1985). Many of these degradative changes in the lignin polymer can be attributed to ligninase; however, some cannot, which indicates that other enzymes participate in polymer breakdown.

1.1.1. Summary

Certain ascomycetes and fungi imperfecti, particularly those that cause soft rot and the few that cause a white rot of wood, can degrade lignin. Growing evidence indicates that ligninase (lignin peroxidase) is the key lignin-degrading enzyme of white-rot fungi. Other studies indicate that enzymes other than ligninases are also involved, such as Mn peroxidases and other phenol-oxidising enzymes. Many aerobic bacteria, including actinomycetes, can degrade certain lignocellulose materials. They possess non-specific intracellular systems that probably degrade lignin fragments produced abiotically and by other microbes.

1.2. Biodegradation of aromatic compounds in the biosphere

Many organic compounds that are degraded slowly, if at all, by plants and animals can be oxidised to carbon dioxide by soil microorganisms, and so serve as sources of carbon and energy for their growth. This being the case for the breakdown of lignin initiated by fungi, with the smaller aromatic fragments thus generated being further degraded by both fungi and bacteria.

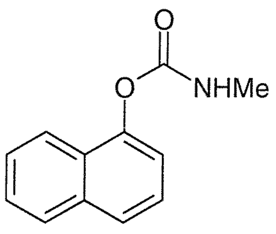
A rapidly increasing number of man-made aromatic compounds have found their way into the environment, for example as pesticides, detergents, oils, solvents, paints or explosives. A man-made compound will be biodegradable if it is susceptible to attack by the enzymatic apparatus acquired by microbes during the course of evolution. This, in turn, depends upon two factors: first, the ability of microbial enzymes to accept as substrates compounds having similar structures to, but not identical with, those found in nature; and second, the ability of these novel compounds to induce the expression of the necessary degradative enzymes. For instance, biodegradability is less likely to be achieved by incorporation of structural features into a pesticide molecule that are never encountered in natural products.

However, many man-made compounds are fairly readily degraded by microorganisms using the same enzymes used for degradation of naturally occurring compounds (Dagley, 1975). An example being the carbamate insecticide carbaryl (**4**), which can be degraded *via* the bacterial naphthalene degradative pathway (Davies and Evans, 1964). Unfortunately, a significant number of these unnatural compounds (xenobiotics) are degraded very slowly, or not at all, by microorganisms, and can persist in the environment. It is of considerable environmental concern if toxic synthetic chemicals persist in the environment as they can cause contamination of freshwater and accumulate in the food chain.

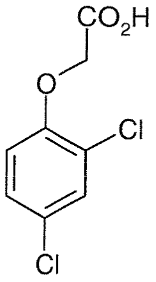
1.3. Xenobiotics in the biosphere: chloro- and nitro-aromatics

Many of the more persistent man-made aromatic hydrocarbons contain chlorine substituents (Reineke, 1988). Chlorination affects the chemical reactivity of the ring and reduces solubility. The presence of a chlorine substituent can either block the normal degradation step, or give rise to a reactive metabolic intermediate which is no longer processed. Thus, the biodegradation of chlorinated aromatics is generally slower and more complex than that of the parent hydrocarbons (Reineke, 1988).

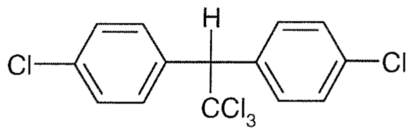
Chlorinated phenols are produced industrially on a large scale, and can be degraded by several strains of bacteria by an adaptation of the normal phenol degradation pathway (Steiert and Crawford, 1985). This pathway is also used for the degradation of the herbicide 2,4-D (**5**), which is degraded *via* 2,4-dichlorophenol (Evans *et al.*, 1971).



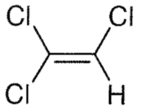
4



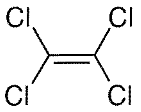
5



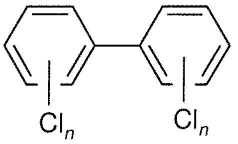
6



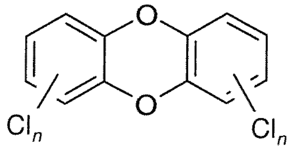
7



8



9



10

However, there are several classes of polychlorinated hydrocarbons which show high persistence in the environment (Reineke, 1988). One well-known example is the insecticide dichlorodiphenyltrichloroethane (DDT, **6**), whose widespread use in the 1950's led, *via* bioaccumulation in the food chain, to the severe poisoning of higher animals (Carson, 1962). Most of the organochlorine insecticides are no longer used in the Western world. Trichloroethylene (TCE, **7**) and tetrachloroethylene (PCE, **8**) are common contaminants of freshwater arising from the manufacture of PVC plastics, and are degraded slowly (Maymo-Gayell *et al.*, 1997).

Two important classes of polychlorinated aromatic pollutants are the polychlorinated biphenyls (PCBs, **9**) and polychlorinated benzodioxins (PCDDs, **10**). PCBs are found in a range of industrial applications, such as transformer oils, heat-transfer fluids, plasticizers and as intermediates in the synthesis of fine chemicals. PCB degradation is very slow, due to their low solubility and ring reactivity, but is made more complicated as the production of these compounds is normally as a complex mixture of isomers. Their industrial synthesis involves reaction of biphenyl with chlorine gas under high temperature and pressure, resulting in a possible 208 isomers. Since enzymes are inherently selective in their action, the presence of such a range of isomers further reduces the rate of degradation (Abramowicz, 1990).

Small quantities of PCDDs arise from the manufacture of (chlorophenoxy)acetic acid herbicides and other chlorinated phenolic chemicals (Horrad and Jones, 1992). They are especially hazardous environmental pollutants, since they have carcinogenic and teratogenic properties.

Another class of man-made aromatics are the nitro-aromatics, used widely as explosives. The presence of a nitro group also reduces the reactivity of the aromatic ring, hindering degradation, and leading to the persistence of nitroaromatic explosives in contaminated soil (Suen and Spain, 1993).

1.4. Bioremediation

Environmental pollution is a direct consequence of human population growth, and the industrial development needed to sustain a continually expanding population has been accompanied by a new quality of environmental pollution, namely: (a) the industrial production and widespread dispersal of chemicals never or rarely seen before in the biosphere (xenobiotics); (b) the localised concentration and release of large quantities of chemicals such as heavy metals and radionuclides ordinarily present in only trace or low concentrations in the biosphere; and (c) the large-scale mobilisation and resulting increase in bioavailability of toxic chemicals that were previously available in only small quantities in the biosphere (Timmis *et al.*, 1994).

The design of efficient microorganisms for bioremediation processes to destroy toxic wastes is a new and rapidly developing branch of waste treatment with considerable promise. The environmental application of genetically engineered organisms requires the use of cloning systems that function under environmental, rather than the normal laboratory setting, are stably maintained, non-transmissible, cost effective, environmentally friendly, and do not allow the spread of resistance to antibiotics.

Biological activities that degrade, detoxify, or concentrate pollutants generally consist of multi-step pathways. For example, degradation of the chemically simple pollutants known as BTX (benzene/toluene/xylene) can involve 10 or more enzymatic reactions. Therefore, in designing strategies to improve catalyst performance, one must determine which enzymatic or regulatory steps are rate limiting under process conditions. The activity of a rate-limiting protein may then be elevated by increasing transcription or translation of its gene or by increasing its stability or its kinetic properties (Timmis *et al.*, 1994).

Protein engineering, ie. by altering its amino acid sequence, may improve enzyme stability, substrate specificity and kinetic properties (Medynski, 1992). Rational design of protein activity requires an understanding of structure-function relationships in the molecule, which in turn necessitates a detailed knowledge of its three-dimensional structure obtained with NMR analysis or X-ray crystallography.

Another means of improving the efficiency of catabolic pathways would be to create bifunctional or multi-functional hybrid enzymes. Though impressive results have been obtained by fusing β-galactosidase with galactose dehydrogenase (Ljungcrantz, 1989), the creation of hybrid enzymes to improve substrate flux through catabolic pathways for aromatic pollutants has not been reported thus far.

Considerable evidence indicates that metabolic routes are biochemically and genetically modular in nature. That is, they evolved as independent units and under prevailing selection pressures subsequently became combined and coordinated (Van der Meer *et al.*, 1992). Typically, central pathways, eg. for ring cleavage and the channelling of simple aromatic compounds to Krebs cycle intermediates, consist of a few modules, whereas the peripheral pathways leading from diverse initial substrates to common ring-cleavage intermediates consist of many modules. The use of molecular genetic methods to clone, characterise, and manipulate the expression of genetic modules encoding useful metabolic sequences will increasingly allow the rational design and assembly of new and improved metabolic routes (Timmis *et al.*, 1994).

1.4.1. Field Application Vectors

Field application vectors (FAVs), which are a combination of a selective substrate, a host, and a cloning vector, have been developed for the purpose of expressing foreign genes in non-sterile, competitive environments in which the gene products provide no advantage to the host (Lajoie *et al.*, 1992). Such gene products are exemplified by the enzymes for the co-metabolism of PCBs through the biphenyl degradation pathway (Bedard *et al.*, 1986).

A FAV system has been developed (Lajoie *et al.*, 1992) that utilises a detergent-degrading strain, which is resistant to 1.0% Igepal CO-720, as a host for heterologous genes. When the detergent was applied to soil it created a temporary ecological disequilibrium, thereby offering the FAV a window of opportunity to establish a functioning population. When PCB-degrading genes (*bphABC*) were cloned into the host (Lajoie *et al.*, 1993) the strain was able to degrade PCBs in the soil. The surfactant provides a dual advantage; it potentially increases PCB bioavailability while minimising competition by native (surfactant sensitive) bacteria.

1.4.2. Summary

Biotechnological processes to destroy toxic wastes offer many advantages over physicochemical processes. When successfully operated, they are non-polluting because they can achieve complete destruction of organic wastes, ie. conversion of toxic organic compounds to innocuous C_1 compounds; are characterised by low costs; and offer the possibility of *in-situ* treatment.

Potential limitations of biological treatment processes include susceptibility to the toxic effects of pollutants at high concentrations, the need to operate within fairly narrow ranges of physicochemical conditions (temperature, pH, salinity etc), and the need for bioavailability of target compounds.

1.5. Strategies for bacterial degradation of aromatic compounds

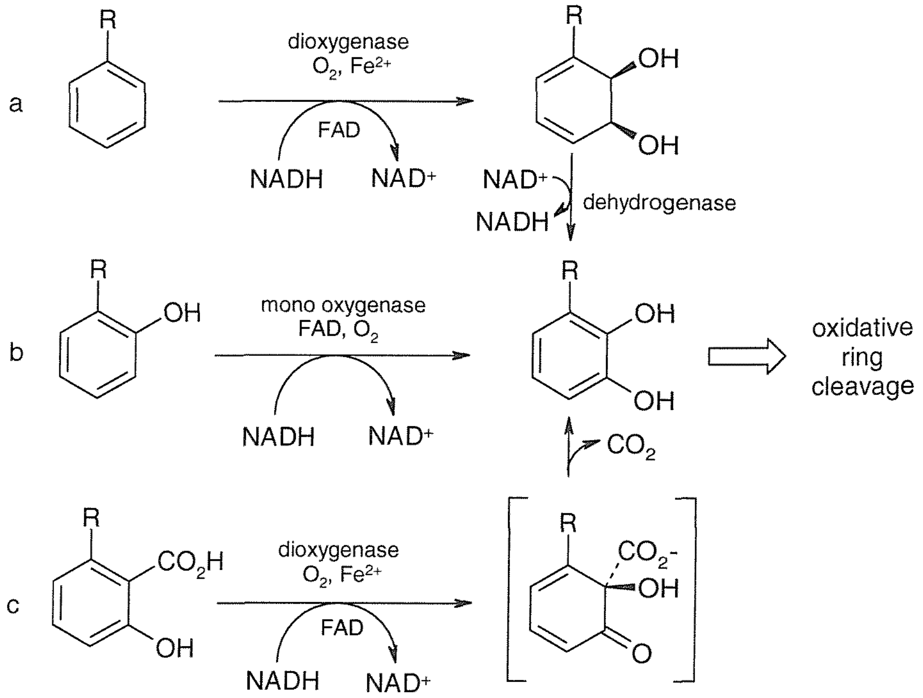
There are two strategies used by microorganisms for aromatic ring breakdown:

(1) by aerobic bacteria, oxidation of the ring to a dihydroxy-aromatic compound (a catechol or a hydroquinone), followed by oxidative cleavage of the ring;
(2) by anaerobic bacteria, reductive hydrogenation of the ring, followed by fragmentation of the cyclohexane ring skeleton (Bugg and Winfield, 1998).

Many simple aromatic compounds such as benzene, toluene, xylenes, benzoic acid, phenylacetic acid and phenylpropionic acid are degraded by aerobic soil bacteria such as *Pseudomonas*, *Acinetobacter* (Gram negative) and *Rhodococcus* (Gram positive) (Dagley, 1975). The latter two compounds are also degraded by *Escherichia coli* (Gram negative), an enteric bacterium. The catabolic pathways of these bacteria are oxidative, proceeding *via* hydroxylation of the ring to give a catechol intermediate, in the following ways.

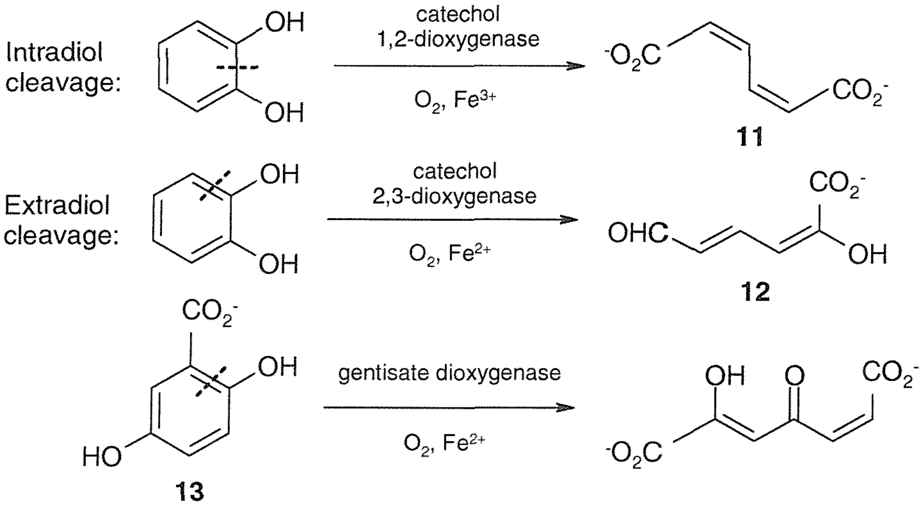
Unsubstituted aromatics such as benzene, naphthalene, and biphenyl are converted into the corresponding 1,2-dihydroxycatechols *via* conversion to the *cis*-1,2-dihydrodiol, followed by oxidation (see Scheme 1a). The dihydroxylation reaction is carried out by a family of three-component dioxygenase enzymes comprising: (1) an NADH-dependent flavin reductase; (2) a ferrodoxin electron transfer protein containing two Rieske [2Fe/2S] clusters; and (3) a terminal dioxygenase subunit, containing a mononuclear iron (II) cofactor and two further [2Fe/2S] clusters (Mason and Cammack, 1992). The subsequent oxidation of the *cis*-dihydrodiol is carried out by a family of NAD^+-dependent dehydrogenases.

Catechols are also commonly produced by *ortho*-hydroxylation of phenols, carried out by FAD-dependent monooxygenases (see Scheme 1b), whose mechanism of action has been reviewed (Massey, 1994). Other oxidative transformations can also give rise to catechols, including the oxidative decarboxylation of salicyclic acid (see Scheme 1c), which occurs *via* a dihydroxylation reaction similar to that in Scheme 1a (Mason and Cammack, 1992).



Scheme 1: Strategies for aromatic degradation

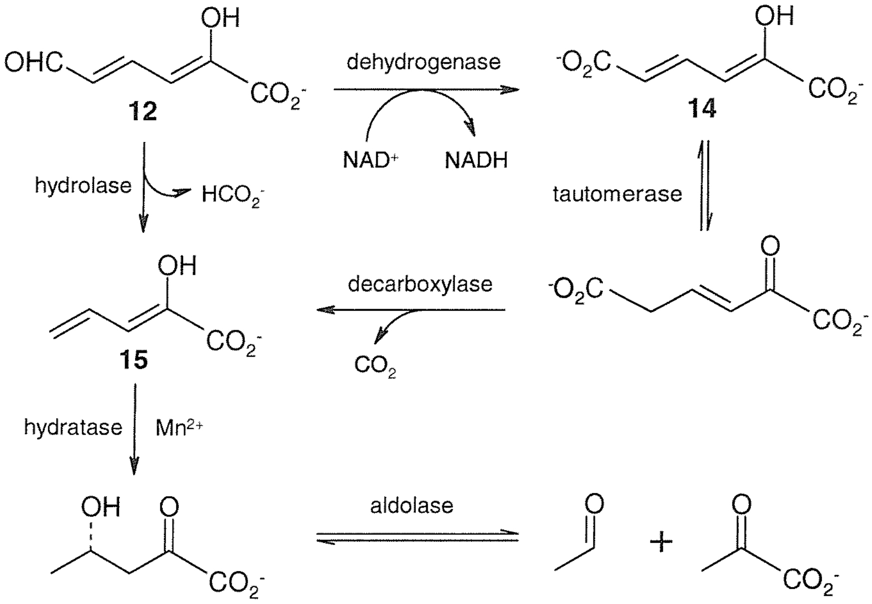
Non-haem iron-dependent dioxygenase enzymes catalyse the oxidative ring cleavage of the catechol intermediates in one of two ways: intradiol (or *ortho*) cleavage to give a muconic acid **11**; or extradiol (or *meta*) cleavage to give a hydroxymuconaldehydic acid derivative **12**, as shown in Scheme 2.



Scheme 2: Dioxygenase catalysed aromatic ring cleavage

Oxidative cleavage is also observed in some cases using a *p*-hydroxyquinone substrate, such as gentisic acid **13**. Thus there is a requirement for a 1,2- or 1,4-dihydroxyaromatic moiety for oxidative ring cleavage (Bugg and Winfield, 1998).

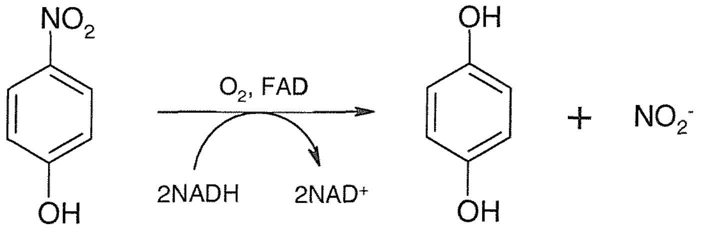
There are two branches for completion of the *meta*-cleavage pathway for catechol, as shown in Scheme 3 (Dagley, 1975). The ring fission product 2-hydroxymuconaldehyde acid (**12**) can be oxidised to give 2-hydroxymuconic acid (**14**) which is decarboxylated through the combined action of a tautomerase enzyme, and a decarboxylase enzyme to give **15**. Alternatively, 2-hydroxymuconaldehydic acid can be cleaved hydrolytically to give formic acid and 2-hydroxypentadienoic acid **15**. 2-hydroxypentadienoic acid is further broken down by hydration to 4-hydroxy-2-oxopentanoic acid, followed by aldolase-catalysed cleavage to give acetaldehyde and pyruvic acid.



Scheme 3: Catechol *meta*-cleavage pathway

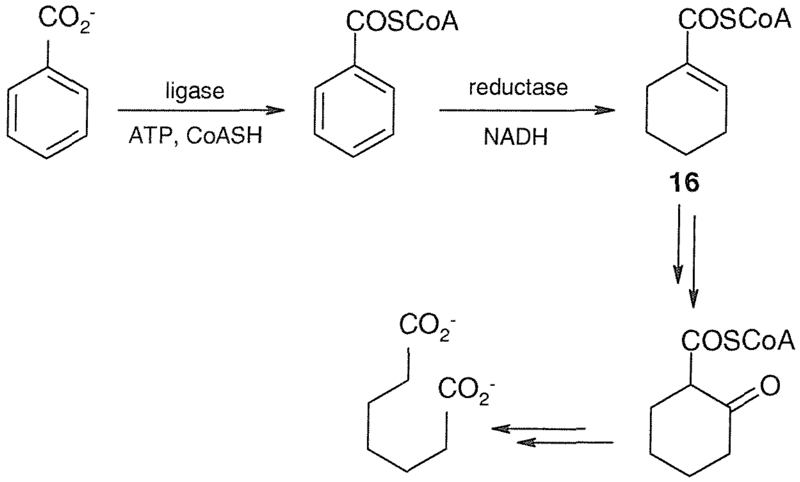
Longer aliphatic side chains attached to aromatic rings are typically broken down *via* β-oxidation into short side chains, followed by the *ortho*- or *meta*-cleavage pathways. The *ortho*-cleavage pathway can be used for the degradation of simple chlorinated aromatics by some *Pseudomonas* spp. Also, nitro aromatics are degraded by two strategies: (1) elimination of nitrite from a nitrophenol intermediate by a

FAD-dependent monooxygenase (Scheme 4); (2) reduction to an amino substituent, followed by oxidative cleavage (Suen and Spain, 1993).



Scheme 4: Enzymatic conversion of *p*-nitrophenol into hydroquinone

Aromatic compounds can also be degraded anaerobically by a reductive strategy (Scheme 5). In a denitrifying *Pseudomonas* strain, benzoic acid is converted into benzoyl CoA (Heider and Fuchs, 1997). The aromatic ring is then reduced to give cyclohexenyl CoA derivative **16** which is then a substrate for hydration and β-cleavage, analogous to β-oxidation.



Scheme 5: Anaerobic degradation of benzoic acid

1.5.1. Mechanistic aspects of the catechol dioxygenases

The intradiol and extradiol oxidative cleavage reactions are catalysed by non-haem iron-dependent dioxygenase enzymes, which incorporate both atoms of molecular oxygen into the ring fission product (Hayaishi, 1966).

The intradiol dioxygenases utilise a mononuclear non-haem iron (III) cofactor for substrate binding and catalysis, whereas the extradiol dioxygenases utilise a mononuclear iron (II) cofactor for substrate binding and catalysis. Thus, these two families of enzyme catalyse similar yet distinct oxidative cleavage reactions and have similar but distinct cofactor requirements (Bugg and Winfield, 1998).

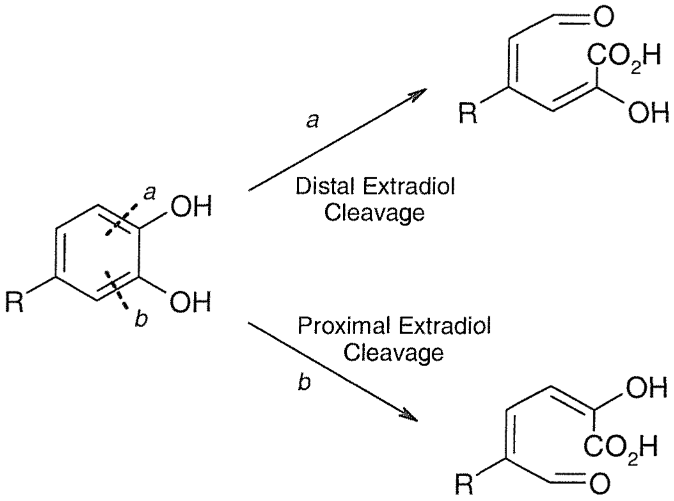
1.5.1.1. Intradiol catechol dioxygenases

The intradiol dioxygenases catalyse the transformation of catechols into *cis,cis*-muconic acids by direct incorporation of both atoms of dioxygen into the substrate resulting in aromatic ring cleavage between the two hydroxy groups. The best-studied of this group of enzymes are the catechol-1,2-dioxygenases (1,2-CTD) and protocatechuate-3,4-dioxygenases (3,4-PCD), which have been extensively characterised as a result of the rich spectroscopic properties of the non-haem iron (III) centre. Detailed structural information is now available for 3,4-PCD from *Pseudomonas putida*, whose X-ray crystal structure has been solved with and without bound substrate (Ohlendorf *et al.*, 1988; Ohlendorf *et al.*, 1994; Orville, Elango *et al.*, 1997; Orville *et al.*, 1997).

The catechol-1,2-dioxygenases have been divided into Type I and Type II enzymes (Dorn and Knackmuss, 1978). Type I are chromosomal enzymes characterised by their narrow substrate range and relatively high specificity for catechol. Type II are plasmid-mediated enzymes induced when the bacteria are grown on media containing halogenated aromatic compounds. They are relatively non-specific with a wider substrate range, catalysing faster breakdown of chloro and methyl substituted catechols but with lower catalytic efficiency for catechol than Type I. The protocatechuate-3,4-dioxygenase enzymes are chromosomally encoded proteins which display a narrow substrate specificity, but will metabolise halogenated catechols albeit with extremely reduced catalytic efficiency (Walsh and Ballou, 1983).

1.5.1.2. Extradiol catechol dioxygenases

The extradiol dioxygenases catalyse the transformation of catechols into 2-hydroxymuconaldehyde acids by direct incorporation of both atoms of dioxygen into the substrate. In contrast to the intradiol dioxygenases, ring fission occurs adjacent to the two hydroxyl groups of the catechol. Enzymes cleaving substituted catechols can be subdivided into the proximal and distal extradiol dioxygenases to distinguish the position of ring cleavage in relation to the substitituent group (Bugg and Winfield, 1998). Protocatechuate-2,3-dioxygenase (2,3-PCD) and protocatechuate-4,5-dioxygenase (4,5-PCD) provide examples of proximal and distal dioxygenases respectively (Scheme 6).



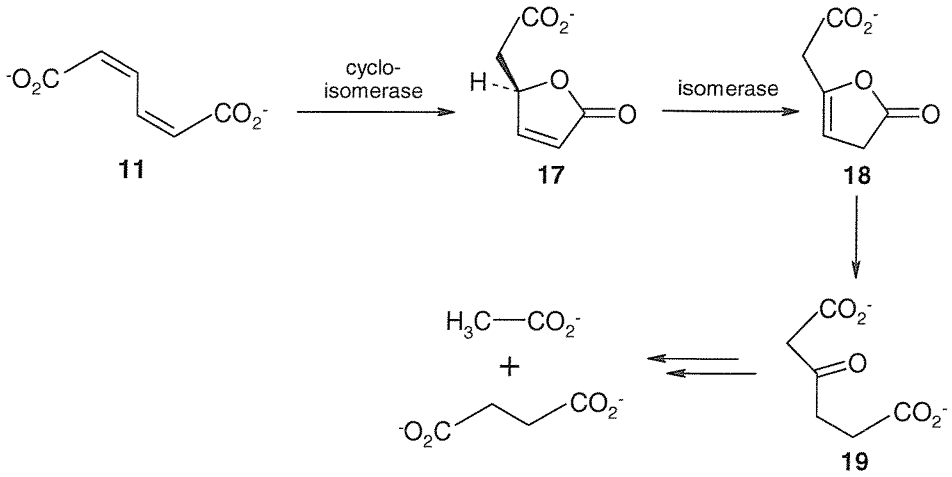
Scheme 6: The two modes of cleavage by extradiol dioxygenases

The crystal structures of 2,3-dihydroxybiphenyl-1,2-dioxygenase (BphC) from *Pseudomonas cepacia* LB400 and *Pseudomonas* sp. KKS102 have been solved (Han *et al.*, 1995; Senda *et al.*, 1996). There is no apparent relationship to the intradiol family of dioxygenases and comparison of the two reveal completely different structures.

1.5.2. Subsequent enzymatic steps in oxidative cleavage pathways

1.5.2.1. Enzymatic steps following intradiol cleavage

The products of intradiol (or *ortho*) cleavage are *cis,cis*-muconic acids. The *ortho*-cleavage pathway for catechol in *Pseudomonas putida* is illustrated in Scheme 7.

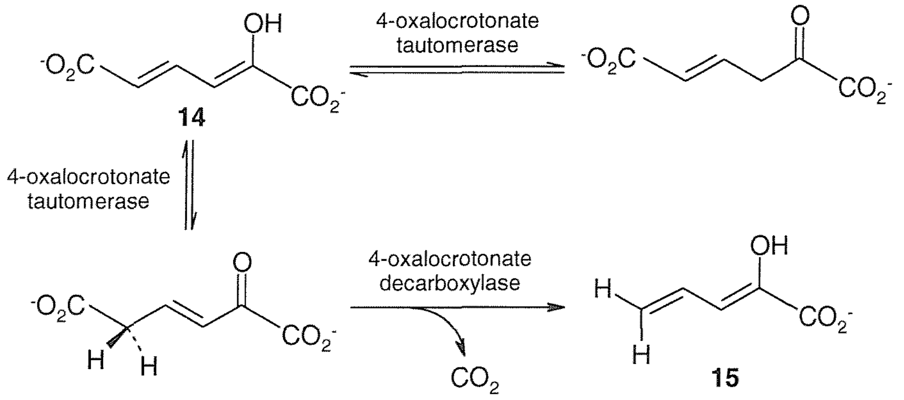


Scheme 7: Catechol *ortho*-cleavage pathway

The intradiol cleavage product **11** is cyclised enzymatically by muconate cycloisomerase to give muconolactone **17**, which is then converted by a Δ-isomerase enzyme into the unsaturated lactone **18**, which is hydrolytically cleaved to give 3-oxoadipic acid (β-ketoadipate, **19**). 3-oxodipic acid is subsequently cleaved to give acetic acid and succinic acid, which can be utilised for growth *via* the tricarboxylic acid (TCA) cycle (Bugg and Winfield, 1998).

1.5.2.2. Enzymatic steps following extradiol cleavage

There are two strategies used for the further breakdown of extradiol ring cleavage products, which have been illustrated in Scheme 3 for the catechol *meta*-cleavage pathway of *Pseudomonas putida.* The ring fission product can be oxidised by a NAD$^+$-dependent dehydrogenase to give 2-hydroxymuconic acid **14**. This dienol is a substrate for oxalocrotonate tautomerase, which catalyses the interconversion of 2-oxohex-3-ene-1,6-dioic acid and 2-oxohex-4-ene-1,6-dioic acid *via* the intermediacy of 2-hydroxymuconic acid **14** (Whitman *et al.*, 1991). The tautomerised product is then a substrate for decarboxylation *via* the enzyme 4-oxalocrotonate decarboxylase, to give 2-hydroxypentadienoic acid **15**, as shown in Scheme 8.



Scheme 8: Reaction of oxalocrotonate tautomerase and oxalocrotonate decarboxylase

The alternative strategy for *meta*-cleavage is *via* direct hydrolytic cleavage of a dienol ring fission product by a mechanistically unusual family of C-C bond hydrolases. This is followed by a series of enzymatically catalysed hydration and aldol-cleavage reactions to furnish the primary metabolites succinate, pyruvate and acetaldehyde (Bugg and Winfield, 1998).

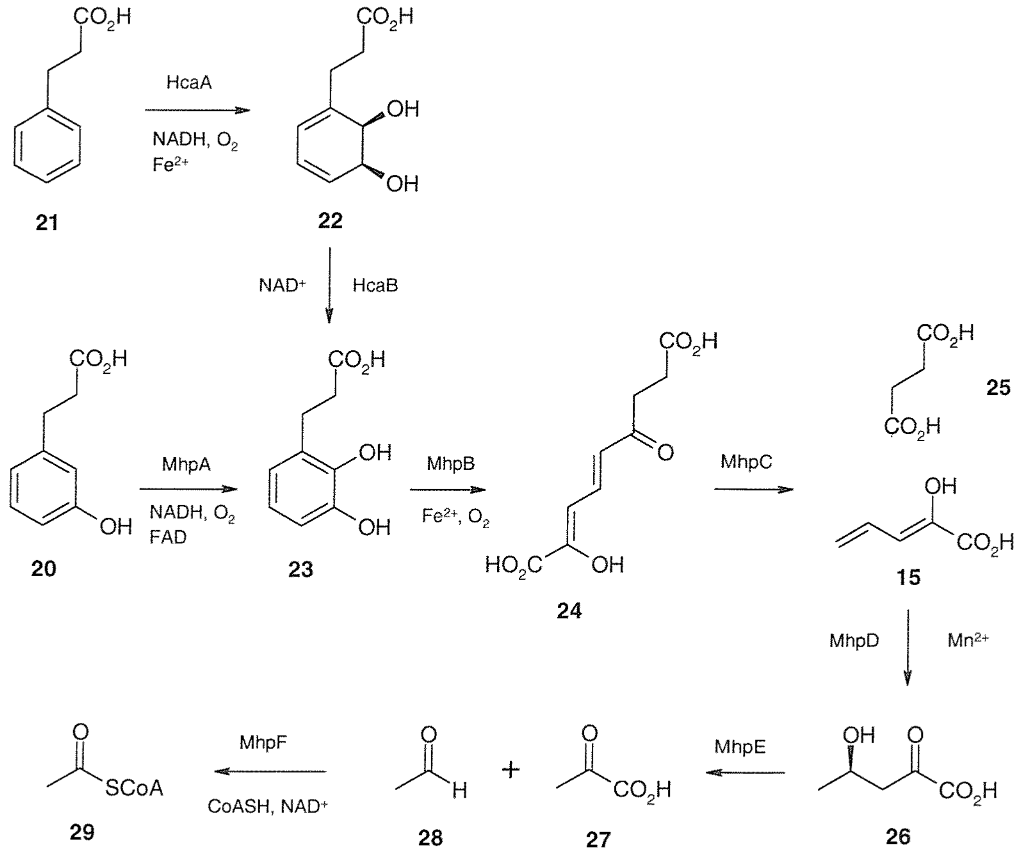
1.6. The catabolic pathway of phenylpropionic acid

Most early knowledge of bacterial aromatic catabolism stemmed from investigations with the genus *Pseudomonas*. Enteric organisms had also been studied on occasions, but the ability of *Escherichia coli* to degrade certain aromatics was not realised until a strain of this organism was grown with 3- and 4-hydroxyphenylacetic acids (Cooper and Skinner, 1980). It is now known that many laboratory strains and clinical isolates of *E. coli* can grow with various aromatic acids. In particular, *E. coli* K12 can grow with 3-phenylpropionic acid or 3-(3-hydroxyphenyl)propionic acid as sources of carbon (Burlingame and Chapman, 1983), and the degradation of these compounds has been studied extensively.

Phenylpropanoid compounds are widely available in natural environments, and they can originate from putrefaction of proteins in soil or as breakdown products of several constituents of plants, such as lignin, various oils, and resins (Andreoni, 1986; Barnes, 1997; Dagley, 1965; Ferrandez, 1997). Microbial catabolism of phenylpropanoid compounds plays an important role not only in the natural degradative cycle of these aromatic molecules but also in their practical applications such as wine-making, aging and storage (Cavin, 1997). In particular, degradation of cinnamic acid, 3-phenylpropionic acid, and their hydroxylated derivatives, has been reported in several bacteria, including *Acinetobacter* sp. (Dagley, 1965), *Pseudomonas* sp. (Blakey and Simpson, 1984; Andreoni, 1986; Strickland, 1973), *Arthobacter* sp. (Strickland, 1973), *E. coli* (Cooper and Skinner, 1980; Burlingame, 1983), *Rhodococcus globelerus* (Barnes, 1997), *Achromobacter* (Dagley, 1965), *Alcaligenes eutrophus* (Kabisch and Fortnagel, 1990) and *Conomonas testosteroni* (Arai *et al.*, 1999).

Since the early 1980s at least six aromatic acids have been shown to be degraded by *E. coli* including 3-phenylpropionic acid (or hydroxycinnamic acid, HCA) and 3-(3-hydroxyphenyl)propionic acid (or *m*-hydroxyphenylpropionic acid, MHP). *E. coli* living in the animal gut encounters phenylpropionic acid and its derivatives as a result of the action of intestinal microflora on plant constituents, the amino acids phenylalanine and tyrosine, and some of their metabolites (Burlingame and Chapman, 1983).

The phenylpropionate degradative pathway is known to be a *meta*-cleavage pathway, where the C-C bond to be broken is situated adjacent to the two hydroxyl groups, and the product of ring fission degraded by the hydrolytic route (Scheme 9). A total of eight enzymes are reported to be included, making this one of the longest catabolic sequences described for *E. coli* (Burlingame *et al.*, 1986).



Scheme 9: The phenylpropionate catabolic pathway of *E. coli*

The catabolic metabolite 2,3-dihydroxyphenylpropionic acid (DHP, **23**) can be formed by two separate routes. Biochemical studies and the isolation and characterisation of mutants defective in the catabolism of 3-phenylpropionic acid and 3-(3-hydroxyphenyl)propionic acid revealed that in *E. coli* the aerobic degradation of these compounds proceeds by two initially separate routes that converge into DHP (Bugg, 1993; Burlingame and Chapman, 1983; Burlingame *et al.*, 1986).

3-phenylpropionic acid (PP, **21**) can be dihydroxylated by the dioxygenase enzyme HcaA to afford the *cis*-dihydrodiol (**22**), which is subsequently oxidised by the dehydrogenase enzyme HcaB to yield DHP. The alternative branch involves the monohydroxylation of 3-(3-hydroxyphenyl)propionic acid (3HPP, **20**) by the flavoprotein monooxygenase, MhpA, directly producing DHP. The benzene nucleus of DHP can then be cleaved by the dioxygenase MhpB, to produce the dienol ring fission product (RFP) 2-hydroxy-6-oxo-nona-2,4-diene-1,9-dioic acid **24**. Subsequent degradation of **24** is initiated by the hydrolase enzyme MhpC which furnishes succinic acid (**25**) and the common intermediate 2-hydroxypenta-2,4-dienoic acid (HPD, **15**). This intermediate, which is common to many *meta*-cleavage pathways, is then hydrated by MhpD to afford 4-hydroxy-2-ketopentanoic acid (HKP, **26**) prior to a reverse aldol cleavage reaction catalysed by MhpE. Products of this reaction are pyruvic acid (**27**) and acetaldehyde **28**.

Recent studies on the genetics of the phenylpropionic acid pathway have highlighted the presence of one further open reading frame associated with the phenylpropionate catabolic operon (Ferrandez *et al.*, 1997). This gene has been sequenced and by comparison with homologous sequences it has been proposed that the protein encoded by this gene is acetaldehyde dehydrogenase (acylating), MhpF. The reaction it is thought to catalyse is the conversion of **28** to acetyl-CoA **29**.

Enzyme assays and respirometry have shown that the syntheses of enzymes required to convert the two initial growth substrates, PP and 3HPP, into DHP are inducible and under separate control (Burlingame and Chapman, 1983; Burlingame *et al.*, 1986).

The complete nucleotide sequence of the gene cluster encoding the 3HPP catabolic pathway of *E. coli* K12 has been reported (Ferrandez *et al.*, 1997). Sequence analysis revealed the existence of eight genes that map at minute 8 of the chromosome, between the *lac* and *hemB* regions. Six enzyme-encoding genes account for a flavin-type monooxygenase (*mhpA*), the extradiol deoxygenase (*mhpB*), and the *meta*-cleavage pathway (*mhpCDFE*). The order of these catabolic genes, with the sole exception of *mhpF*, parallels that of the enzymatic steps of the pathway. Enzymes that catalyse the early reactions of the pathway, MhpA and MhpB, showed the lowest level of sequence similarity to analogous enzymes of other aromatic catabolic pathways. However, the genes *mhpCDFE* present the same organisation and appear to be homologous to the *Pseudomonas xylFJQK*, *dmpDEFG*, and *nahNLOM* genes, from the toluene-xylene, phenol and naphthalene catabolic pathways respectively (Platt *et al.*, 1995; Williams and Sayers, 1994). As these three *meta*-cleavage pathways have been shown to share a common ancestry (Williams and Sayers, 1994), the *mhpCDFE* genes represent the first example of this type of catabolic module outside of the genus *Pseudomonas*. Interestingly, while the three *Pseudomonas meta*-cleavage pathways are plasmid encoded, the *E. coli mhp* pathway is located in a chromosomal region.

The *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of PP in *E. coli* K12 has also been identified, cloned and sequenced (Diaz *et al.*, 1998). This cluster maps at minute 57.5 of the chromosome and is composed of five catabolic genes arranged as a putative operon (*hcaA1A2CBD*). Sequence comparisons revealed that while *hcaA1A2CD* genes encode the four subunits of the 3-phenypropioniate dioxygenase, the *hcaB* gene encodes for the corresponding *cis*-dihydrodiol dehydrogenase.

The HcaA1A2CD and HcaB enzymes are not only able to oxidise PP to DHP but also cinnamic acid (CI) to its corresponding 2,3-dihydroxy derivative. Further catabolism of DHP requires the *mhp*-encoded *meta*-fission pathway for the mineralisation of 3HPP (Ferrandez *et al.*, 1997).

The *hca* cluster maps at min 57.5 of the *E. coli* chromosome and therefore far from min 8, where the *mhp* cluster responsible for DHP degradation is located. Interestingly, the third aromatic catabolic pathway characterised so far in *E. coli* at the molecular level, the *hpa* cluster for 4-hydroxyphenylacetate degradation, was shown to map at minute 98 (Prieto *et al.*, 1996). Thus, while in some *Pseudomonas* and *Acinetobacter* species a supraoperonic clustering of the aromatic catabolic genes has been observed in a limited region of the chromosome (Eaton, 1996; Eaton, 1997; Williams and Shaw, 1997), in *E. coli* the aromatic catabolic clusters are dispersed throughout the genome.

It is known that *E. coli* K12 is also able to grow with 3-hydroxycinnamic acid (3HCI) as the sole carbon and energy source (Burlingame and Chapman, 1983). Since growth with 3HCI induces the synthesis of enzymes MhpA and MhpB, responsible for initial attack upon 3HPP, it was suggested that the same enzymes are used for catabolising these two compounds. Expression in *Salmonella typhimurium* of the *mhp* genes alone or in combination with the *hca* cluster allowed the growth of the recombinant bacteria in 3HCI and CI, respectively. Thus, the convergent *mhp*- and *hca*-encoded pathways are also functional in *S. typhimurium*, and they are responsible for the catabolism of different phenylpropanoid compounds (3HPP, 3HCI, PP and CI) widely available in nature (Diaz *et al.*, 1998).

1.7. 2-hydroxy-6-oxo-nona-2,4-diene-1,9-dioic acid 5,6-hydrolase (MhpC)

Hydrolytic cleavage of a carbon-carbon bond adjacent to a carbonyl is a rare class of reaction in chemistry and biochemistry. There are a small number of enzyme-catalysed reactions which involve hydrolytic C-C cleavage, including kynureninase from *Pseudomonas fluorescens* (Philips and Dua, 1991) and dioxo acid hydrolase from beef liver (Hsiang *et al.*, 1972). In the case of the bacterial *meta*-cleavage pathways for the degradation of aromatic compounds, ring fission products containing a δ-keto dienol are formed by oxidative cleavage of catechol substrates (Bugg and Winfield, 1998). These ring fission products are then substrates for hydrolytic C-C cleavage by a family of hydrolase enzymes, as illustrated in Figure 1.

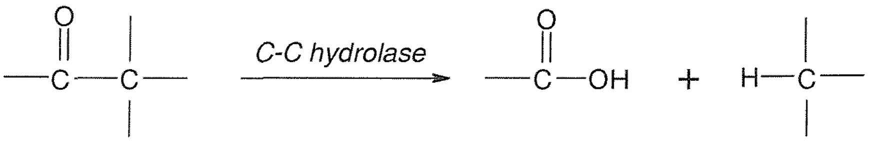
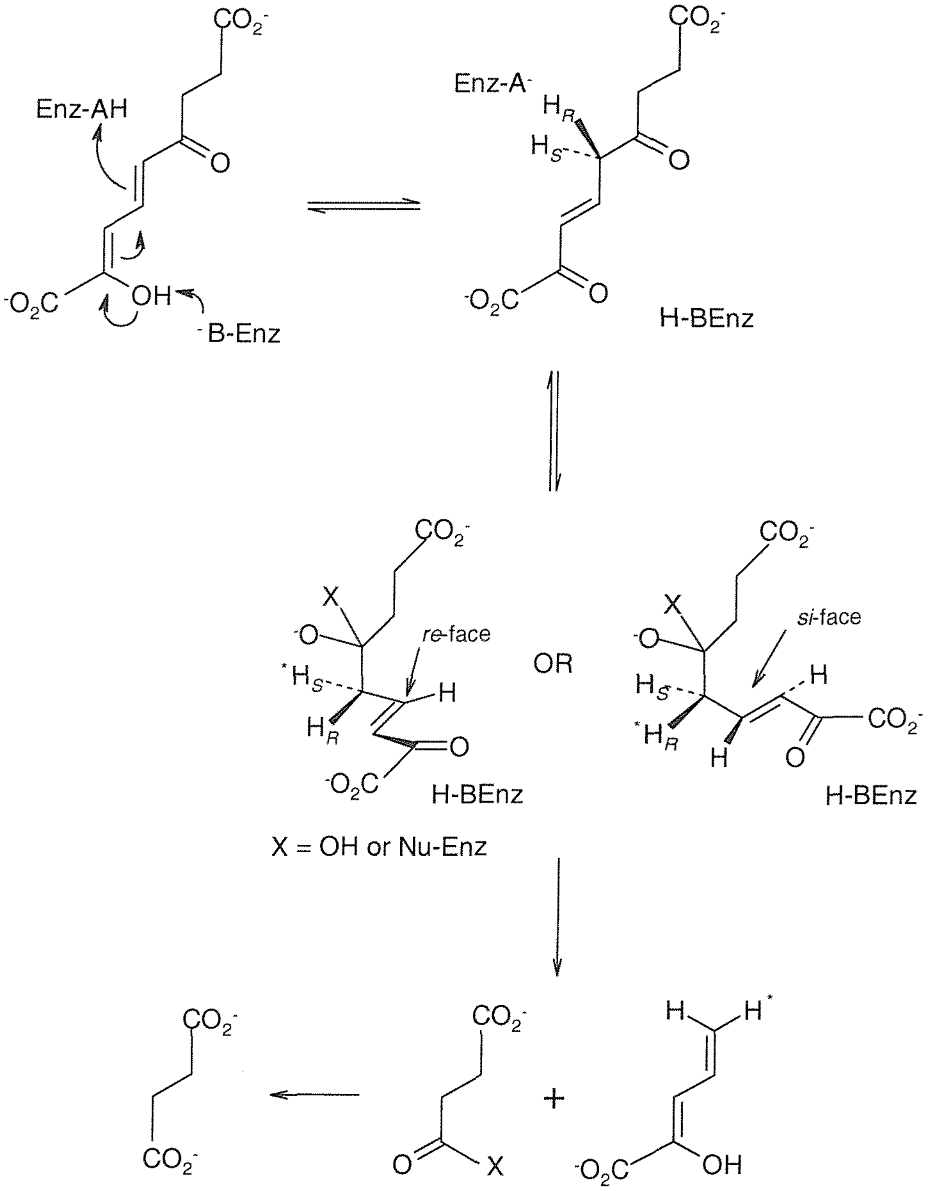


Figure 1: Reaction catalysed by C-C bond hydrolases

Detailed studies of the reaction catalysed by MhpC on the phenylpropionate catabolic pathway of *E. coli* have provided some insight into the catalytic mechanism (Lam and Bugg, 1997; Lam and Bugg, 1994; Henderson and Bugg, 1997). Extradiol cleavage of 2,3-dihydroxyphenylpropionate by dioxygenase MhpB gives the stable dienol 2-hydroxy-6-oxonona-2,4-diene-1,9-dioic acid, which has been found to exist in the *trans,transoid* conformation, as shown in Scheme 10 (Lam and Bugg, 1994).



Scheme 10: Mechanistic and stereochemical course of reaction for MhpC

Hydrolytic cleavage of the C5-C6 bond is catalysed by hydrolase MhpC, a 29kDa homodimeric protein requiring no cofactors for activity. Isotope labelling experiments have demonstrated that MhpC inserts one atom of ^{18}O from $H_2{}^{18}O$ into the first product succinic acid (Sanovisin *et al.*, 1995), and inserts a ^{2}H atom from $^{2}H_2O$ into the H_{5E} position of the second product 2-hydroxypentadienoic acid (Lam and Bugg, 1997).

The catalytic mechanism of the MhpC-catalysed reaction has been examined by stopped flow pre-steady-state kinetics (Henderson and Bugg, 1997). The disappearance of substrate at 394nm at pH5.0 proceeds at a rate constant, *k*, of 10 s^{-1}, whereas the appearance of product 2-hydroxypentadienoic acid at 270nm proceeds at a rate constant, *k*, of 3 s^{-1}, indicating the existence of a kinetically competent intermediate. Evidence from ^2H isotope exchange is consistent with an initial ketonisation of the ring fission product (RFP) to give a keto-intermediate (RFP^K), formed by protonation of the dienol at C-5 (Lam and Bugg, 1997). The overall stereochemical course of the reaction, as shown in scheme 10, is consistent either with protonation at the C-5 *pro-R* position, followed by fragmentation onto the *si*-face of the 3,4-double bond; or protonation at the C-5 *pro-S* position, followed by fragmentation onto the *re*-face of the 3,4-double bond (Bugg and Winfield, 1998). The keto-intermediate contains an α/β-unsaturated ketone functional group, which serves as an electron sink for a carbanion arising from C-C cleavage. Mechanistic proposals for the C-C cleavage step have involved attack upon the C-6 carbonyl of either water or an active site nucleophile, as illustrated in scheme 10, giving rise in the latter case to a subsequent acyl enzyme intermediate (Lam and Bugg, 1997; Lam and Bugg, 1994; Henderson and Bugg, 1997).

The third gene of the *mhp* cluster, *mhpC*, encodes a protein of 31,937 Da whose deduced amino acid sequence shows significant identity to that of hydrolases cleaving carbon-carbon bonds of the *meta*-cleavage products of aromatic compounds (Ferrandez *et al.*, 1997). It shows 50% identity with gene products for BphD and BphD_LB from the biphenyl pathways of *Pseudomonas* sp. strains KKS102 (Kikuchi *et al.*, 1994) and LB400 (Hofer *et al.*, 1994), respectively. Also, 35% identity with XylF, from the toluene-xylene pathway of *P. putida* mt-2 (Harayana and Rekik, 1993); 34% identity with TodF and DmpD from the toluene pathway of *P. putida* F1 (Wong *et al.*, 1995) and the phenol-dimethylphenol pathway of *Pseudomonas* sp. strain CF600 (Shingler *et al.*, 1992), respectively; and 27% identity with CmtE from the *p*-cumate pathway of *P. putida* F1 (Eaton, 1996). Hence, the MhpC protein was ascribed to the 2-hydroxy-6-oxonona-2,4-diene-1,9-dioate hydrolase of the 3HPP pathway (Ferrandez *et al.*, 1997).

The enzyme MhpC also hydrolyses the 3-methylcatechol ring fission product, the natural substrate of the XylF hydrolase. Sequence alignments between MhpC and isofunctional enzymes from the *meta*-cleavage biphenyl (BphD), xylene (XylF), toluene (TodF), phenol (DmpD), and *p*-cumate (CmtE) biodegradative pathways strongly suggest that MhpC is a serine hydrolase that belongs to the α/β-hydrolase fold family (Diaz and Timmis, 1995). These comparisons also point to residues S114, D239, and H267 as the putative catalytic triad of MhpC (Ferrandez *et al.*, 1997). The sequence GNSMGG around the potential active-site serine of MhpC fits very well in the nucleophilic motif present in all members of the α/β-hydrolase fold family (Ollis *et al.*, 1992). Moreover, the conserved dipeptide HG at the NH_2 terminus of the MhpC molecule can be tentatively assigned as the central dipeptide that characterises the oxyanion hole in the α/β-fold hydrolases (Diaz and Timmis, 1995).

Extensive sequence similarity has also been detected between the C-C hydrolases and the sequences of *P. putida* atropinesterase, *Moraxella* sp. lipase 3, *Pseudomonas* sp. KWI-56 esterase V, and *Lactobacillus delbrueckii* proline iminopeptidase (Hofer *et al.*, 1993; Diaz and Timmis, 1995). The conserved sequence motif G-X-S-X-G observed at the position of the suspected serine nucleophile, is also found in the serine proteases, as in Figure 2.

...H L L	G N **S** M G	G H S...	MhpC – C-C hydrolase from *E. coli* phenylpropionate catabolic pathway
...H L V	G N **S** M G	G A G...	BphD – C-C hydrolase from *Pseudomonas* sp. KKS102 biphenyl catabolic pathway
...H L V	G N **S** M G	G A T...	BphD – C-C hydrolase from *Pseudomonas* sp. LB400 biphenyl catabolic pathway
...H L V	G N **S** M G	G A G...	BphD – C-C hydrolase from *Comomonas testosteroni* TK102 biphenyl catabolic pathway
...H L V	G N **S** M G	G A T...	BphD – C-C hydrolase from *P. putida* KF715 biphenyl catabolic pathway
...H I V	G N **S** M G	G A V...	BphD – C-C hydrolase from *Rhodococcus* sp. RHA1 biphenyl catabolic pathway
...G F V	G N **S** M G	G H T...	BpdF – C-C hydrolase from *Rhodococcus* sp. M5 biphenyl catabolic pathway
...D L V	G N **S** F G	G G I...	DmpD – C-C hydrolase from *Pseudomonas* sp. CF600 phenol catabolic pathway
...D I V	G N **S** F G	G G L...	XylF – C-C hydrolase from *P. putida* xylene catabolic pathway
...D L V	G N **S** F G	G A L...	TodF – C-C hydrolase from *P. putida* F1 toluene catabolic pathway
...S L L	G W **S** D G	G I T...	Bph-rp – Biphenyl hydrolase related protein from a human breast carcinoma
...T V I	G H **S** M G	S M T...	TPES – Atropinesterase from *P. putida*
...S C M	G D **S** G G	P L V...	CTRA – Chymotrypsinogen A from *Bos taurus* (Bovine)

Figure 2: Sequence alignment showing the homologous 'nucleophile elbow' motif around the proposed catalytic serine residue.

There is a large body of biochemical and crystallographic evidence collected for members of the α/β-hydrolase fold family to suggest that the active site serine group acts as a nucleophile (Schrag and Cygler, 1997), although acyl enzyme intermediates have only been characterised using unnatural substrates such as *p*-nitrophenylacetate (Chapus *et al.*, 1976).

In the C-C hydrolase family, catalytic activity of *P. putida* XylF was lost upon treatment with phenylmethylsulfonyl fluoride (PMSF) or di-isopropylfluorophosphate (DIFP), both active-site-serine specific reagents (Diaz and Timmis, 1995). Furthermore, serine-to-alanine site-directed mutants in this position of *P. putida* XylF or *Conomonas testosteroni* BphD were found to be devoid of catalytic activity (Diaz and Timmis, 1995; Ahmad *et al.*, 1995). Thus, the available biochemical evidence heavily favours a mechanism involving an active site serine nucleophile.

However, no stable acyl enzyme intermediate (X=Nu-Enz) could be detected in the MhpC reaction cycle by stopped flow kinetics (Henderson and Bugg, 1997), even at pH4.0, under which conditions the serine proteases accumulate an acyl-enzyme intermediate (Figure 3) at their active site (Schrag and Cygler, 1997; Chapus *et al.*, 1976).

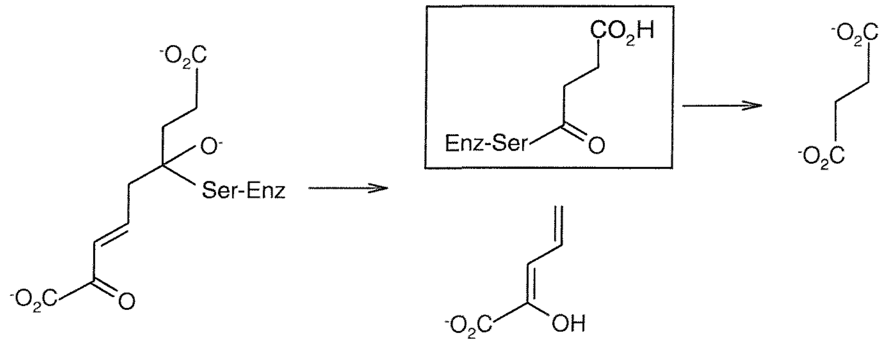


Figure 3: Proposed mechanism for MhpC, illustrating the involvement of an acyl-enzyme intermediate.

Attempts to trap the putative acyl-enzyme intermediate using hydroxylamine failed to give any hydroxamic acid product (Lam and Bugg, 1997). Synthesis of *p*-nitrophenylsuccinate as an alternative substrate for pre-steady-state kinetics revealed that this compound was not a substrate for MhpC (Lam and Bugg, 1997).

Despite the implication of an active site serine residue in the α/β-hydrolase fold family, attempts to verify a putative acyl-enzyme intermediate by radiochemical trapping methods using a ^{14}C-labelled substrate yielded a stoichiometry of <1% covalent intermediate, which could be accounted for by non-enzymatic processes (Fleming *et al.*, 2000). In contrast, incorporation of 5-6% of two atoms of ^{18}O from $H_2{}^{18}O$ into succinic acid was observed using the natural substrate, consistent with the reversible formation of a *gem*-diol intermediate (Fleming *et al.*, 2000). Furthermore, time-dependent incorporation of ^{18}O from $H_2{}^{18}O$ into the carbonyl group of a non-hydrolysable analogue 4-ketonona-1,9-dioic acid (KNDA, Figure 4), was observed in the presence of MhpC, consistent with enzyme-catalysed attack of water at the ketone carbonyl (Fleming *et al.*, 2000).

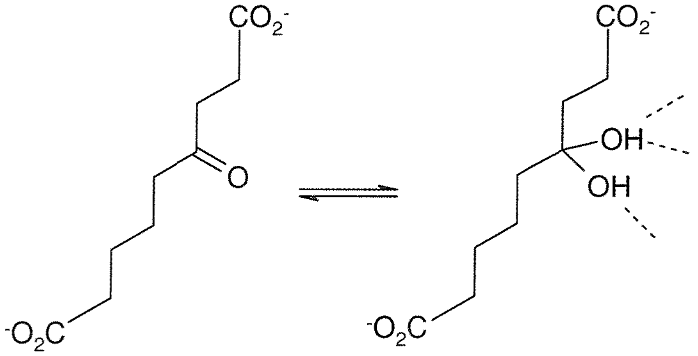


Figure 4: Proposed reaction of substrate analogue KNDA with MhpC to form a *gem*-diol intermediate

Therefore, these latest results favour a catalytic mechanism involving base-catalysed attack of water (Figure 5) as seen for the aspartic proteases (Veerapandian, 1992), rather than nucleophilic attack of an active site serine.

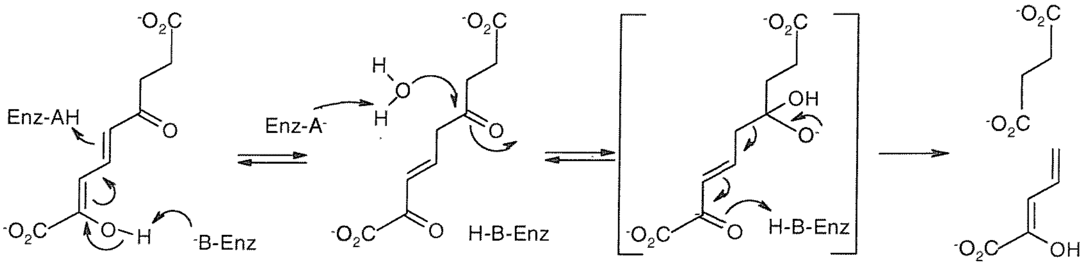


Figure 5: Proposed general base mechanism for the MhpC catalysed reaction

1.7.1. Research Aims

The primary purpose of this research is to help understand the catalytic mechanism of *E. coli* MhpC. Firstly, to determine if MhpC is a member of the α/β-hydrolase fold family by solving its three-dimensional structure using X-ray crystallography; and secondly, once the structure has been found, to conduct a more detailed investigation of the mechanism of action of this C-C hydrolase, and discover whether the active site nucleophile is the postulated serine residue or an activated water molecule. It is hoped that by determining the mechanism of action of this unusual class of enzyme, mutant enzymes capable of novel biotransformations could be produced and also that it would contribute to the understanding of the biodegradation of environmental pollutants.

1.8. The α/β hydrolase fold superfamily

The α/β hydrolase fold family of proteins is an important, diverse, widespread group of enzymes, where all the members are related by divergent evolution and yet where – unlike the serine proteases – the level of divergence is extremely high and the structural similarity between related proteins is often very low. The fold is common to several hydrolytic enzymes of widely differing phylogenetic origin and catalytic function.

The core of each enzyme is similar: an α/β sheet, not barrel, of eight β-sheets connected by α-helices. They all have a catalytic triad, the elements of which are borne on loops and are the most structurally conserved feature of this family. Only the histidine in the nucleophile-histidine-acid catalytic triad is completely conserved.

In the early 1990's, five apparently unrelated hydrolases turned out to have the same protein fold, named the α/β hydrolase fold (Ollis *et al.*, 1992). The enzymes whose structures were compared were acetylcholinesterase (AchE) from *Torpedo californica* (Sussman *et al.*, 1991), carboxypeptidase II (CPW) from wheat – a member of a large family of serine carboxypeptidases (Thomas *et al.*, 1990), dienelactone hydrolase (DLH) from *Pseudomonas* sp. B13 (Pathak and Ollis, 1990), haloalkane dehalogenase (HAL) from *Xanthobacter autotrophicus* (Francken *et al.*, 1991), and lipase (GLP) from *Geotrichum candidum* (Selvrag *et al.*, 1991). The only features in common were a few aspects of structure and reaction: the sequences were unrelated; they did not operate on similar substrates; nor did they use the same nucleophile (Table 1).

Protein	Subunit		Oligomer structure	Modifications	Source	Substrate
	Mol. Wt. (kDa)	Sequence length (amino acids)				
Acetylcholine esterase (AChE)	60	537	Dimer	Glycosylated	Fish: *Torpedo*	Acetylcholine
Carboxypeptidase II (CPW)	60	423	Dimer	Glycosylated	Plant: wheat	Peptide
Dienelactone hydrolase (DLH)	25	236	Monomer	None	Bacterium: *Pseudomonas*	Dienelactone
Dehalogenase (HAL)	35	310	Monomer	None	Bacterium: *Xanthobacter*	1,2-dichloroethane
Lipase (GLP)	60	544	Monomer	Glycosylated	Fungus: *Geotrichum*	Triacylglycerol

Table 1: Comparison of the five proteins first discovered to possess the α/β hydrolase fold, showing the differences in their physical properties and sources. *Source*: Ollis *et. al.*, 1992.

The α/β hydrolase fold can be said to have the most plastic of protein folds, tolerating large insertions into a single-domain protein, so that domain size can vary from as little as 197 residues in *Fusarium solani* cutinase (FsCUT) to 582 residues in mouse acetylcholinesterase (mACE) as shown in Figure 6 (Heikinheimo *et al.*, 1999). There have been upward of 48 α/β hydrolase fold structures solved thus far, with more being discovered constantly (Cousin *et al.*, 1998).

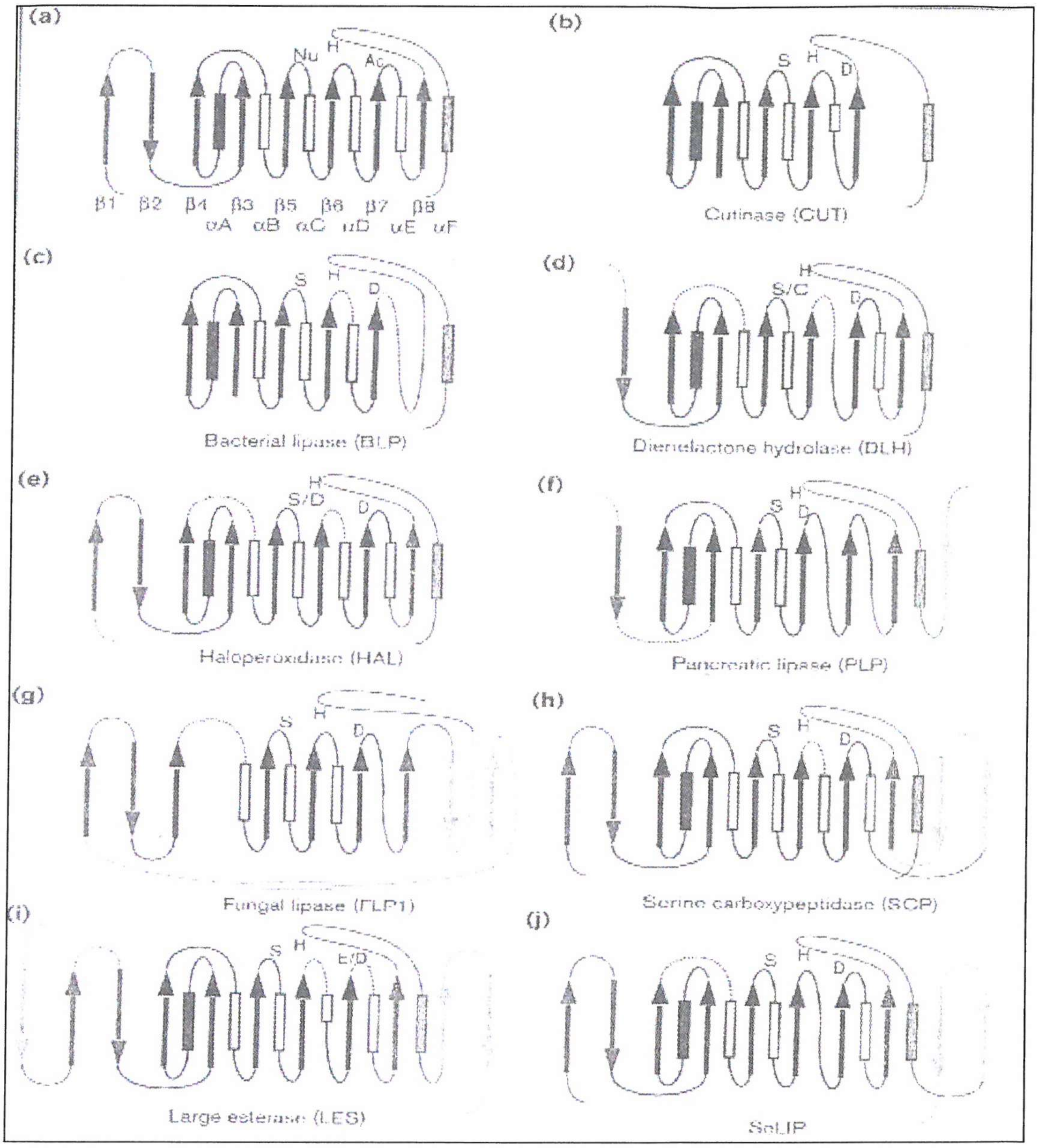


Figure 6: Topological diagrams of the canonical, minimal and variations of the α/β hydrolase fold. The minimal fold is indicated in black, the canonical fold in dark grey, and extensions to the core beyond the canonical fold are shown in light grey. Strands are shown by arrows and helices by rectangles. Filled rectangles are below the plane of the β sheet (on the convex surface); open rectangles are above (on the concave surface). *Source:* Heikinheimo *et al.*, 1999.

1.8.1. Features of an α/β hydrolase fold protein

1.8.1.1. The canonical structure

The canonical α/β hydrolase fold is an eight-stranded mostly parallel α/β structure with strand β2 antiparallel to the rest, and the connection of the strands being +1,+2, -1x, +2x, +1x, +1x (strand order 12435678; Figure 7). The substrate-binding crevice is at the cross-over connection on strand β5; and, as is typical of the active sites in α/β proteins, the nucleophile is also located at the strand cross-over point of the parallel β-sheet (Branden, 1980).

The sheet is highly twisted and bent so that it forms a half-barrel with the first and the last strands at almost 90° to one another (Ollis *et al.*, 1992; Heikinheimo *et al.*, 1999). In the canonical protein, of which *Pseudomonas fluorescens* carboxylesterase (PfCES) is an example (Kim *et al.*, 1997), the strands are connected by α-helices A to F. The first and last helices (αA and αF) pack onto the convex face of the half-barrel, whereas helices αB-αE pack onto the concave surface (Figure 7). The structurally conserved helices occur immediately before the structurally conserved β-strand; the model being α-turn-β, not β-turn-α. Insertions occur after one or more of the β-strands β3, β4, β6 and β7. Insertions following β8 are also common, but can occur before or after αF. The substrate-binding domains are often made of several insertions and form a cap on top of the active site. Many, but not all, of the α/β hydrolase fold enzymes have a left-handed cross-over following strand β8 (Heikinheimo *et al.*, 1999).

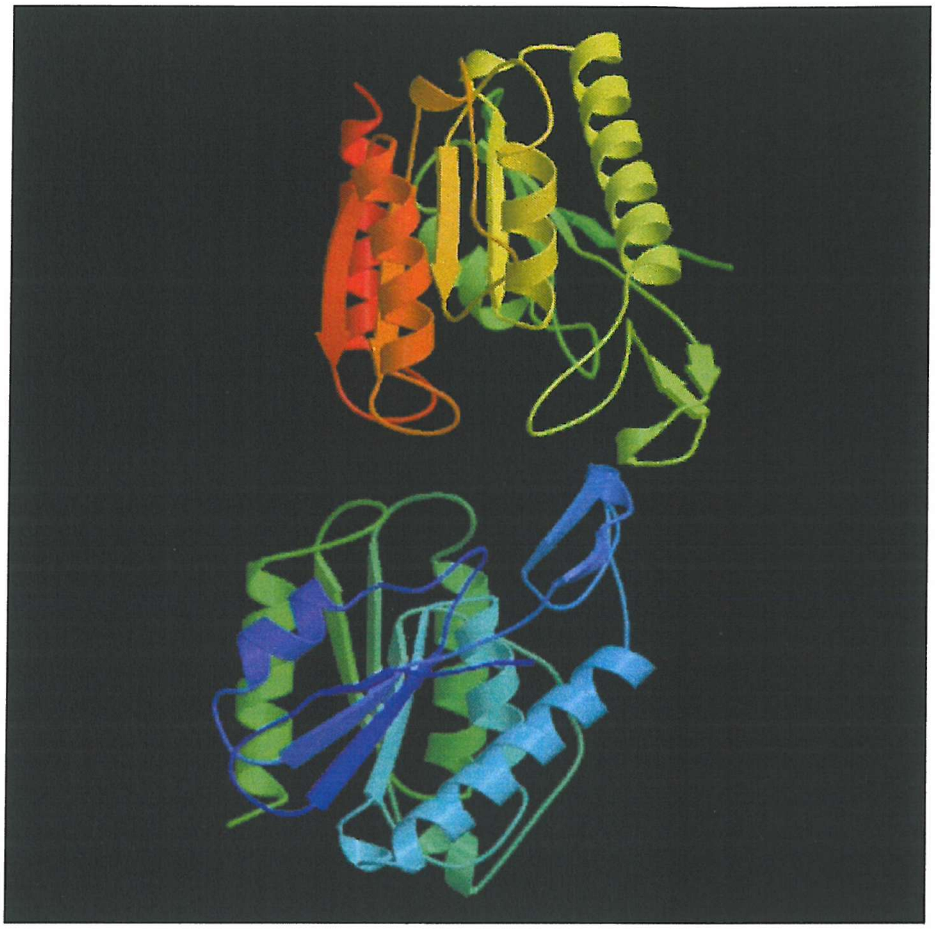


Figure 7: Structure of *Pseudomonas fluorescens* carboxyl esterase (PfCES), which is an example of a canonical α/β hydrolase fold protein. The model is shown from two orientations: perpendicular to (top) and parallel with (bottom) the β-sheet.

1.8.1.2. Enzymatic properties of the α/β hydrolase fold proteins

The catalytic triad – nucleophile, acid and a histidine – forms the basis of the enzymatic properties of the α/β hydrolases. Unlike the other protease families the nature of the nucleophile varies: it can be serine, cysteine or aspartate. However, it is always the central residue in an extremely sharp, γ-like (Matthews, 1972), turn between strand β5 and helix αC. The strand-nucleophile-helix feature – the 'nucleophile elbow' – is the most conserved structure within the α/β hydrolase fold (Ollis *et al.*, 1992).

The potential steric problems expected to arise at the nucleophile elbow are avoided by having residues with small side-chains at four key positions:

(i) Nu –2 (on strand β5);

(ii) Nu +2 (on helix αC);

(iii) Nu +3 (on helix αC) to avoid steric overlap with sheet 4; and

(iv) The residue immediately following sheet 6: its side-chain points towards helix C and so must be small.

One, and usually both, of the residues at Nu –2 and Nu +2 must be glycine. Consequently, the sequence around the nucleophile is Sm-X-Nu-X-Sm-Sm (where Sm = small residue); this is different from G-X-Nu-G-G, the pattern seen in the trypsin-like proteases (Brenner, 1988).

The second member of the triad, which can be either Asp or Glu, is on a loop following strand β7. The α/β hydrolase fold family includes the only example of enzymes in which glutamate is utilised as the acid in the triad. While the sequence of the aspartatic acid turn varies, the local structure – from the end of strand β7 to two residues past the acid – is almost identical. At the end of strand seven, two reverse turns occur in the space of seven amino acids with the acid as the joining residue: last residue of the first turn and first residue of the second turn. An hydrogen bond forms between the Asp side-chain oxygen distal to the His and the backbone nitrogen of the residue two amino acids towards the C-terminus from the Asp. Thus, the second reverse turn stabilises the position of the Asp side-chain, a role fulfilled by Ser214 in the serine proteases (Meyer *et al.*, 1988).

The loop bearing the Glu and the loop bearing the Asp in other α/β hydrolase fold enzymes are quite similar. In the glutamic peptides, the Glu occupies the same position as the Asp does: it is the fourth residue in a reverse turn. However, in both types of 'acid turn', the peptide is structurally conserved only as far as the acid and the residues which form hydrogen bonds to the acid residue side-chain. The Glu side-chain distal to the His appears to form a hydrogen bond with the backbone nitrogen of the residue three amino acids before it in the sequence (Ollis *et al.*, 1992).

The structure of the peptide containing the catalytic triad histidine in α/β hydrolase fold enzymes is highly conserved despite the fact that their topologies in this region may be completely different. The His residue, which lies at the end of strand β8, comprises a turn, one amino acid and then the histidine which is the first residue in a reverse turn, but the structural conservation only extends up to the histidine. In some peptides of the α/β hydrolase fold family, extended loops bring the histidine to a position appropriate for forming the triad, and similar to that in the other proteins (Ollis *et al.*, 1992).

1.8.1.3. Active site design

The twist of the central β-sheet allows short pieces of peptides to bring the triad residues together to form a catalytically active enzyme and creates good geometry for the 'oxyanion hole' (Henderson, 1970); the oxyanion residues usually being located on strands β5 and β3 (Heikinheimo *et al.*, 1999).

In the α/β hydrolase fold family the catalytic triad residues always occur in the same order in the primary sequence: nucleophile, acid, histidine; this order is different from that observed in any other proteins containing catalytic triads. The catalytic triad residues in all the α/β hydrolase fold enzymes have similar topological and three-dimensional positions despite the lack of sequence homology. The parallel β-sheet has the normal left-handed twist (Richardson, 1976) found in α/β proteins, which means that the very short loops beyond the ends of strand β5 (nucleophile), β7 (acid) and β8 (histidine) are sufficient to position these residues correctly to form a catalytic triad (Ollis *et al.*, 1992).

The nucleophile and histidine of a catalytic triad must stand proud of the rest of the active site surface. This is not easy to achieve with the serine side-chain, because it is so short. Placing the nucleophile on an extremely sharp turn, as occurs in the α/β hydrolase fold, achieves this effectively and allows easy access on one side by His and on the other by substrate (Ollis *et al.*, 1992). The sharp turn in the peptides also optimally positions the nucleophile at the end of helix αC so that the helix dipole can help stabilise both the tetrahedral intermediate in the catalytic process and the ionised form of the nucleophile.

In α/β hydrolase fold enzymes, the twist of the sheet imposes a 'handedness' on the catalytic triad, which is approximately a mirror-image of that seen in the serine proteases. It is known that the triad residues of chymotrypsin and papain have the opposite handedness (Garavito *et al.*, 1977), and therefore the handedness of papain and the enzymes of the α/β hydrolase fold is the same (Ollis *et al.*, 1992). Therefore, the stereochemistry of attack will also be the mirror-image of that seen in the serine protease and subtilisin families.

Finally, the candidate for the oxyanion hole occurs in a similar place in most enzymes in a turn between strand β3 and helix αA. The backbone amides of residues in this loop and of the amides of the residues immediately following the nucleophile (β5) point into a small cavity between the loop and the nucleophile. Consequently, this cavity, like the oxyanion hole in chymotrypsin, appears to be well designed to stabilise the tetrahedral intermediate. This is another example of the way in which the α/β hydrolase fold active site is a 'mirror-image' of that of the serine proteases. Thus, the role played by amide nitrogen of the active site serine in the serine proteases (Garavito *et al.*, 1977) is played by the amide nitrogen of the residue following the nucleophile (first residue in αC) in the α/β hydrolase fold enzymes (Ollis *et al.*, 1992).

1.8.2. Structural relationships of the α/β hydrolase fold proteins

The α/β hydrolase superfamily can be arranged into families, as illustrated in Figure 6, that are clearly related by sequence identity over the whole molecule (siblings), groups of families with only partial sequence identity (cousins), and extended groupings where the sequences only match around the nucleophile elbow, if at all (distant relatives).

Before the structures of AChE and GLP were determined, sequence comparisons suggested that the two enzymes had similar structures (Selvrag *et al.*, 1991; Shimada *et al.*, 1990; Slabas *et al.*, 1990), but comparing the sequence of AchE/GLP with DLH, HAL or CPW gave no indication of similarity. Furthermore, when structural alignments were used as a basis for sequence comparisons, there was no significant sequence similarity except around the nucleophile (Ollis *et al.*, 1992). Although it was impossible to predict the similarity of these enzymes based on their sequences, they were clearly members of a large class of similar proteins.

1.8.2.1. Siblings

The family of large esterases (LES) are siblings, with ~30% sequence identity across most, or all, of the molecule. The LES family includes acetylcholinesterases (eg. *Torpedo californica* ACE, TcACE [Rawes *et al.*, 1997]), fungal lipases (eg. *Geotrichum candidum* FLP2, GcFLP2 [Schrag and Cygler, 1993]) and the bovine bile-salt-activated cholesterol esterases (bBACE)[Chen *et al.*, 1998]. This family includes the largest α/β hydrolases and large insertions follow β1, β6 and β7. Except in bBACE, the triad acid is glutamate.

Even here precise distinctions have been made; some have grouped the fungal type B lipases (FLP2s) with the acetylcholinesterases (ACEs) because of the high sequence identity (eg. 30% between TcACE and GcFLP2), because they have very similar structures, and because almost all have the same unusual Ser-His-Glu catalytic triad (Heikinheimo *et al.*, 1999). However, others have grouped all the fungal lipases together separate from ACE, presumably by catalytic function and origin (Murzin *et al.*, 1995).

Nonetheless, despite the high sequence identity ACEs and FLP2s have significant differences in substrate preference and binding affinity: ACE has been optimised for the diffusion-controlled catalysis of a small substrate, whereas the same structures in FLP2 form the lid that achieves interfacial activation (Heikinheimo *et al.*, 1999).

1.8.2.2. Cousins

The haloperoxidase (HAL) family includes a wide variety of different enzymes, such as haloalkane dehalogenases (eg. *Xanthobacter autotrophicus* HAL, XaHAL [Verschueren *et al.*, 1993]), epoxide hydrolases from *Agrobacterium radiobacter* (ArEHY) and mouse (mEHY), bacterial bromo- and chloroperoxidases, proline iminopeptidase from *Xanthomonas campestris*, and thioesterase from *Vibrio harveyi*. *Bacillus subtilis* Brefeldin A esterase (BsBES; Wei *et al.*, 1999) is also included in this family, although it lacks αD.

Members of this family demonstrate poor conservation at the nucleophile elbow; however, the similarity between EHYs and HAL provide an example of cousins. Prior studies had found a low level of sequence similarity between HAL and an epoxide hydrolase (Jansen *et al.*, 1989). Epoxide hydrolases are generally involved in the detoxification of harmful aromatic compounds in mammals, and the structures of two EHYs have been solved: ArEHY (Nordini *et al.*, 1999) and mEHY. The sequence identity between ArEHY and XaHAL is 24%, which made it hard to construct an accurate model of either cousin from sequence alignment alone. As in HAL, there is an α-helical insertion between strands β6 and β7 in EHYs, which partially caps the active site cavity and contributes to the hydrophobic binding pocket. The catalytic residues are Asp333^{nuc}-His523-Asp495 (mEHY) and Asp107^{nuc}-His275-Asp246 (ArEHY), where the last residue of the triad is least important. Also, the N-terminal domain of mEHY (not present in ArEHY) has a six-stranded all-parallel α/β structure that is a vestigial haloacid dehalogenase (Heikinheimo *et al.*, 1999). Two different domains, both involved in haloalkane degradation, were thus fused into a single gene to form mEHY, with the haloacid dehalogenase becoming non-functional.

The dienelactone hydrolase (DLH) family includes porcine proline oligopeptidase (pPOP; Fulop *et al.*, 1998), *Pseudomonas* sp. B13 dienelactone hydrolase (P13DLH; Pathak and Ollis, 1990) and *Pseudomonas fluorescens* carboxylesterase (PfCES; Kim *et al.*, 1997). *Candida antarctica* lipase (CaFLP3; Uppenberg *et al.*, 1995) is also a member of this family although it lacks αE instead of αD. This family also provides another example of a cousin-type relationship of the α/β hydrolase fold, ie. the relationship between pPOP, which degrades short peptides, and P13DLH. The C-terminal domain of pPOP, like mEHY, is a canonical α/β hydrolase fold with structural similarity to P13DLH; the sequence identity from the structural alignment being 22% (Heikinheimo *et al.*, 1999). However, the N-terminal domain of pPOP is an unusual seven-bladed β-propellor domain, the function of which is to prevent all except small peptides from reaching the active site. P13DLH, on the other hand, uses two short peptides to bind its relatively small substrate.

1.8.2.3. Distant relatives

Some other lipases have structures similar to the α/β hydrolase fold, and thus are deemed to be distant relatives of the α/β hydrolase fold class of enzymes.

Rhizomucor miehei lipase (MLIP; Brady *et al.*, 1990) has a slightly different topology compared to other α/β hydrolase fold enzymes: ignoring strand one of its nine-stranded sheet, it is +1,+1, $(+1x)_4$, +1, while the α/β hydrolase fold is +1, +2, -1x, +2x, $(+1x)_3$. Although it is clear that the topologies are somewhat different, its catalytic triad is on the same strands as the α/β hydrolase folds: serine after strand five, aspartate after strand seven, and histidine after strand eight. The protein has a nucleophile elbow and the catalytic triad has the same hand as that of the α/β hydrolase fold enzymes (Brady *et al.*, 1990).

The topology of the first domain of human pancreatic lipase (hLIP) is even more like that of the α/β hydrolase fold enzymes: ignoring the first two strands, it is –1, +3, -1x, +2x, $(+1x)_3$, +1x, which differs from the α/β hydrolase fold only in the connection of the first two strands (Winkler *et al.*, 1990). The pancreatic lipase (PLP) family includes the human (Egloff *et al.*, 1995) and other mammalian PLPs; however, the catalytic triad residues are organised differently than in the canonical α/β hydrolase fold. The active site Ser and His are in the expected places, but the acid residue follows β6, not seven. Also, the PLP family members have a large C-terminal domain.

The serine carboxypeptidase (SCP) family includes serine carboxypeptidases from, for example, wheat (wSCP; Bullock *et al.*, 1996), the yeast kex1 protein and human protective protein. Two additional strands are present between αE and β8. They are also examples of distant relatives as members of this family, such as carboxypeptidase A (CBPA; Rees *et al.*, 1983), show some similarity to enzymes of the α/β hydrolase fold. The topology of the mixed sheet in CBPA is +1, +2, +1x, +2x, +2, +1x, -2, so the topology of the first five strands is the same as the α/β hydrolase fold enzymes, but the strands in its sheet all tend to be much shorter. Nevertheless, certain features of the active site of CBPA are similar to the α/β hydrolase fold. The metal ion is bound by residues located on loops at the C-termini of strands three and five; equivalent loops in the α/β hydrolase fold form the oxyanion pocket and hold the nucleophile.

The structure of BsBES (Wei *et al.*, 1999), which hydrolyses a macrocyclic fungal toxin, is distantly related by sequence to the important human hormone-sensitive lipase (hHSL), which performs the rate-limiting step of hydrolysing fat stored in adipocytes. The fold of BsBES (and by inference the unsolved structure of hHSL) is closest to that of the HAL family. The catalytic triad of BsBES is Ser202-His338-Asp308, and so led to the identity of the catalytic triad of hHSL as Ser423-His733-Asp703 (Heikinheimo *et al.*, 1999).

Other enzymes have even greater differences in topology to the canonical α/β hydrolase fold, and classifying these has thus been harder still. One such enzyme is FsCUT, the structure of which has been determined (Longhi *et al.*, 1997), whose function is closely related to that of lipases and so should have evolved from the large esterase (LES) family. FsCUT, however, lacks β1, β2, β8 and αE of the canonical fold. Also, the catalytic His is located in a loop-like structure that replaces the omitted αE and β8 and that is followed by a bent and distorted αF. FsCUT does, though, have some α/β hydrolase fold features: the nucleophile elbow on strand β5, the inverted geometry of attack as compared to the serine proteases, and the correct core topology. FsCUT had previously been grouped with flavodoxin (Murzin *et al.*, 1995) which also has a three-layered α/β/α fold with strand order 21345. However, as there is no mechanistic relationship between FsCUT and flavodoxin, the cutinases have been grouped together as the CUT family, which contains the smallest members of the α/β hydrolase fold (Heikinheimo *et al.*, 1999).

Sequence comparisons can be used to identify additional hydrolase domain proteins whose structures have not yet been determined. A search of sequence databases revealed sequence similarity between HAL and two C-C hydrolases, 2-hydroxymuconic semialdehyde hydrolase (XylF; Nordland and Shingler, 1990) and 2-hydroxy-6-oxo-6-phenyl-hexa-2,4-dienoate hydrolase (BphD; Kimbara *et al.*, 1989), that are involved in catabolic degradation pathways of aromatic compounds (cf. chapter 1.5.). These three enzymes are most similar over the first half of the α/β hydrolase fold (strands β2 to β6). Thus, though these two proteins may have had α/β hydrolase folds after strand seven, it was not proven until their structures were solved and they were shown to be proteins of the α/β hydrolase fold superfamily (Horn *et al.*, 1991; Mitsui *et al.*, 1997).

1.8.2.4. Summary

There seem to be four requirements to distinguish α/β hydrolase fold proteins from others:

(i) the sequence order of the catalytic triad is nucleophile-acid-histidine, with the nucleophile on canonical strand β5;

(ii) there is a nucleophile elbow at the top of canonical strand β5, with a sequence pattern that is often G-X-Nuc-X-G;

(iii) the structure starts at strand β3 and is at least five strands long including the cross-over connection at the nucleophile (strands 43567; figure 6b); and,

(iv) a long loop at the end of strand β7 allows the side-chains of the triad residues to hydrogen bond.

Of the last three features, at least two are present in all α/β hydrolase fold enzymes (Heikinheimo *et al.*, 1999).

A distinguishing feature of the α/β hydrolase fold is its ability to be modified while retaining catalytic function. Attached to this hydrolytic domain are different substrate-binding loops with sizes that vary from simple turns to 80-residue insertions. However, these loops are integrated into the single-domain structure, rather than added on as complete domains.

Among the five classes of α/β hydrolases that include lipases (Figure 6), many different kinds of changes are seen: the fold of bacterial lipases starts at strand β3; *Candida Antarctica* lipase (CaFLP3) starts at strand β2; LES family lipases have the complete canonical fold; and the PLPs contain eight β-strands with an additional strand before the separate C-terminal domain. This is consistent with the idea that the minimal fold (and possibly the evolutionary ancestor) started from strand β3 (Heikinheimo *et al.*, 1999), but not with the idea that the α/β hydrolase fold is a two-domain protein (Cygler and Doctor, 1993).

Unlike most superfamilies, where the size of the fold varies over a rather narrow range, the fold of this family ranges widely (from 197 to 583 residues).

It seems clear that the α/β hydrolase fold enzymes have diverged from a common ancestor and that they have evolved to preserve the positions of key catalytic components. They share a similarity in their central catalytic domains, have similar core topology, a conserved sequence order for catalytic triad residues, and conserved loops for the catalytic triad and oxyanion hole. Their three-dimensional structure similarity is also remarkable given the lack of sequence similarity between many of the proteins.

There is no global similarity between the α/β hydrolase fold enzymes and the other enzymes that have catalytic triads: the serine proteases, subtilisins and the cysteine proteases. Thus, there are now four different examples of the catalytic triad which are related by convergent, not divergent, evolution possibly reflecting how central hydrolysis is to biochemical pathways, and how few solutions are possible, at the level of individual amino acid side-chain chemistry, to the problem of hydrolysing esters and amides (Ollis *et al.*, 1992).

1.9. Serine proteases

Proteolytic enzymes dependent on a serine residue for catalytic activity are widespread and very numerous. Serine proteases are found in viruses, bacteria, and eukaryotes, and they include exopeptidases, endopeptidases, oligopeptidases, and omega peptidases (Table 2).

Table 2
A Selection of Serine Proteases

Enzyme	Source	Function
Trypsin	Pancreas	Digestion of proteins
Chymotrypsin	Pancreas	Digestion of proteins
Elastase	Pancreas	Digestion of proteins
Thrombin	Vertebrate serum	Blood clotting
Plasmin	Vertebrate serum	Dissolution of blood clots
Kallikrein	Blood and tissues	Control of blood flow
Complement C1	Serum	Cell lysis in the immune response
Acrosomal protease	Sperm acrosome	Penetration of ovum
Lysosomal protease	Animal cells	Cell protein turnover
Cocoonase	Moth larvae	Dissolution of cocoon after metamorphosis
α-lytic protease	*Bacillus sorangium*	Possibly digestion
Proteases A and B	*Streptomyces griseus*	Possibly digestion
Subtilisin	*Bacillus subtilis*	Possibly digestion

Source: Stroud (1974).

The serine proteases are the most thoroughly understood family of enzymes as a result of their extensive examination over a 40-year period by kinetic, chemical, and physical techniques (Voet and Voet, 1995).

Interest in this family has been maintained in part by an increasing recognition of their involvement in a host of physiological processes. In addition to the biological role played by digestive enzymes such as trypsin, serine proteases also function broadly as regulators through the proteolytic activation of precursor proteins (Neurath, 1984; Van de Ven *et al.*, 1993). An example of this regulation being the processing of trypsinogen by enteropeptidase to produce active trypsin (Huker and Bode, 1978). Another reason for the continued interest in serine proteases has been their emergence as one of the major paradigms for the understanding of enzymic rate enhancements and of structure-activity relationships (Perona and Craik, 1995).

1.9.1. Families of serine proteases

Over 20 families of serine proteases have been recognised, and most of these can be grouped together into about six clans (Table 3), based on three-dimensional studies and some less direct evidence (Rawlings and Barrett, 1994).

Table 3

Clans and Families of Serine Proteases[a]

Clan	Family	Representative enzyme	Known catalytic residues
SA	S1	Chymotrypsin	------H-----D----------------------S---------
“	S2	α-Lytic endopeptidase	--------H---D--------------------S-----------
“	S3	Sindbis virus core endopeptidase	------H---D------------------S--------------
“	S5	Lysyl endopeptidase	-------H-------D---------------S------------
“	S6	IgA-specific serine endopeptidase	--------S--------------------------------------
“	S30	Tobacco etch virus 35kDa endopeptidase	---------------------------H-D---S----------
“	S7	Yellow fever virus NS3 endopeptidase	---HD--S-------------------------------------
“	S29	Hepatitis C virus NS3 endopeptidase	-----------S-----------------------------------
“	S31	Cattle diarrhea virus p80 endopeptidase	------------S----------------------------------
“	S32	Equine arteritis virus putative endopeptidase	--
SB	S8	Subtilisin	----D----H---------------------------S------
SC	S9	Prolyl oligopeptidase	--------------------------------------S---DH
“	S10	Carboxypeptidase C	---------------S-----------------------D----H
“	S28	Lysosomal Pro-X carboxypeptidase	--
“	S15	*Lactococcus* X-Pro-peptidase	------------------S--------------------------

[a] The arrangement of catalytic residues in the representative enzymes is illustrated diagrammatically by use of lines of normalised length to depict the polypeptide chains, although in reality the chains vary considerably in length.

In the early 1990's it was thought that all of the known serine protease enzymes fell into one of two distinct structural classes: the chymotrypsin-like and subtilisin-like clans (Matthews, 1977). However, the crystal structure of wheat serine carboxypeptidase II (Liao and Remington, 1990; Liao *et al.*, 1992) revealed the conservation of the essential features of the catalytic apparatus within a third distinct protein fold. This homodimeric protein possesses the α/β hydrolase fold found also in a number of other enzymes that share hydrolytic activity as their only common feature (cf. chapter 1.8.; Ollis *et al.*, 1992). The structure for *Escherichia* D-Ala-D-Ala peptidase has shown that this clan is also unrelated to the other three with known structures.

The three serine protease clans, SA, SB, and SC, are distinguished by the absence of any conserved secondary and tertiary motifs, but in each case, the catalytic serine, histidine, and aspartate residues maintain an identical geometric orientation (Figure 8). To a lesser extent, adjacent groups that stabilise the transition state are also similarly arranged (Wright *et al.*, 1969; Robertus *et al.*, 1972; Liao *et al.*, 1992). Thus, it appears that nature has arrived at the same biochemical mechanism by separate avenues: the chymotrypsin, subtilisin, and serine carboxypeptidase clans of serine proteases are a classic example of convergent enzyme evolution (Matthews *et al.*, 1977; Liao *et al.*, 1992). The resemblance of serine carboxypeptidase to other members of the α/β hydrolase fold family also indicates the operation of divergent evolution within this structural framework (Ollis *et al.*, 1992).

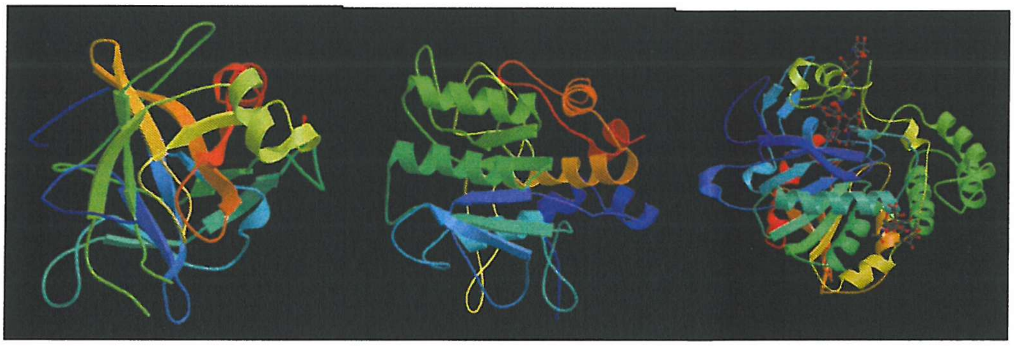


Figure 8: Diversity of structural motifs in which the common catalytic apparatus of serine protease is embedded. Shown are ribbon diagrams of chymotrypsin, subtilisin BPN', and wheat serine carboxypeptidase (left to right).

However, the catalytic mechanisms of clans SE (*Escherichia* D-Ala-D-Ala peptidase A) and SF (repressor LexA) are already known to be very different, and there is little doubt that some of the other families of serine proteases also will prove to have distinctive mechanisms of action, without the 'classic' Ser-His-Asp triad (Rawlings and Barrett, 1994).

The arrangements of the catalytic residues in the linear sequences of members of the various families commonly reflect their relationships at the clan level (Table 2).

The presence of several glycine residues in the vicinity of the catalytic serine residue, to form the motif Gly-X-Ser-X-Gly, led to the expectation that this would be found in all serine proteases (Brenner, 1988). However, although most families do show conserved glycine residues near the essential serine, this is not generally the case for all serine proteases (Table 4), and the exact positions are variable (Rawlings and Barrett, 1994).

Clan	Family	
SA	S1	- - G - S G - - - -
	S2	- - G - S G - - - -
	S3	- - G - S G - - - -
	S5	- - G - S G - - - -
	S6	- - G - S G - - - -
	S30	- - G - S G - - - -
	S7	- - G - S G - - - -
	S29	- - G - S GG - - -
	S31	- - G - S G - - - -
SB	S8	- - G - S - - - - -
SC	S9	- - G - S - GG - -
	S10	- - G - S - - G - -
	S15	- - G - S - - G - -

Table 4: Glycine residues totally conserved in the vicinity of the catalytic serine residues in members of various families of serine proteases. The codes for clans and families are as in Table 3.

1.9.1.1. Chymotrypsin clan (SA)

In the chymotrypsin clan of serine proteases, comprising some ten families of endopeptidases, the three-dimensional structures of chymotrypsin (S1), α-lytic endopeptidase (S2), Sindbis virus core endopeptidase (S3), and *Achromobacter* lysyl endopeptidase (S5), are known to be similar. The order of catalytic residues in the polypeptide chain in clan SA is His-Asp-Ser.

The members of the chymotrypsin family (EC 3.4.21) are almost entirely confined to animals and are inherently secreted proteins, either directly (eg. proenzymes of coagulation factors in the liver) or packaged in vesicles for regulated secretion (eg. chymotrypsin in the pancreas). They may also be retained in leucocyte granules (eg. elastase, chymase, or granzymes in polymorphonuclear leucocytes, mast cells, or cytotoxic lymphocytes, respectively).

There are about 200 complete amino acid sequences known, more than for any other known family of serine protease. Most are monomers, but granzyme A is a disulfide-bonded homodimer and tryptase a homotetramer (Rawlings and Barrett, 1994).

The essential catalytic unit of these proteases is a polypeptide chain of about 220 amino acid residues. However, many members of the family have N-terminal extensions, with the catalytic unit almost invariably forming the C-terminal domain (Rawlings and Barrett, 1994).

Proteolytic cleavage at the N-terminus of the catalytic domain of a proenzyme in the chymotrypsin family forms a new N-terminal amino acid residue with a hydrophobic side-chain. The new terminal α-amino group forms a salt bridge with Asp194, which leads to the assembly of the functional catalytic site (Huber and Bode, 1978). The N-terminal residue is commonly Ile, but may be Leu, Val or Met, and the salivary plasminogen activator from the vampire bat contains Ser in this position (Kratzschmar, 1991).

The fact that members of family S1 occur in both eukaryotes and prokaryotes has been the subject of considerable debate. The *Actinomycete* trypsins have sequences too similar to eukaryotic trypsins to be consistent with divergence 3500 million years ago, which is when the common ancestor of prokaryotes and eukaryotes is generally thought to have lived. Instead, it is thought that 1300 million years ago, at the time that mitochondria were introduced into eukaryotes, a horizontal gene transfer occurred from the protomitochondrion to the eukaryote host (Young *et al.*, 1978).

1.9.1.2. Subtilisin clan (SB)

The subtilisin family (S8) is the second largest family of serine proteases so far identified; and is extremely widespread, members having been found in eubacteria, archaebacteria, eukaryotes, and viruses. The majority of these enzymes are endopeptidases (Rawlings and Barrett, 1994).

Crystallographic structures have been determined for several members of the family, and these show a catalytic triad composed of the same residues as in proteases of the chymotrypsin clan. However, these occur in a different order in the sequence (Asp-His-Ser), and the three-dimensional structures of these molecules bear no resemblance to that of chymotrypsin, so that it is clear that subtilisin and chymotrypsin are not evolutionarily related.

A distinct subfamily, termed the kexin subfamily, comprises proprotein-processing endopeptidases kexin, furin, and related enzymes, from yeast and animals, that preferentially cleave C-terminally to paired basic amino acids. The members of this subfamily can be identified by the subtly different motifs around the active site residues. The kexin subfamily have Asp in place of Ser or Thr in position 139, Arg in place of His in position 172, and Ala in place of Met in position 324 (Brenner *et al.*, 1994; Nakayama, 1994; Seidah & Chretien, 1994).

In subtilisins, the oxyanion hole is formed by the active site Ser and Asn262 (Brenner *et al.*, 1994).

Some members of the subtilisin family have a requirement for thiol activation, such as endopeptidases R, T, and K from the yeast *Tritirachium* as well as members of the kexin subfamily. This thiol dependence is attributable to Cys173 near the active-site histidine (Jany *et al.*, 1986). However, all but one of the bacterial subtilisins has no Cys residues whatsoever.

The sole viral member of family S8 is a 56-kDa protease from the herpes virus 1 that infects the channel catfish (Davison, 1992). Herpes viruses are related to baculoviruses, both of which groups encode proteins homologous to mammalian endopeptidases.

The subtilisin family contains some mosaic proteins. The secreted endopeptidase from *Vibrio alginolyticus* (Deane *et al.*, 1989) and *Xanthomonas campestris* contains homologous C-terminal domains similar to some found in metalloproteases (Rawlings and Barrett, 1994).

Family S8 contains several enzymes with N- or C-terminal extensions that show no relationship to other proteins. The kexin subfamily includes endopeptidases with a variety of C-terminal extensions (Seidah & Chretien, 1994).

1.9.1.3. Carboxypeptidase clan (SC)

Tertiary structures available for enzymes of the carboxypeptidase C family (S10) are of the α/β hydrolase fold type (Ollis *et al.*, 1992), also seen for a wide variety of other enzymes including acetylcholinesterase, lipases and *Xanthomonas* haloalkane dehalogenase. The sequence of *Neisseria* prolyl aminopeptidase (family S33) shows it to be homologous to the haloalkane dehalogenase, so it has been assumed that families S10 and S33 have similar tertiary folds and linear arrangements of catalytic residues (Ser-Asp-His) and form a clan (Rawlings and Barrett, 1994).

The distinctive Ser-Asp-His arrangement of catalytic residues is also found in family S9 (prolyl oligopeptidase), which raises the possibility that family S9 is also a member of clan SC (Polgar, 1992).

The proteases of family S9 differ biologically from those of families S10 in that the homologues of prolyl oligopeptidase are either cytosolic or integral membrane proteins, whereas the carboxypeptidases are secreted or lysosomal enzymes.

The active site Ser, Asp, and His residues have been identified in family S9 (David *et al.*, 1993), and in all known members of the family these residues are within about 130 residues of the C-terminus. The N-terminal parts of these molecules are more or less variable, with the membrane-bound members of this family containing membrane-spanning domains near the N-terminus.

The carboxypeptidase C family (EC 3.4.16) includes carboxypeptidases that are unusual among serine-dependent enzymes in that they are maximally active at acidic pH. The family appears to be restricted to eukaryotes, with enzymes known from protozoa, fungi, plants, and animals (Rawlings and Barrett, 1994). The fungal and plant enzymes can be divided into variants of carboxypeptidase C (EC 3.4.16.5), with a preference for a hydrophobic amino acid, and those of carboxypeptidase D (EC 3.4.16.6), which releases a C-terminal arginine or lysine (Remington & Breddam, 1994).

In all the proteases, the residue preceding the catalytic Ser is a Glu residue that is believed to be responsible for the acidic pH optimum of the enzymes (Remington & Breddam, 1994). The differences in substrate specificity between carboxypeptidase C and D have been attributed to two residues that are Glu in carboxypeptidase D but are both hydrophobic in character (Phe, Leu, or Met) in carboxypeptidase C (Remington & Breddam, 1994).

The crystal structures of several members of family S10 have been determined (Remington & Breddam, 1994), with the fold of these proteins bearing no relation to those of chymotrypsin, subtilisin, or D-Ala-D-Ala proteases. However, the tertiary structures of serine carboxypeptidases do strongly suggest a distant relationship to a variety of non-protease enzymes.

The non-protease enzymes that show no amino acid sequence similarity to carboxypeptidase C, but are thought to be distant relatives because of their similar tertiary structures, include cholinesterases, haloalkane dehalogenases, and dienelactone hydrolases in the α/β hydrolase fold superfamily (cf. chapter 1.8; Ollis *et al.*, 1992; Cygler *et al.*, 1993).

Prolyl aminopeptidase (family S33) is a 35kDa protease from *Nisseria gonorrhoea* that selectively hydrolyses N-terminal Pro residues (Albertson and Koomey, 1993). The amino acid sequence shows no relationship to that known for any other protease, but is homologous to those of *Pseudomonas* 2-hydroxymuconic semialdehyde hydrolase (XylF), *Xanthobacter* haloalkane dehalogenase (XaHAL), and human and rat epoxide hydrolases (EHYs) (Rawlings and Barrett, 1994). These enzymes, in turn, are known to be structurally related to acetylcholinesterase and carboxypeptidase C (family S8) (Ollis *et al.*, 1992). Most of the α/β hydrolase fold enzymes have Ser at the active site, and prolyl aminopeptidase also has Ser at this position.

1.9.2. Catalytic mechanism of the serine proteases

The vast majority of early studies on the serine proteases focused on the elucidation of the chemical and kinetic mechanisms of catalysis.

1.9.2.1. Ester hydrolysis as a kinetic model

The chemical mechanisms of ester and amide hydrolysis are almost identical; thus, it was not surprising when it was found that chymotrypsin can act as an esterase as well as a protease. In fact, the study of chymotrypsin's esterase activity has led to important insights concerning this enzymes catalytic mechanism.

Kinetic measurements of the chymotrypsin-catalysed hydrolysis of *p*-nitrophenylacetate (Hartley *et al.*, 1954) indicated that the reaction occurs in two phases:

(i) the 'burst phase', in which the highly coloured *p*-nitrophenylate anion is rapidly formed in amounts stoichiometric with the quantity of enzyme present;

(ii) the 'steady state phase', in which *p*-nitrophenylate is generated at a reduced but constant rate that is independent of substrate concentration.

These observations have been interpreted in terms of a two-stage reaction sequence in which the enzyme (1) rapidly reacts with the *p*-nitrophenylacetate to release the *p*-nitrophenylate anion forming a covalent acyl-enzyme intermediate that (2) is slowly hydrolysed to release acetate (Figure 9).

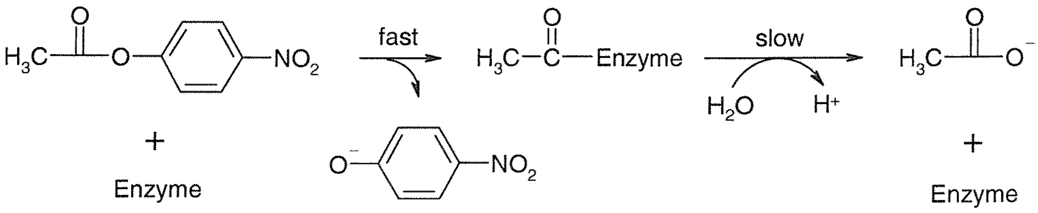


Figure 9: Chymotrypsin-catalysed hydrolysis of *p*-nitrophenylacetate

Chymotrypsin-catalysed amide hydrolysis has been shown to follow a reaction pathway similar to that of ester hydrolysis but with the first step of the reaction, enzyme acylation, being rate determining rather than the deacylation step (Voet and Voet, 1995).

1.9.2.2. Identification of the catalytic residues

Chymotrypsins catalytically important groups have been identified by chemical labelling studies.

The presence of an active site serine residue in serine proteases was discovered by their reaction with diisopropylphosphofluoridate (DIPF), which irreversibly inactivates the enzyme (Figure 10). Other Ser residues, including those on the same protein, do not react with DIPF. DIPF reacts only with Ser195 of chymotrypsin, thereby demonstrating that this residue is catalytically important.

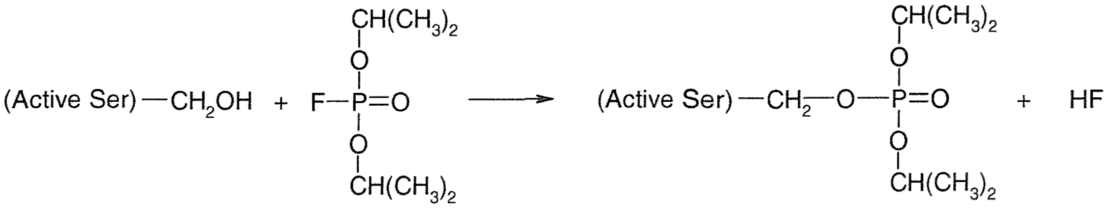


Figure 10: Reaction of chymotrypsin with diisopropylphosphofluoridate (DIPF)

The use of DIPF as an enzyme inactivating agent came about through the discovery that organophosphorus compounds are potent nerve toxins. The neurotoxicity of DIPF arises from its ability to inactivate acetylcholinesterase, a serine esterase that catalyses the hydrolysis of acetylcholine, a neurotransmitter.

A second catalytically important residue was discovered through affinity labelling, when a substrate analogue bearing a reactive group specifically binds at the enzymes active site, where it reacts to form a stable covalent bond with a nearby susceptible group. In the case of chymotrypsin, a phenylalanine analogue, tosyl-L-phenylalaninechloromethylketone (TPCK), specifically binds, and the chlorophenylketone group, a strong alkylating agent, reacts only with His57, thereby inactivating the enzyme (Figure 11).

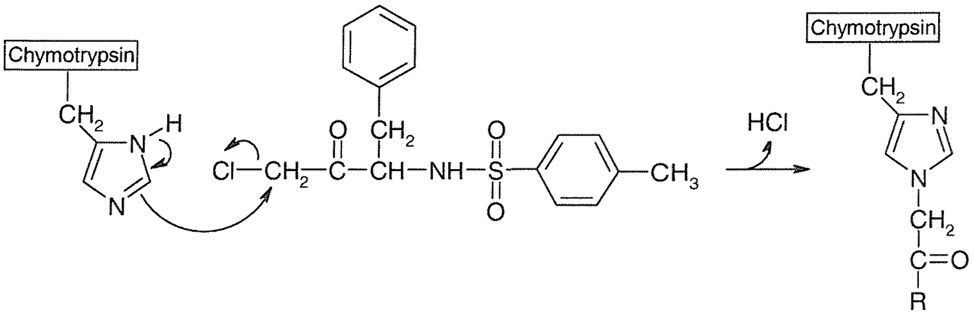


Figure 11: The reaction of TPCK with His57 of chymotrypsin

The TPCK reaction is inhibited by β-phenylpropionate, a competitive inhibitor of chymotrypsin that presumably competes with TPCK for its enzymatic binding site. Moreover, the TPCK reaction does not occur in 8M urea, a denaturing agent, or with DIP-chymotrypsin, in which the active site is blocked. These observations established that His57 is also an essential active site residue of chymotrypsin.

1.9.2.3. The catalytic mechanism

Hydrolysis of ester and amide bonds proceeds by an identical acyl transfer mechanism in all enzymes of the subtilisin and chymotrypsin families (Figure 12). Michaelis complex formation is followed by attack on the carbonyl carbon atom of the scissile bond by the eponymous serine of the catalytic triad, which is enhanced in nucleophilicity by the presence of an adjacent histidine functioning as a general base catalyst. Proton donation by the histidine to the newly formed alcohol or amine group then results in dissociation of the first product and concomitant formation of a covalent acyl-enzyme complex. The deacylation reaction occurs via the same mechanistic steps, with the attacking nucleophile provided by a water molecule that approaches from the just-vacated leaving group side. Each step proceeds through a tetrahedral intermediate, which resembles the high-energy transition state for both reactions.

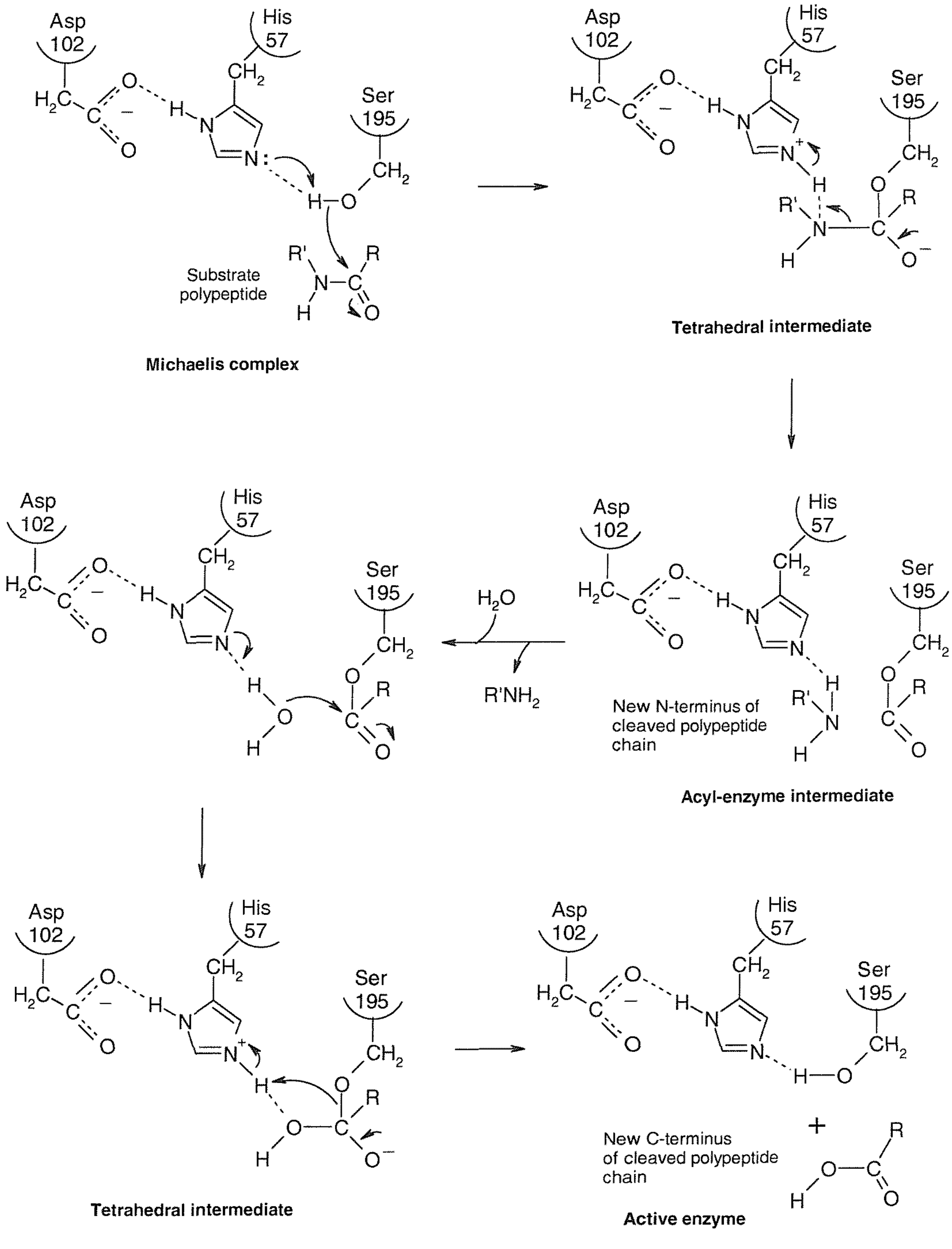


Figure 12: Mechanism of serine proteases

This mechanism is capable of accelerating the rate of peptide bond hydrolysis by a factor of more than 10^9 relative to the uncatalysed reaction (Kahne and Still, 1988).

A third member of the catalytic machinery is the conserved aspartate residue, which aids the formation of the imidazolium ion by the polarising effect of its unsolvated carboxylate ion which is hydrogen bonded to the histidine.

Extensive structural evidence obtained from X-ray crystallographic and NMR investigations has provided conclusive corroboration of the essential features of this mechanism (Steitz and Shulman, 1982). The investigations have been favoured by the availability of good ground-state and transition-state substrate analogues, which have been used to obtain high-resolution images of these interactions. The scissile bond of the substrate is bound directly adjacent to the Ser-His catalytic couple in all complexes studied. A strong hydrogen bond between these two amino acids, necessary to subsequent proton transfer, is formed only after substrate is bound (Perona and Craik, 1994). A binding site for the oxyanion of the intermediate is formed by the Gly193 and Ser195 backbone amide nitrogens in the chymotrypsin-like enzymes, by one amide nitrogen and the Asn155 side chain in the subtilisin family, and by the backbone amides of Tyr147 and Gly53 in the serine carboxypeptidases (Liao *et al.*, 1992).

Mutational analysis of both subtilisin and trypsin has confirmed the essential roles of Ser195 and His57 in providing rate acceleration (Carter and Wells, 1987, 1988; Corey and Craik, 1992). Studies have also shown that enzyme variants providing an alternative general base were ineffective, further underscoring the importance of the native catalytic triad geometry (Corey and Craik, 1992). The mutational data are in agreement with early chemical modification experiments, which also indicated that Ser195 and His 57 play crucial roles in catalysis (cf. chapter 1.9.2.2.; Dixon *et al.*, 1956; Shaw *et al.*, 1965).

Considerable controversy originally surrounded the role of the conserved buried aspartate residue, first described in the crystal structure of chymotrypsin (Blow *et al.*, 1969). Mutation of this residue confirmed its essential role, because all variants of trypsin and subtilisin in which the aspartate is absent are decreased in catalytic efficiency by at least a factor of 10^4 (Corey and Craik, 1992).

The early suggestion that the Asp-polarised His57 side-chain directly abstracts a proton from the Ser195, thereby converting it's weakly nucleophilic $-CH_2OH$ group to a highly nucleophilic alkoxide ion $-CH_2O^-$, is largely unsupported by the bulk of the experimental (Bachovchin and Roberts, 1978; Markley, 1979; Kossiakoff and Spencer, 1981) as well as theoretical (Warshel *et al.*, 1989) evidence. In this two proton transfer model, the anionic charge of Asp102 was thought to be transferred, via a tautomeric shift of His57, to Ser195. The catalytic triad was therefore originally named the 'charge relay system' (Figure 13).

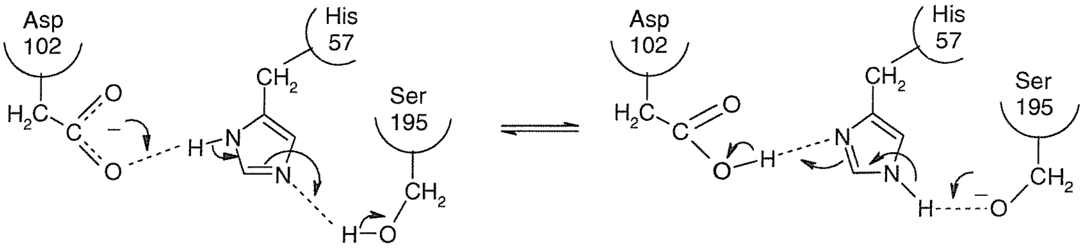


Figure 13: The "charge relay system".

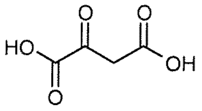
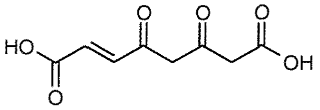
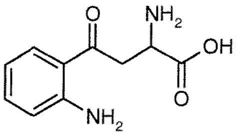
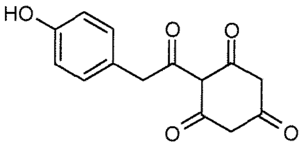
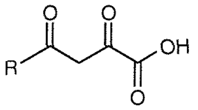
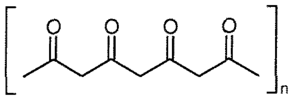
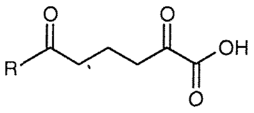
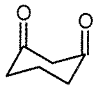
Instead, one role of the conserved Asp appears to be stabilisation of the tautomer and rotamer of the catalytic His (Craik *et al.*, 1987). In addition, because the His imidazole ring acquires a proton in the transition state, the Asp carboxylate can provide compensation for the developing positive charge. This role is similar to that of the hydrogen bond donor groups in the oxyanion hole, which compensate the developing negative charge on the substrate carboxyl oxygen atom (Warshel *et al.*, 1989).

Comparative analysis of the structures of chymotrypsin, subtilisin, and serine carboxypeptidase shows that the precise geometric orientation of the Asp is not conserved relative to the Ser-His catalytic dyad (Liao *et al.*, 1992). It appears that the Asp can occupy virtually any position relative to the Ser-His dyad (Perona and Craik, 1994). Therefore, it may be more accurate to regard the operation of the serine protease catalytic machinery as two dyads – Ser/His and His/Asp – that operate in concert, rather than as a single catalytic triad (Liao *et al.*, 1992).

1.10. Carbon-carbon bond hydrolases

Hydrolases catalyse the hydrolysis of various bonds, such as carbon-carbon (C-C), carbon-oxygen (ester) and carbon-nitrogen (amide) bonds. Some have very wide specificity, while others will only hydrolyse one specific substrate. There are very few hydrolases known that act solely on C-C bonds; and these mostly catalyse the hydrolysis of 3-oxo-carboxylic acids (or β-ketoacids; hence, they are also known as β-ketolases). There are a number of enzymes reported to fall into this category, but at present only ten have been classified by the International Union of Biochemistry and Molecular Biology (IUBMB), as shown in Box 1 (Webb, 1992).

Box 1

Substrate	C-C bond hydrolase	EC number
	Oxaloacetase	3.7.1.1
	Fumarylacetoacetase	3.7.1.2
	Kynureninase	3.7.1.3
	Phloretin hydrolase	3.7.1.4
	Acetopyruvate hydrolase (Acylpyruvate hydrolase)	3.7.1.5(6)
	β-diketone hydrolase	3.7.1.7
	2,6-dioxo-6-phenylhexa-3-enoate hydrolase (2-hydroxymuconate-semialdehyde hydrolase)	3.7.1.8(9)
	Cyclohexane-1,3-dione hydrolase	3.7.1.10

These enzymes hydrolyse C-C bonds in pathways that metabolise the aromatic compounds biphenyl, phloretin, phenylalanine, tyrosine, tryptophan and toluene. Carbon-carbon bond hydrolases also act on oxaloacetate and related acylpyruvates in microbial metabolism (Lenz *et al.*, 1976).

β-ketolases hydrolyse carbon-carbon bonds of 1,3-dioxo compounds. These enzymes are frequently encountered in intermediary metabolism and also hydrolyse 1,5-dioxovinylogous analogues (Pokorny *et al.*, 1997). They are very common enzymes in microbial metabolism and in the liver, where they were first described in the hydrolysis of 2,4-dioxoalkanoates and acetoacetone (Meister and Greenstein, 1948).

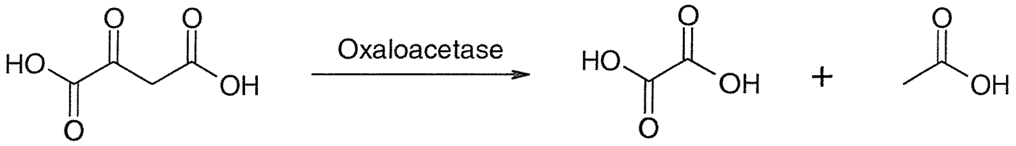
Natural products derived from the various polyketide and chorismate enzymic routes of biosynthesis may represent the most diverse group of natural compounds in terms of their chemical structure and biological function (Pokorny *et al.*, 1997). Catabolic reaction sequences, found mostly in microorganisms, use enzymes to reverse these biosynthetic processes by hydrolysis as an essential part of the carbon cycle (cf. chapter 1.2).

The provision of 1,3-dioxo compounds (and their vinylogous 1,5-dioxo and 3-amino-1-oxo analogues) by enzymes (Box 1) is an essential feature for easy C-C bond cleavage, and is readily achieved by decarboxylases, aldolases and some transferases (Abeles *et al.*, 1992; Stryer, 1994).

β-ketolases are involved in diverse reaction pathways; for instance, they can cleave branched-chain β-diketones, as in the thymol catabolic pathway of *Pseudomonas putida* (Chamberlain and Dagley, 1968). However, when γ-diketones (1,4-diketones) are formed, as in camphor metabolism by *Pseudomonas* and *Mycobacteria*, then Baeyer-Villiger lactone forming enzymes are responsible for the C-C bond cleavage. Thus, metabolic pathways that yield 1,4-diketones do not include C-C bond hydrolysis reactions (Pokorny *et al.*, 1997).

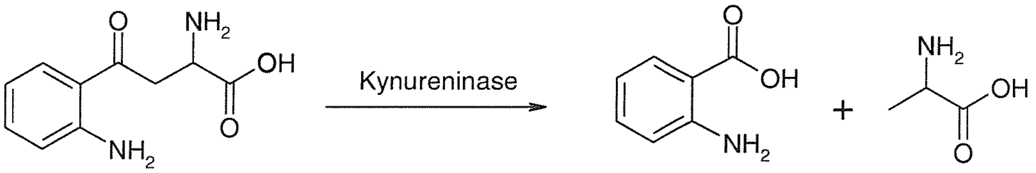
1.10.1. Reactions of carbon-carbon bond hydrolases

Oxaloacetase (EC 3.7.1.1) from *Aspergillus niger* catalyses the hydrolysis of oxaloacetate to oxalic acid and acetic acid (Hayaishi *et al.*,1956):

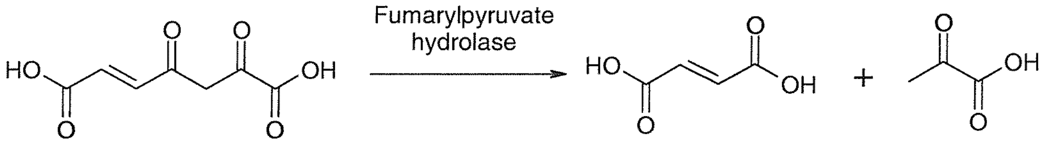


Chymotrypsin (EC 3.4.21.1), although a serine protease (cf. chapter 1.9), will also catalyse the hydrolysis of a carbon-carbon bond in ethyl 5-(*p*-hydroxyphenyl)-3-ketovalerate to give 3-(*p*-hydroxyphenyl)propionate and ethyl acetate (Doherty, 1955). In an analogous manner, trypsin (EC 3.4.21.4) cleaves ethyl 5-(*p*-aminophenyl)-3-ketovalerate (Roget and Calvet, 1962), illustrating that reactions catalysed by C-C bond hydrolases are in some respects similar to the hydrolysis of an ester (Hsiang *et al.*, 1972).

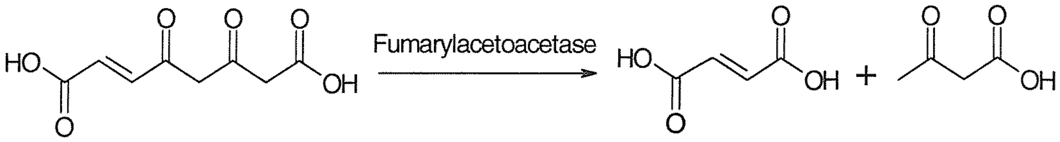
L-kynurenine hydrolase (EC 3.7.1.3) catalyses the formation of anthranilate and alanine from L-kynurenine (Longenecka and Snell, 1955):



In a bacterial system, the pathway for the degradation of gentisic acid includes the enzyme fumarylpyruvate hydrolase, an example of an acylpyruvate hydrolase (EC 3.7.1.6), which hydrolyses fumarylpyruvic acid to fumarate and pyruvate (Lack, 1961):



The metabolic pathway for the degradation of tyrosine and phenylalanine, in mammalian systems, utilises the enzyme fumarylacetoacetate fumarylhydrolase (EC 3.7.1.2) to hydrolyse fumarylacetoacetate to fumaric acid and acetoacetic acid (Ravdin and Crandall, 1951):



This 2,4-diketo acid hydrolase has been shown to be identical to a β-ketolase isolated from rat liver (Meister and Greenstein, 1948) and purified to homogeneity (Brock and Williamson, 1968), as well as a protein partially purified from beef liver that hydrolyses triacetic acid to acetoacetic acid and acetic acid (Connors and Stotz, 1949; Hsiang *et al.*, 1972).

Loss of the fumarylacetoacetate hydrolase (FAH) activity required for phenylalanine and tyrosine degradation is lethal to humans (Lindblad *et al.*, 1977; Tanguay *et al.*, 1996). Several compounds that accumulate upstream of FAH are toxic alkylating agents (Jorquera and Tanguay, 1997) that can induce apoptosis (Endo and Matsuda, 1997). Thus, FAH deficiency causes the fatal metabolic disease hereditary tyrosinemia type I (HT1), an autosomal recessive disorder that leads to liver failure, liver cirrhosis, liver cancer, renal tubular damage and neurological crises (Lindblad *et al.*, 1977; Tanguay *et al.*, 1996).

The reaction catalysed by the diketo acid hydrolase (FAH) is similar to the hydrolysis of an ester, an idea reinforced by the ability of chymotrypsin to catalyse the hydrolysis of a C-C bond (Doherty, 1955). As many esterases contain either a catalytically essential thiol group (cysteine proteases) or serine residue (serine proteases) in their active sites, a study of sulfhydryl-specific and serine-specific inhibitors on FAH was conducted (Hsiang *et al.*, 1972). Thiol-specific inhibitors, such as hydroxymercuribenzoate, were shown to inhibit the enzyme, with the exception of iodoacetate; whereas, serine-specific inhibitors, such as DIPF and phenylmethylsulfonyl fluoride (PMSF), had no inhibitory effect whatsoever. Thus, a serine esterase type activity for FAH seemed unlikely, and the presence of an active-site cysteine more likely (Nagainis *et al.*, 1981). However, the crystal structure of FAH has recently been determined using the multiwavelength anomalous diffraction (MAD) method (Timm *et al.*, 1999). The FAH structure was resolved at 1.9Å resolution and the active site was shown to contain no cysteine residues.

FAH represents the first structure of a hydrolase that acts specifically on C-C bonds and also defines a new class of metalloenzymes characterised by a unique α/β fold. Therefore, FAH is structurally distinct from other C-C bond hydrolases, such as 2-hydroxymuconic semialdehyde hydrolase (XylF; EC 3.7.1.9) and 2,6-dioxo-6-phenylhexa-3-enaote (BphD; EC 3.7.1.8), known to be members of the α/β hydrolase fold family (Nandhagopal *et al.*, 1997; Diaz and Timmis, 1995).

Crystal structures of FAH and FAH complexed with its physiological products, fumarate and acetoacetate, have been determined (Timm *et al.*, 1999). The FAH structure consists of a 120-residue N-terminal domain of unknown function and a 300-residue C-terminal domain defined by a novel β-sandwich roll structure that forms an active site in close proximity to the dimer interface (topology shown in Figure 14).

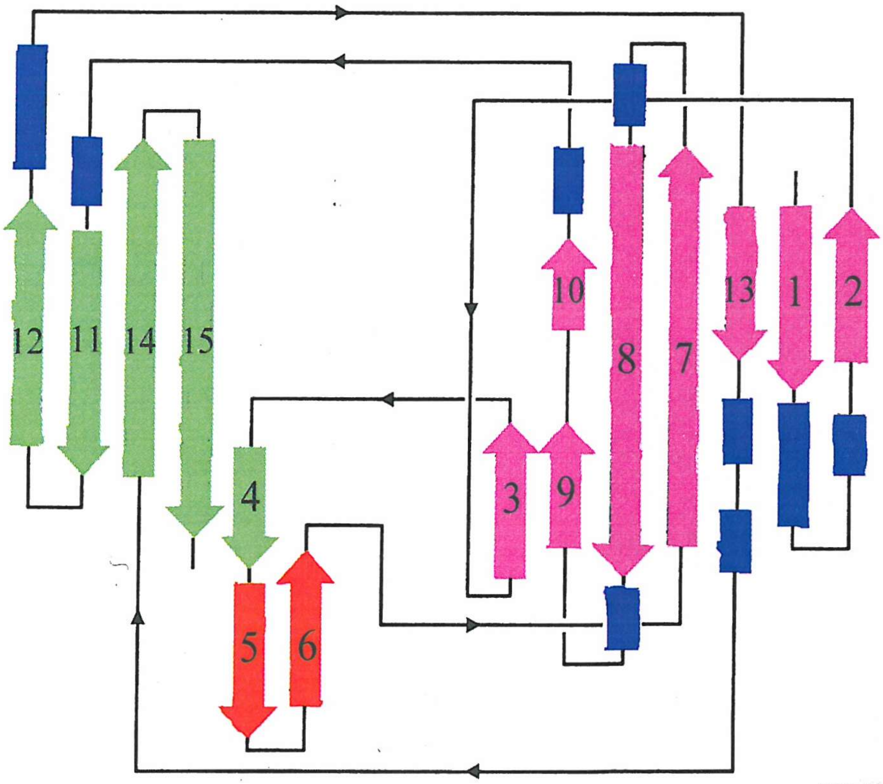


Figure 14: A topology diagram of the novel FAH β-strand arrangement. β strands are numbered according to their sequential occurrence in the polypeptide chain. Sheets A, B and C are respectively coloured in purple, green and red. α helices are represented by blue rectangles.

The FAH active site occurs in a solvent-filled cavity that is complementary in shape and charge to the fumarylacetoacetate substrate (Timm *et al.*, 1999). The base of the active site is formed by a metal ion coordinated by four side-chain carboxyl groups at the edge of the β-roll, whereas the sides are formed by helices and turns located above the β-roll. The product acetoacetate binds to FAH by providing two oxygen ligands to a Ca^{2+} ion present in the metal ion binding site. Fumarate binding involves an Arg and two Tyr side-chains near the entrance to the active site. These structural observations, along with the finding that a D233V substitution causes HT1 (St. Louis and Tanguay, 1997), led to a proposed mechanism involving active site His133 (as part of a Glu/His dyad) as a general base catalyst in the direct activation of a nucleophilic water molecule (Timm *et al.*, 1999).

The side-chains of Arg237, Gln240, and Lys253 are proposed to stabilise the formation of a tetrahedral alkoxy transition state intermediate, and the metal ion is proposed to function in binding substrate and stabilising an acetoacetate carbanion leaving group. The formation of a carbanion intermediate also accompanies the C-C bond hydrolysis catalysed by kynureninase during tryptophan degradation (Bild and Morris, 1984).

Studies on FAH with the phosphinoyl compound 4-(hydroxmethylphosphinoyl)-3-oxo-butanoic acid (HMPOBA), an FAH inhibitor, demonstrate the potential for this class of molecule to effectively inhibit a hydrolase that acts on C-C bonds (Bateman *et al.*, 2001). Phosphorus-containing compounds have previously been characterised as noncovalent inhibitors and transition state analogues of other hydrolases, including aspartyl- and metalloproteases (Phillips *et al.*, 1992; Fraser *et al.*, 1992). The FAH:HMPOBA complex shares many features with the proposed mechanism of FAH (Timm *et al.*, 1999), with the HMPOBA acetoacetyl group interacting directly with Ca^{2+} present in the catalytic metal binding site, and the tetrahedral phosphinoyl group occupying the position predicted for the alkoxy intermediate (Bateman *et al.*, 2001).

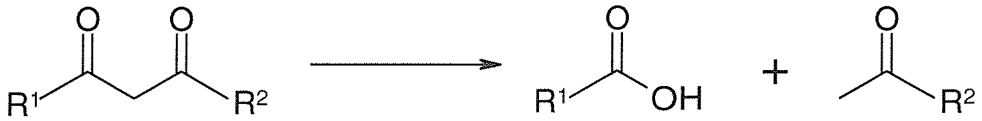
Despite the FAH structure being clearly different from the α/β hydrolase fold canonical structure, FAH does share features in common with the Ser-His-Asp catalytic triad (Timm *et al.*, 1999). The distances and angles made between the proposed FAH nucleophilic water, the His133 imidazole N atoms and the Glu364 carboxyl O atoms are very similar to those of the serine protease catalytic triad.

1.10.2. Uses of carbon-carbon bond hydrolases in organic synthesis

Since the mid-1980s the synthesis of homochiral substances by enzymatic transformations has become very important and more widespread (Tramper *et al.*, 1985). This field of study has been dominated by hydrolytic enzymes, including lipases, esterases, amidases, glucosidases and nucleases, to give enantiomerically pure compounds not easily obtainable by conventional organic synthesis. This is due to enzymes regio- and stereoselectivities which are useful for kinetic resolutions and asymmetric synthesis (Porter and Clark, 1985). Hydrolases have now become essential reagents for many specific carbon-heteroatom bond fissions and fusions in chemistry, and are very widely available as they are produced by all free-living organisms. Thus, many of the carbon-heteroatom hydrolases are available commercially and under appropriate reaction conditions, such as low water activity, can catalyse synthetic and transfer reactions.

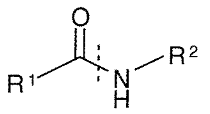
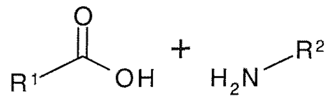
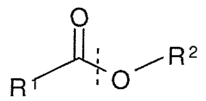
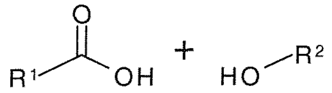
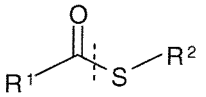
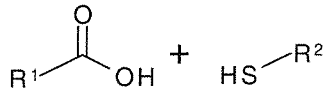
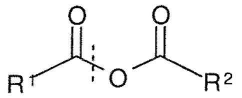
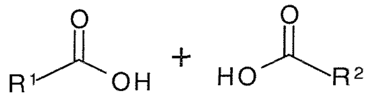
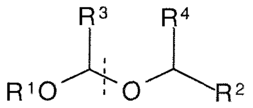
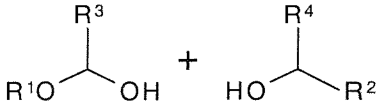
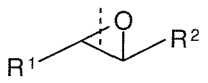
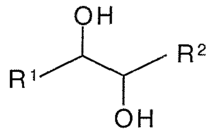
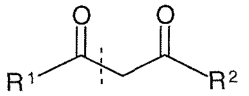
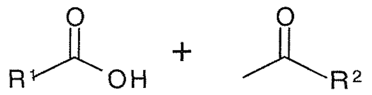
In contrast to carbon-heteroatom bond cleavage, C-C bond hydrolytic cleavage (analogous to a retro-Claisen condensation with carboxylate anions) is considered a relatively rare enzymatic reaction type, and has been little studied (Pokorny *et al.*, 1997).

In mode of action, β-ketolases may not differ much from the carbon-heteroatom hydrolases (Box 2). The methylene carbon situated between two adjacent carbonyls assumes the equivalent of a carbanion structure, and this resembles the oxygen and nitrogen in esters, anhydrides and amides by providing an electron pair for the formation of the methyl ketone and carboxylate products:

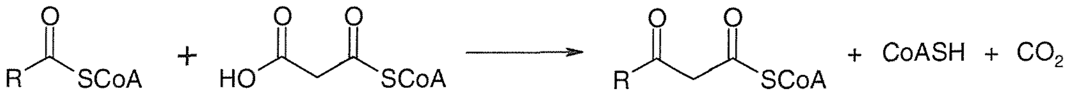


β-ketolases are not available commercially, and therefore, unlike carbon-heteroatom hydrolases, have not been used in organic synthesis despite their potential. Nevertheless, the β-ketolase group of enzymes, and their vinylogous counterparts, are possibly as widespread as other hydrolases, and could conceivably be useful for the production of polyketides (Pokorny *et al.*, 1997).

Box 2

Substrate type	Products	Enzymes (generic)
		Proteases Peptidases Amidases
		Lipases Esterases
		
		
		Glycosidases
		Epoxide hydrolases
		β-Ketolases

Natural products are an important class of compounds derived from polyketide biosynthetic pathways. These biosynthetic reactions, catalysed by synthetases using thioester substrates, are driven towards synthesis by concomitant decarboxylation of the donor malonyl thioester and its addition to the growing β-ketide chain:



The variety of products formed is dependent on chain elongation with aceto-units, similar in principle to the C-C bond formation in fatty acid biosynthesis (Pokorny *et al.*, 1997).

The specificities of the enzymes involved allow stereospecific modifications of the growing β-oxochain. The incorporation of branched chains, by using alternative thioester precursors, at precise positions in the extension process, gives stereochemically discrete asymmetric centres. Reductases then provide secondary alcohols. Aromatase enzymes and other cyclases catalyse the formation of ring structures from these aliphatic polyketides, leading to many compounds of biological and pharmaceutical interest, such as antibiotics, feed additives and anticancer agents (McDaniel *et al.*, 1995), as shown in Figure 15.

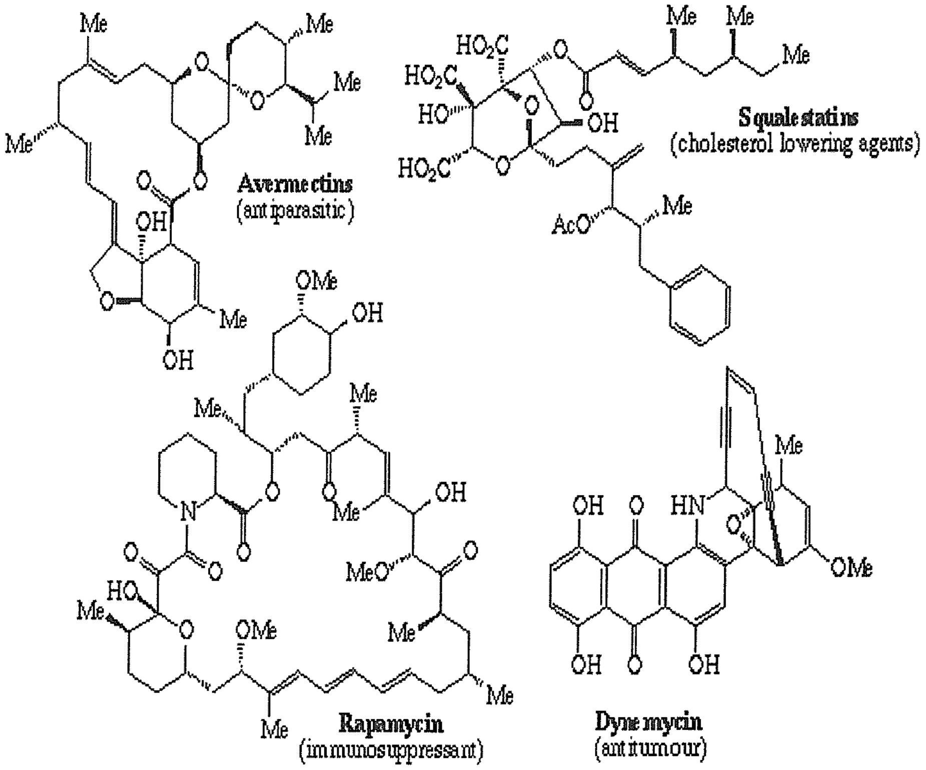


Figure 15: Polyketide derived metabolites (Simpson, 1995).

CHAPTER 2

METHODS AND MATERIALS

2.1. Crystallographic methods

2.1.1. Brief introduction to Crystallography

X-ray crystallography is an experimental technique in which a detailed picture of a molecule can be obtained by the interpretation of the diffraction of X-rays from many identical molecules in an ordered array, like a crystal. Many thousands of protein and nucleic-acid structures have been obtained by this method.

Visible light cannot produce an image of individual atoms in protein molecules, in which bonded atoms are about 1.5Å apart, whereas X-rays are of the correct wavelength and so are diffracted by even the smallest molecules. X-ray analysis of proteins seldom resolves the hydrogen atoms; however, their positions can be deduced based on the assumption that bond lengths, bond angles, and conformational angles are like those in small organic molecules (Rhodes, 2000).

Although individual atoms do diffract X-rays, it is not possible to produce a focused image of a molecule as, firstly, X-rays cannot be focused by lenses, and secondly, a single molecule is a very weak scatterer of X-rays. The first problem is overcome by measuring the directions and intensities of the diffracted X-rays, and then using a computer to simulate an image-reconstructing lens. The second problem is met by analysing diffraction from crystals, rather than individual molecules, as most X-rays will pass through a single molecule without being diffracted producing diffracted beams too weak to detect. A crystal of a protein contains many ordered molecules in identical orientations, so each molecule diffracts identically, and the diffracted beams augment each other to produce strong, detectable X-ray beams.

Therefore, the use of X-ray crystallography to determine a protein's structure involves the growing of high-quality crystals of the purified protein, measuring the directions and intensities of X-ray beams diffracted from the crystals, and using a computer to simulate the effects of an objective lens and thus produce an image of the crystal contents. Finally, that image must be interpreted by the building of a molecular model, using computer graphics, that is consistent with the image (Figure 16).

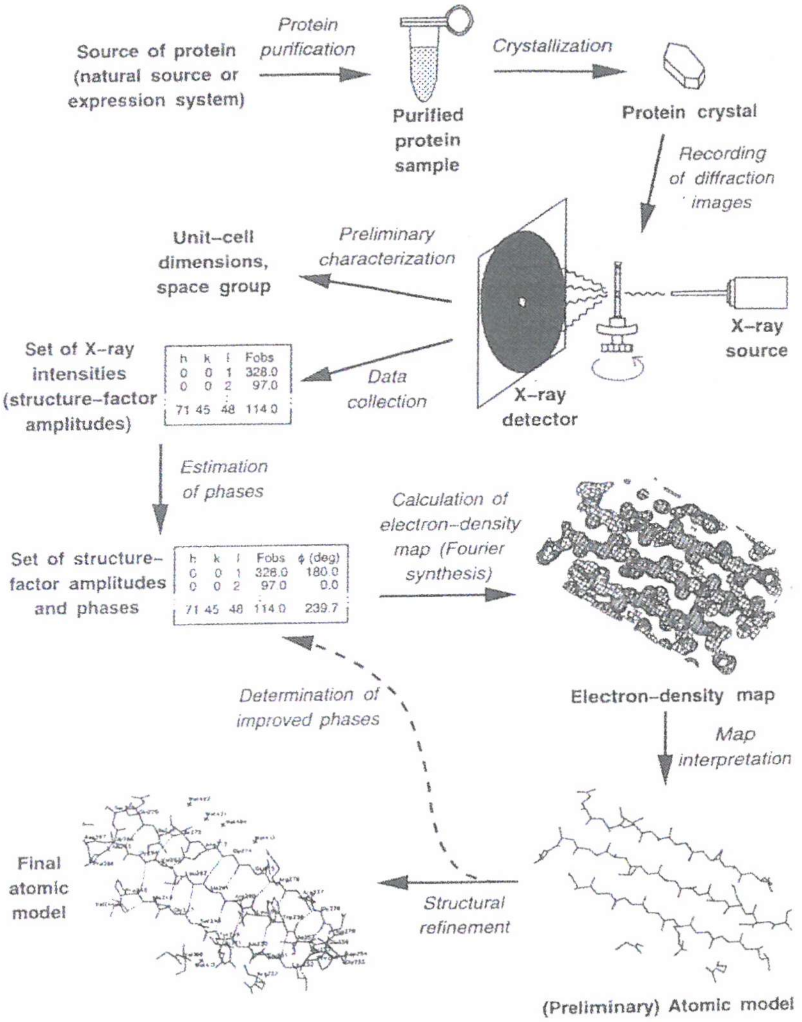


Figure 16: An overview of protein structure determination by X-ray crystallography (*Source:* Louie, 1996).

2.1.2. Crystals and unit cells

Crystals are an orderly three-dimensional array of molecules, held together by non-covalent interactions. The smallest and simplest volume element that is completely representative of a whole crystal is the unit cell. The whole crystal, therefore, is an efficiently packed array of many unit cells stacked beside and on top of one another. Figure 17a shows such a crystalline array of molecules. The array of points at the corners of unit cells is called the lattice. From crystallography, we obtain an image of the electron clouds that surround the molecules in the average unit cell in the crystal. This image should then allow the location of all atoms in the unit cell, these atoms then being given a set of three-dimensional Cartesian coordinates, x, y, and z. One of the vertices (a lattice point or any other convenient point) is used as the origin of the unit cell's coordinate system and is assigned the coordinates x=0, y=0, and z=0, usually written (0,0,0). See Figure 17b.

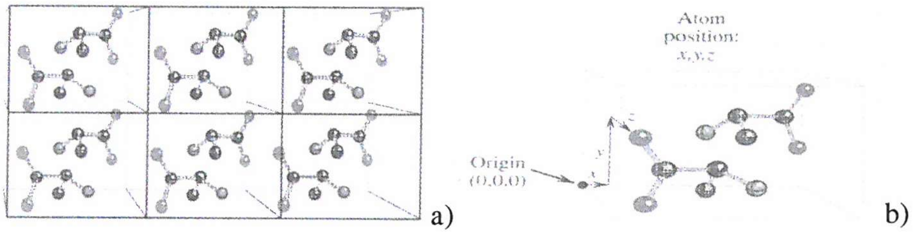


Figure 17: a) Six unit cells in a crystalline lattice. Each unit cell contains two molecules of alanine (hydrogen atoms not shown) in different orientations; b) One unit cell from 17a. The position of an atom in the unit cell can be specified by a set of spatial coordinates x, y, z (Rhodes, 2000).

2.1.3. Practically producing crystals

Protein crystals are generally grown by slow, controlled precipitation from aqueous solution under conditions that do not denature the protein. One commonly used technique is vapour diffusion, in which the protein/precipitant solution is allowed to equilibrate in a closed container with a larger aqueous reservoir whose precipitant concentration is optimal for producing crystals (Blundell & Johnson, 1976). It is generally impossible to grow protein crystals as large as 1mm in their shortest dimension. In addition, larger crystals tend to be twinned, where two or more crystals have grown into each other at different orientations, and therefore unusable. A crystal of at least 0.2mm in the shortest dimension is normally required for successful crystallography. Protein crystals are usually harvested, examined, and mounted for crystallography within their mother liquor, the solution in which they formed (Rhodes, 2000).

The derivative crystals needed for phase determination and for studying protein-ligand interactions can be obtained by two methods – co-crystallising protein and ligand, or soaking preformed crystals in mother liquor solutions containing ligand.

Many variables influence the formation of macromolecular crystals, such as protein purity, pH, and temperature, as well as more subtle ones like vibration and sound, convection, source and age of the protein, and the presence of ligands.

2.1.4. Lattices: Real and Reciprocal Space

A diffraction pattern consists of reflections (spots) in an orderly array (Figure 18). There is a simple inverse relationship between the spacing of unit cells in the crystalline lattice, called the real lattice, and the spacing of reflections in a diffraction pattern, which because of its inverse relationship to the real lattice, is called the reciprocal lattice. Because the real lattice spacing is inversely proportional to the spacing of the reflections, then the unit cell dimensions, in Ångstroms, can be calculated from the spacings of the reciprocal lattice. This is the basic relationship which connects the submicroscopic dimensions of the crystal to the macroscopic dimensions of the diffraction pattern (Rhodes, 2000).

Figure 18: Diffraction pattern from a crystal of the MoFe (molybdenum-iron) protein of the enzyme nitrogenase from *Clostridium pasteurianum*. The reflections lie in a regular pattern, but their intensities (darkness of spots) are highly variable (Rhodes, 2000).

Each reflection can be assigned three coordinates or indices in the imaginary three-dimensional space of the diffraction pattern, called reciprocal space. The position of an individual reflection in the reciprocal space of the diffraction pattern is designated by the symbols *h*, *k*, and *l*. The central reflection is taken as the origin in reciprocal space and assigned the coordinates $(h, k, l) = (0, 0, 0)$, usually written $hkl = 000$. The other reflections are assigned whole-number coordinates counted from this origin. Thus, the parameters which can be measured and analysed from an X-ray diffraction pattern are the position *hkl* and the intensity I_{hkl} of each reflection. The position of a reflection is related to the angle by which the diffracted beam diverges from the source beam. For a unit cell of known dimensions, the angle of divergence uniquely specifies the indices of a reflection.

The dimensions of a unit cell are designated by six numbers: the lengths of three unique edges **a**, **b**, and **c**; and three unique angles α, β, and γ (Figure 19a). There are seven different types of cells, depending on the relationship between their unique cell edges and their unique angles, such as triclinic, orthorhombic and monoclinic. If the unit cell contents are symmetrical, then the reciprocal lattice is also symmetrical and certain sets of reflections are equivalent. Therefore, awareness of unit cell symmetry can greatly reduce the magnitude of data collection, as only one member of each set of equivalent reflections needs to be measured. The symmetry of a unit cell is described by its space group, which is represented by a symbol, such as $P2_12_12_1$, in which a capital letter indicates the lattice type and other symbols represent symmetry operations that can be carried out on the unit cell without changing its appearance.

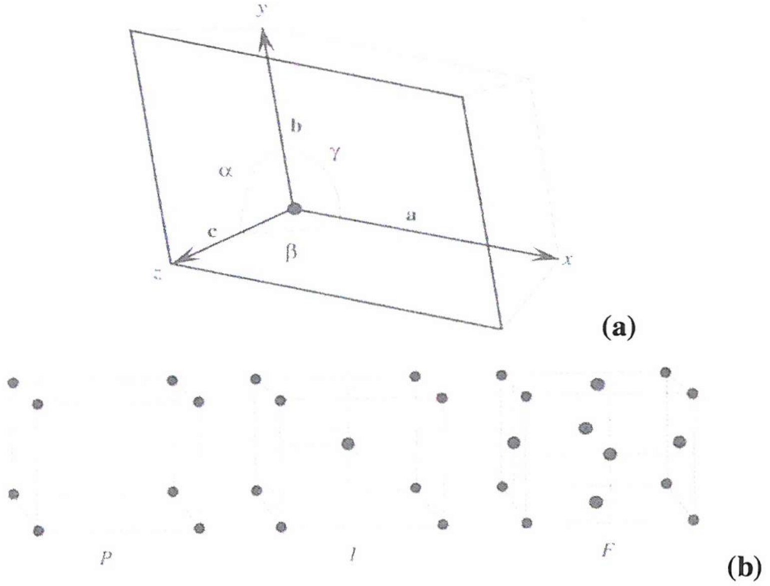


Figure 19: (a) General (triclinic) unit cell, with edges **a, b, c** and angles α, β, γ; **(b)** *P* (primitive lattice), *I* (body-centred or internal lattice), and *F* (face-centred lattice) unit cells (Rhodes, 2000).

There are exactly 230 possible space groups. There are three lattice types: primitive lattices containing a lattice point at each corner of the unit cell; body-centred lattices, with an additional lattice point in the centre of the cell; and face-centred lattices, with additional lattice points on the centre of each face (Figure 19b). An example of a symmetry operation is rotation of an object about an axis, with the axis itself becoming a symmetry element. Protein molecules are inherently asymmetric, being composed of chiral amino acid residues coiled into larger chiral structures such as helices or twisted sheets. If one protein molecule occupies a unit cell, then the cell itself is chiral, and there are no symmetry elements. However, in most cases, the unit cell contains several identical molecules or complexes in an arrangement that produces symmetry elements. In the unit cell, the largest aggregate of molecules that possesses no symmetry elements, but can be juxtaposed on other identical entities by symmetry operations, is called the asymmetric unit. In the simplest case, the asymmetric unit is a single protein molecule. The simplest symmetry operations needed to describe unit cell symmetry are translation, rotation, and reflection. Combinations of these operations can produce more complex elements, such as screw axes which are rotations and translations combined. Certain symmetry elements in the unit cell announce themselves in the diffraction pattern by causing specific reflections to be missing (intensity of zero). These patterns of missing reflections are called systematic absences, and they allow the determination of the space group of a crystal by looking at a few crucial planes of reflections (Rhodes, 2000).

Summarising briefly, in crystallography there are two different coordinate systems. The first system is the unit cell (real space), where an atom's position is described by its coordinates x, y, and z. A vertex of the unit cell, or any other convenient position, is taken as the origin, with coordinates $x, y, z = (0, 0, 0)$. Coordinates in real space designate real spatial positions within the unit cell and are given in Ångstroms or nanometers, or in fractions of unit cell dimensions. The second system is the three-dimensional diffraction pattern (reciprocal space), where a reflection's position is described by its indices *hkl*. The central reflection is taken as the origin with the index 000. The position of a reflection is designated by counting reflections from 000, so the indices h, k, and l are integers. Distances in reciprocal space, measured in reciprocal Ångstroms or reciprocal nanometers, are used to judge the potential resolution that the diffraction data can yield.

2.1.5. The Mathematics: Finding electron density from reflected waves

Each reflection is produced by a beam of electromagnetic radiation, so the lens-simulating computer programs have to treat the reflections as waves and recombine them to produce an image of the molecules in the unit cell. As the reflections result from the complicated diffraction of molecules in the unit cell then the resulting wave will be complicated also. Waves can be described by three constants - the amplitude F, the frequency λ, and the phase α. Even the most complicated waves can be described with this same simplicity using a sum called a Fourier series, where each simple sine or cosine function in the sum is called a Fourier term. A method called Fourier synthesis is used to compute the sine and cosine terms that describe a complex wave (Rhodes, 2000).

Each diffracted X-ray that produces a recorded reflection can also be described as the sum of the contributions of all scatterers in the unit cell. The sum that describes a diffracted ray is called a structure-factor equation. The computed sum for the reflection *hkl* is called the structure factor F_{hkl}. The structure-factor equation implies that each reflection is the result of diffractive contributions from all atoms in the unit cell. That is, every atom in the unit cell contributes to every reflection in the diffraction pattern. The structure factor is a wave created by the superposition of many individual waves, each resulting from diffraction by an individual atom (Rhodes, 2000).

When a crystal is exposed to an X-ray beam the actual diffractors of the X-rays are the clouds of electrons in the molecules of the crystal. Diffraction therefore reveals the distribution of electrons, or the electron density, of the molecules, which in turn reflects the molecules shape. Because protein molecules are ordered, and because in a crystal they are in an ordered array, the electron density in a crystal can be described by a periodic function. The surface features and overall shape of all molecules in the unit cell, as described by the electron density, is therefore a three-dimensional periodic function. This function can be shortened to $\rho\ (x, y, z)$, which specifies a value ρ for electron density at every position x, y, z in the unit cell. An electron-density map is a graph of this function, which is basically a drawing of a surface along which there is constant electron density (a contour map). This map is therefore a basic image of the molecules in the unit cell, and the goal of crystallography (Figure 20).

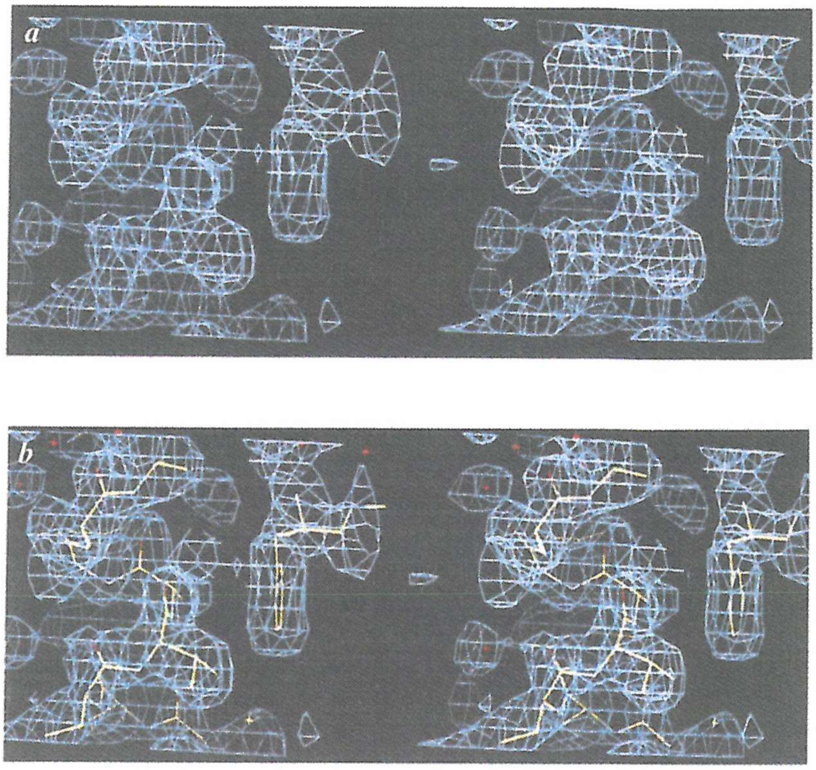


Figure 20: (a) Small section of molecular image (electron density map) displayed on computer graphics terminal; **(b)** Image (a) is interpreted by building a molecular model within the image. Computer graphics programs allow parts of the model to be added and their conformations adjusted to fit the image (Rhodes, 2000).

A mathematical operation called the Fourier transform is used to solve the structure-factor equations for the desired function $\rho\ (x, y, z)$. The Fourier transform describes precisely the mathematical relationship between an object and its diffraction pattern. Alternatively it could be said that the Fourier transform is the lens-simulating operation that a computer performs to produce an image of molecules (or more precisely, of electron clouds) in the crystal. This view of $\rho\ (x, y, z)$ as the Fourier transform of the structure factors implies that the measurement of just three parameters - amplitude, frequency, and phase - of each reflection will give the function $\rho\ (x, y, z)$ and hence enable the production of electron-density maps (Rhodes, 2000).

A structure factor describes one diffracted X-ray, which produces one reflection received at the detector. A structure factor F_{hkl} can be written as a Fourier series in which each term gives the contribution of one atom to the reflection *hkl*. The phase of a wave is implicit in the exponential formulation of a structure factor and depends only upon the atomic coordinates of the atom. All atoms on a set of equivalent, parallel lattice planes diffract in phase with each other. Each diffracted ray is a complicated wave, the sum of diffractive contributions from all atoms in the unit cell. For a unit cell containing *n* atoms, the structure factor F_{hkl} is the sum of all the atomic f_{hkl} values for individual atoms. Thus, the structure factor for F_{hkl} can be written as follows:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)} \qquad \textbf{(1)}$$

In words, the structure factor that describes reflection *hkl* is a Fourier series in which each term is the contribution of one atom, treated as a simple sphere of electron density. So the contribution of each atom *j* to F_{hkl} depends on what element it is, which determines f_j, the amplitude of the contribution, and its position in the unit cell (x_j, y_j, z_j), which establishes the phase of its contribution.

Alternatively, F_{hkl} can be written as the sum of contributions from each volume element of electron density in the unit cell. The electron density of a volume element centred at (*x*, *y*, *z*) is, roughly, the average value of ρ(*x*, *y*, *z*) in that region. This can be written:

$$F_{hkl} = \int_{V} \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dV \qquad \textbf{(2)}$$

where the integral over *V*, the unit cell volume, is shorthand for the integral over all values of *x*, *y*, and *z* in the unit cell. Basically, F_{hkl} is the Fourier transform of ρ(*x*,*y*,*z*), or more precisely, F_{hkl} is the transform of ρ(*x*,*y*,*z*) on the set of real-lattice planes (*hkl*). All of the F_{hkl}s together compose the transform of ρ(*x*,*y*,*z*) on all sets of equivalent, parallel planes throughout the unit cell. Importantly, because the Fourier transform operation is reversible, the electron density is in turn the transform of the structure factors, as follows:

$$\rho(x,y,z)=\frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}e^{-2\pi i(hx+ky+lz)} \tag{3}$$

where V is the volume of the unit cell. Since F_{hkl} is a periodic function, it possesses amplitude, frequency, and phase. It is a diffracted X-ray, so its frequency is that of the X-ray source. The amplitude of F_{hkl} is proportional to the square root of the reflection intensity I_{hkl}, so structure amplitudes are directly obtainable from measured reflection intensities. But the phase of F_{hkl} is not directly obtainable from a single measurement of the reflection intensity. In order to compute $\rho(x,y,z)$ from the structure factors the phase of each diffracted ray must be obtained, as well as the intensity of each reflection. This is the so-called phase problem, and the phases can be obtained, normally, by using one of three crystallographic techniques.

2.1.6. Solving the Phase Problem

The most demanding element of macromolecular crystallography is the phase problem, which is that of determining the phase angle for each reflection. There are a number of common methods for overcoming this obstacle, including isomorphous replacement (the heavy atom method), anomalous scattering (also called anomalous dispersion), and molecular replacement. Each of these techniques yields estimates of phases, which must be improved before an interpretable electron density map is obtained.

2.1.6.1. Isomorphous Replacement

Each atom in the unit cell contributes to every reflection in the diffraction pattern. A specific atom contributes to some reflections strongly and to some weakly or not at all. If one or a very small number of atoms are added to identical sites in all unit cells then it would be expected to see changes in the diffraction pattern, as a result of the additional contribution of the added atom. The slight perturbation in the diffraction pattern caused by an added atom can be used to obtain initial estimates of phases (Rhodes, 2000). In order for these perturbations to be large enough to measure, the added atom must be a strong diffractor, which means it must be an element of high atomic number, a heavy atom such as mercury or gold (Hendrickson & Ogata, 1997). Heavy atom derivatives are prepared most commonly by soaking crystals of the protein in solutions of heavy ions, for instance ions or ionic complexes of Hg, Pt, or Au. In many cases, such ions bind to one or few specific sites on the protein without perturbing its conformation or crystal packing. Unit cell dimensions are quite sensitive to such disturbances, so heavy atom derivatives whose unit cell dimensions are the same as native crystals are probably isomorphous (hence isomorphous replacement). To be useful heavy atom derivatives must cause measurable changes in a modest number of reflection intensities, as this is the means by which the phase estimates are made. Also, the derivative crystal must diffract to a reasonably high resolution, although not as high as that needed for the native data. Because derivatives must be isomorphous with native crystals, the strategy is the same as that for collecting native data (Rhodes, 2000).

2.1.6.2. Anomalous Dispersion

A second means of obtaining phases from heavy atom derivatives takes advantage of the heavy atom's capacity to absorb X-rays of specified wavelength. As a result of this absorption symmetry related reflections are not equal in intensity and this inequality is called anomalous scattering or anomalous dispersion (Rhodes, 2000). An element exhibits anomalous scattering when the X-ray wavelength is near the elements absorption edge. Absorption edges for the light atoms in the unit cell are not near the wavelength of X-rays used in crystallography, so carbon, nitrogen, and oxygen do not contribute appreciably to anomalous scattering. However, absorption edges of heavy atoms are in this range, and if X-rays of varying wavelength are available, such as at synchrotrons, then X-ray data can be collected under conditions that maximise anomalous scattering by the heavy atom. When the X-ray wavelength is near the heavy atom absorption edge, a fraction of the radiation is absorbed by the heavy atom and re-emitted with altered phase. These differences can be used to establish the phase of a reflection in the heavy atom derivative data, thus establishing the phase of the corresponding reflection in the native data. Three developments - variable-wavelength synchrotron X-rays, cryocrystallography, and the production of proteins containing selenomethionine instead of the normal sulphur-containing methionine - have allowed rapid progress in maximising the information obtainable from anomalous dispersion (Ogata, 1998). Proteins lacking functional heavy atoms, such as iron in a globin, can be expressed in *E. coli* containing exclusively selenomethionine. The selenium atoms then serve as heavy atoms. Data sets at different wavelengths, taken at the heavy atoms absorption maximum, the edge, and at wavelengths distant from the maximum, from a single crystal can contain sufficient phasing information to solve a structure. This method is called multiwavelength anomalous dispersion, or MAD, phasing (Walsh *et. al.*, 1999).

2.1.6.3. Molecular Replacement

The phases from structure factors of a known protein can sometimes be used as initial estimates of phases for a new protein. If this is the case, then it may be possible to determine the structure of the new protein from a single native data set. In this method, known as molecular replacement, the known protein is the phasing model and the initial phases are calculated by placing a model of the known protein in the same unit cell as the new protein (Rhodes, 2000). For instance, if a new protein shares significant sequence homology with a known protein then this suggests that it's structure may be similar. Therefore, the known protein might be used as the phasing model for the new protein. Also, having once learnt the structure of a protein it then becomes possible to study protein/ligand complexes of this molecule by using the ligand-free protein as a phasing model for the protein-ligand complex. That is, of course, if the protein/ligand complex is similar in structure to the free protein. If the new crystals and those of the model are isomorphous, the model phases can be used directly as estimates of the desired phases. If not, then the known protein must be superimposed onto the new protein to create the best phasing model. 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 21).

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:

$$X' = [C]X + t \qquad (4)$$

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, while the translation vector (t) is defined by three translational parameters (t_x, t_y, t_z).

It is possible to search for all six parameters at the same time but this is time-consuming even for the fastest computers. A solution was found by splitting the six parameter search into two parts, using Patterson map comparisons to find the best orientation of the model protein, the rotation function, and then again to find the best position of the model protein, the translation function (Rossman & Blow, 1962).

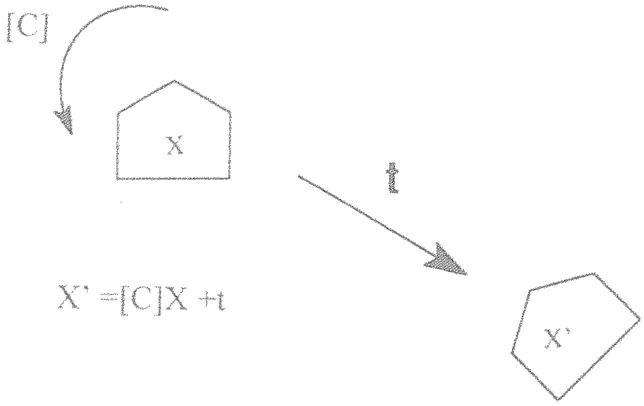


Figure 21: Diagram illustrating the molecular replacement method by superimposing the known structure (X) on the target structure (X') in its unit cell. The transformation of X to X' is described by X'=[C]X+t where [C] is a rotation matrix that rotates co-ordinates X into a new orientation and t is the translation vector defining the position. The success of the procedure depends on determining 3 orientational parameters (rotation function) and 3 translational parameters (translation function). (Mohammed, 2001)

2.1.7. Patterson-based Molecular Replacement

Traditional molecular replacement methods are based on the Patterson function, as this can be computed without phase information. The Patterson-based method is useful as it can be used to make the six-dimensional problem into two three-dimensional problems, thus making the solution easier and quicker to obtain. The orientation (three parameters) is determined first, followed by the translation (also three parameters).

The Patterson map is a vector map, with peaks at the positions of vectors between atoms in the unit cell. The vectors in a Patterson map can be divided into two categories. Intramolecular vectors, from one atom in the molecule to another atom in the same molecule, depend only on the orientation of the molecule, and not on its position in the cell, so these can be exploited in the rotation function. Intermolecular vectors depend both on the orientation of the molecule and on its position so, once the orientation is known, these can be exploited in the translation function (Read, 2000).

2.1.7.1. Rotation function

The Patterson function does not depend on any choice of origin in the molecule. The origin of the Patterson function is the point where vectors of zero length are found. It is this feature that allows it to be used to search for a rotation, independently of any translations that may have occurred. On average, the intramolecular vectors will be shorter than the intermolecular vectors, so the rotation function can be computed using only the part of the Patterson map near the origin.

The rotation function is commonly defined as a product function. When the model Patterson is rotated and superimposed on the observed Patterson, it picks out the parts of the observed Patterson which arise from intramolecular vectors. It does not match the rest of the function. All the peaks that fit must be within a sphere, the radius of which is the maximum dimension of the model Patterson (Read, 2000).

The procedure, then, is to calculate the Patterson function of the model molecule, and to rotate it over the Patterson function of the target protein, keeping the origins at the same point. For every possible rotation, the agreement between the two Pattersons is evaluated, using only the region within a radius r of the origin. If the two agree well within this radius then the orientation of the model and the target protein are probably matched. The value for r is usually chosen to be slightly less than the maximum dimension of the molecule, as this eliminates many cross-vectors.

A rotation function will always have many peaks, some larger than others. Since it is possible for the correct peak to be obscured by noise, statistical tests are usually made. Most frequently, the mean and standard deviation of all points in the rotation function are computed, and the significance of the highest peak is demonstrated by stating its deviation from the mean. Often the magnitude of the next highest peak is stated, which must usually be assumed to represent random noise (Blow, 2002).

2.1.7.2. Translation function

The next step is to decide where the origin of the model molecule must be placed, to generate a crystal structure with diffraction corresponding closely to that of the target protein. This step uses a translation function, which searches for the three positional parameters when the three rotational parameters have been determined. Once again, it uses the Patterson function, so that phase information is not needed. In essence, the search consists in placing the origin of the search model at all positions in the unit cell of the target protein. For each position, the predicted configuration of Patterson vectors is compared to the actual Patterson function of the target protein. Compared to the rotation function, the translation function can be computed more rapidly. In the majority of cases, a correct peak in a translation function is overwhelmingly significant (Blow, 2002).

2.1.8. Collecting diffraction data

X-ray diffraction data is collected by mounting a crystal between a X-ray source and a X-ray detector (Figure 22). The crystal lies in the path of a narrow beam of X-rays coming from the source. A simple detector is X-ray film, which when developed exhibits dark spots where X-ray beams have impinged. The crystal diffracts the source beam into many discrete beams, each of which produces a distinct reflection on the film. These spots are called reflections because they emerge from the crystal as if reflected from planes of atoms. The greater the intensity of the X-ray beam that reaches a particular position the darker the reflection.

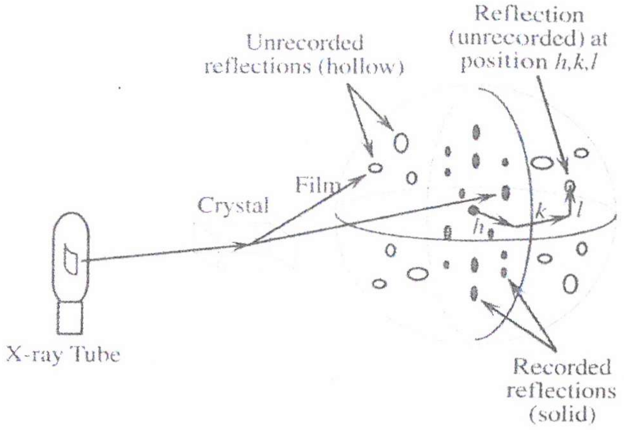


Figure 22: Crystallographic data collection, showing reflections measured at one particular crystal orientation (solid, on film) and those that could be measured at other orientations (hollow, within the sphere but not on the film). (Rhodes, 2000).

Data collection must be carried out as rapidly and efficiently as possible because crystals, especially those of macromolecules, can deteriorate in the X-ray beam. Therefore, as many reflections as possible are captured, during every moment of irradiation. The diffracting power of the crystal can also limit the number of available reflections. Protein crystals that produce measurable reflections down to about 3Å or less are usually suitable for structure determination.

The position and intensity of each reflection is measured. The position of a reflection can be used to obtain the direction in which that particular beam was diffracted by the crystal. The intensity of the reflection is obtained by measurement of the absorbance of the spot, giving a measure of the strength of the diffracted beam that produced the spot. These two parameters, the beam intensity and direction, for each diffracted beam, are required to reconstruct an image of the molecules in the unit cell (Rhodes, 2000).

In producing an image of molecules from crystallographic data, the computer simulates the action of the lens, computing the electron density within the unit cell from the list of indexed intensities. Computation of the Fourier transform is the lens-simulating operation that a computer performs to produce an image of molecules in the crystal.

2.1.8.1. Producing X-rays

X-rays are electromagnetic radiation of wavelengths 0.1-100Å and can be produced by bombarding a metal target (commonly copper or molybdenum) with electrons (Rhodes, 2000). There are three common X-ray sources, X-ray tubes (like a cathode ray tube), rotating anode tubes, and particle storage rings, which produce synchrotron radiation in the X-ray region. In the X-ray tube, electrons from a hot filament (cathode) are accelerated by electrically charged plates and collide with a water-cooled anode made of the target metal. X-rays are produced at low angles from the anode. Output from X-ray tubes is limited by the amount of heat that can be dissipated from the anode (Drenth, 1994). Higher X-ray output can be obtained from rotating anode tubes, in which the target is a rapidly rotating metal disk. The electron bombardment is spread over a larger area and so improves heat dissipation, resulting in rotating anode sources being ten times as powerful as fixed anode sources (Drenth, 1994). Particle storage rings are the most powerful sources of X-rays. In these giant rings, electrons or positrons circulate at velocities near the speed of light and are maintained in circular motion by powerful magnets. When forced into such curved motion a charged particle, like an electron, emits energy (synchrotron radiation). This energy is emitted as X-rays, whose intensity can be increased by additional bending of the beam, after which systems of focusing mirrors and monochromators provide powerful monochromatic X-rays at selectable wavelengths (Giacovazzo *et. al.*, 1992). Advantages of using synchrotron sources are that X-ray data that requires several hours of exposure to a rotating anode source can often be obtained in seconds or minutes at a synchrotron source, and that X-rays of selectable wavelength can be helpful in solving the phase problem.

Whatever the source of X-rays, the beam can be further reduced with systems of metal plates called focusing mirrors and then directed through a collimator which produces a narrow beam. During data collection the direct beam is blocked just beyond the crystal by a piece of metal called a beamstop, which prevents excessive radiation from reaching the X-ray detector, thus obscuring low-angle reflections.

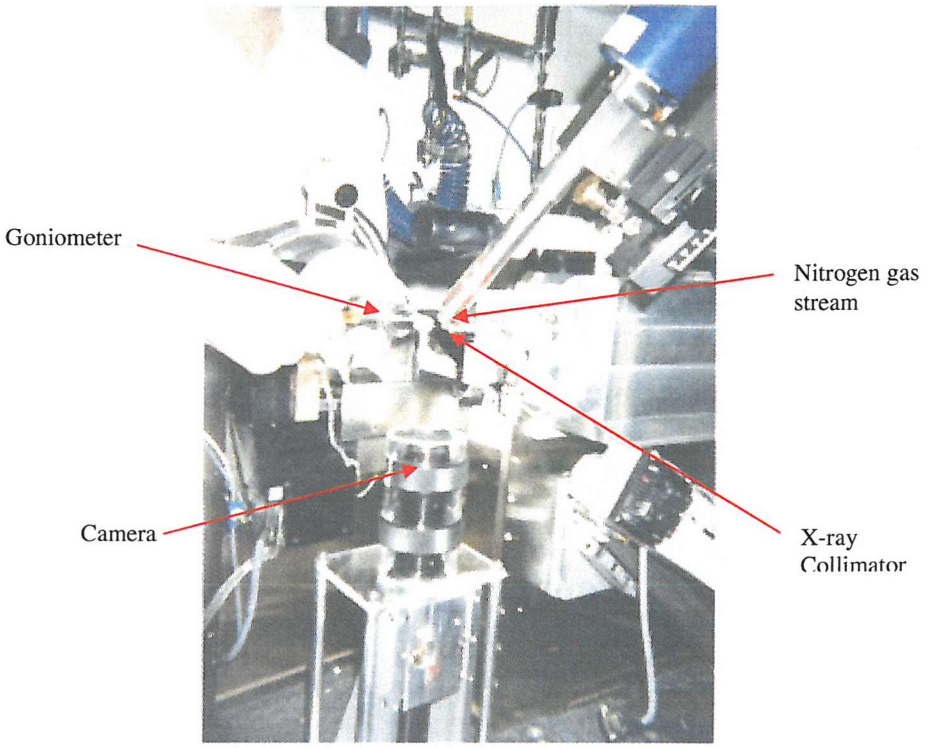


Figure 23: Photo taken at ESRF beamline ID29 illustrating the mounting of protein crystals. The crystal position is adjusted with the aid of a TV camera so that it is within the nitrogen gas stream and so that the X-ray beam will go directly through it. The goniometer automatically rotates the crystal with retention of centring.

2.1.8.2. X-ray detector

The simplest X-ray detector is X-ray sensitive film, but this has almost been completely replaced by image plates and CCD detectors (Rhodes, 2000). Image plate detectors can store diffraction images reversibly and have the capacity to record reflections of widely varying intensity. Image plates are plastic sheets with a coating of small crystals of a phosphor. The phosphor crystals can be stimulated by X-rays into a stable excited state and further stimulation by visible light can cause the production of visible light in proportion to the intensity of the previously absorbed X-rays (Helliwell, 1992). After X-ray exposure, data are read from the plate by a scanner in which a fine laser beam induces luminescence from a very small area of the plate, and a photocell records the intensity of emitted light. The intensities are fed to a computer, which can then reconstruct an image of the diffraction pattern. Image plates can be erased by exposure to bright visible light and reused indefinitely (Garman, 1996). The latest X-ray detectors employ charge-coupled devices (CCDs) as detectors. In effect, CCDs are photon counters, solid-state devices that accumulate charge in proportion to the amount of light that strikes them. For crystallographic data collection, CCDs are coated with phosphors that emit visible light in response to X-rays (Westbrook & Naday, 1997).

The irradiated crystal is mounted on a device for precisely orienting the crystal so as to direct specific reflections towards the detector (Figure 23). The task of this device is to rotate a crystal through a series of known orientations, causing specified reciprocal lattice points to pass through the sphere of reflection and thus produce diffracted X-ray beams. In all forms of data collection, the crystal is mounted on a goniometer head, a device that allows the crystallographer to set the crystal orientation precisely (Rhodes, 2000). Protein crystals, either sealed in capillary tubes with mother liquor or flash-frozen in a fibre loop, are mounted on the goniometer head, which is adjusted to centre one face of the crystal perpendicular to the X-ray beam and to allow rotation of the crystal while retaining centring. Flash-frozen crystals are held in a stream of cold nitrogen gas (cryo-crystallography).

2.1.9. Data processing: A three-step cycle

The goal of data collection is a set of measured, indexed intensities for as many of the reflections as possible. Scaling and postrefinement are the final stages in producing a list of internally consistent intensities for most of the available reflections (Rhodes, 2000).

Obtaining a detailed molecular model of the unit cell contents entails calculating the electron density around the molecule using measured intensities from the native data set and phases computed from heavy atom data, anomalous scattering, or molecular replacement. Once the rough estimates of phases are obtained using one or more of these methods, then an iterative process is used to improve further these phases. This iteration entails a three-step cycle. In the first step, an estimated $\rho(x,y,z)$, that is a crude model of the structure, is computed using the electron density equation 3 with observed intensities (I_{obs}) and estimated phases (α_{calc}). In step two, attempts are then made to improve the model by viewing the electron density map and identifying molecular features such as the molecule-solvent boundaries or specific groups of atoms (map interpretation). In the final step, new structure factors (F_{calc}) are computed with the improved $\rho(x,y,z)$ model, containing only those atoms that can be located with some confidence in the electron density map. Calculation of new F_{calc}s produces a new, hopefully better, set of estimated phases, and the cycle is repeated. In each cycle it is hoped to obtain an improved $\rho(x,y,z)$, which means a more detailed and interpretable electron density map, and thus a more complete and accurate model of the desired structure. The progress of the structure determination can also be assessed by monitoring this iterative process to see if it is converging towards improved phases and improved $\rho(x,y,z)$. The computed structure factors F_{calc} include both the desired phases α_{calc} and a new set of intensities. These new intensities are referred to as calculated intensities (I_{calc}) to distinguish them from the measured reflection intensities (I_{obs}) taken from the diffraction pattern. As the iteration proceeds, the values of I_{calc} should approach those of I_{obs}, so that it is possible to see whether the iteration is converging by comparing the two values at each cycle. When cycles of computation provide no further improvement in correspondence between calculated and measured intensities, then the process is complete (Rhodes, 2000). This iterative process is known as structure refinement and is carried out using a number of computer packages, with the electron density maps being constructed and manipulated with computer graphics software.

2.1.10. Computational software used in data processing and structure refinement

Common computational software used in crystallographic data processing and refinement are the CCP4 (CCP4, 1994) and CNS (Brunger *et. al.*, 1998) packages. Programs from these packages will be described in more detail in the following section, allowing the complete process to be illustrated, from initial diffraction pattern data to the final modelled structure of the macromolecule.

Initial data processing proceeds with MOSFLM (CCP4), which allows auto-indexing of diffraction data and performs subsequent positional refinement and intensity integration of the spots (Leslie, 1992, 1994; Steller *et. al.*, 1997). The spot positions are found and then used to compute the corresponding reciprocal-space lattice vectors. Basically, each spot (reflection) of the diffraction data is assigned a number corresponding to h, k, and l, the Miller Indices. Auto-indexing can be used to determine the cell dimensions and space group of the crystal. Once these are known, predicted positions for all spots in the reciprocal-space lattice can be obtained. Once the spot positions have been determined then the intensities of the individual diffraction spots are calculated. These are then placed in numerical order according to the values of their Miller Indices using the program SORTMTZ (CCP4). The program SCALA (CCP4) then applies scale factors to and merges the data. The output from SCALA is then fed to the program TRUNCATE (CCP4), which estimates the number of molecules per unit cell and calculates the all-important structure factor amplitudes. During data processing the relative intensities of symmetry related reflections are used as a measure of data quality, and are known as the merging R_{factor} (R_{merge}). For crystals yielding high quality data, the R_{merge} is typically between 3 and 10%.

The structure factor amplitudes thus found can be used with phases determined from molecular replacement, anomalous dispersion or isomorphous replacement to give an initial electron density map. A popular package for molecular replacement is MOLREP (Vagin and Teplyakov, 1997) from the CCP4 suite of programs. 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. The rotation and translation functions performed in MOLREP are essentially overlapping functions between target and calculated model Pattersons. The noise level in the rotation function is reduced because the program computes the rotation function for three different orientations of the search model and averages them. 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.

CNS is a package commonly used to refine protein structures, and contains programs that carry out a variety of refinement protocols. These include RIGID which refines the protomer positions within the unit cell as separate rigid bodies, and ANNEAL, which performs simulated annealing, where the model is allowed to move as if at high temperature and then slowly cooled to find its preferred conformation. Temperature factor refinement of all atoms in the model can be carried out with programs such as BGROUP and BINDIVIDUAL. Following refinement electron density maps can be calculated using a program such as MODEL MAP. These maps can be viewed using graphical display packages such as QUANTA (Molecular Simulations Inc.), in which changes can be made to the protein structure so that it fits the electron density more accurately (called model building). Then, further cycles of refinement and model building can be carried out until a final, structurally sound protein model is obtained.

The improvement in the quality of the model during refinement can be followed by monitoring changes in the crystallographic R_{factor}. As refinement converges to the correct structure, the agreement between observed and calculated structure factor amplitudes also improves. Initially, an early model will have an R_{factor} of around 0.4, which should hopefully improve to at least 0.2 following several rounds of refinement and model building, as long as the data was collected to a resolution approaching 2.5Å. An R_{factor} of 0.6 is obtained when a set of observed amplitudes is compared with a set of random amplitudes, and indicates that the model is extremely poor. The crystallographic R_{factor} is almost always monitored alongside that of the free R_{factor} (R_{free}) as this overcomes the problem of model bias during refinement. The R_{free} is calculated using 5 to 10% of the total reflections, which are set aside and not included in the overall refinement (Brunger, 1992).

2.2. MhpC crystallographic data collection and processing

Synchrotron diffraction data from native MhpC crystals, ligand-soaked crystals and MhpC co-crystallised with ligands were collected with a MAR CCD detector on beamlines ID29 and ID14-2 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France and on beamline 9.6 at Daresbury Laboratory, Warrington, UK. Data were also collected 'in-house' using a 300mm MAR Image Plate X-ray detector, mounted on an ENRAF NONIUS FR591 rotating anode X-ray generator, which was operated at 50 kV and 100mA.

Intensity data from all crystals were processed with the program MOSFLM. Programs of the CCP4 suite were used to sort, merge and scale the data, with the intensities eventually converted to structure factor amplitudes. Molecular replacement was carried out with the program MOLREP (Vagin & Teplyakov, 1997) and the MAD MhpC structure (Dunn, 1999) as phasing model, and the resulting electron density maps were refined using programs of the CNS package. Model-building was done using QUANTA, and the iterative process repeated using the same software packages.

2.3. Biochemical and Molecular Biological methods

Spectroscopic Analysis: Ultraviolet/visible spectra were recorded on a Pharmacia Biotech Ultrospec 3000 spectrophotometer using cells with a path length of 10mm.

Growth Media: Luria broth (LB) media was prepared by autoclaving at 120°C for 30 minutes a solution of tryptone (10g), yeast extract (5g), and sodium chloride (10g) in water (1L). LB Agar for culture plates was prepared by the addition of agar (15g) to LB media (1L). All apparatus used to manipulate bacterial cultures was also autoclaved at 120°C for 30 minutes.

Enzyme Kinetics: MhpC was assayed by monitoring the decrease in absorbance at 394nm, as previously described (Henderson & Bugg, 1997), upon addition of enzyme to a solution of RFP in 50mM potassium phosphate buffer (pH8.0). One unit was defined as the amount of enzyme that converted 1 μmol of substrate per min at 22°C. K_m and V_{max} were calculated from Lineweaver-Burk plots (Lineweaver & Burk, 1934).

Protein Concentration Assay: The method of Bradford was followed (Bradford, 1976). The Bradford protein assay uses the Coomassie Blue reagent, made up as follows: to a solution of Coomassie Blue (G250, 10mg) in 95% ethanol (5ml) was added 85% *o*-phosphoric acid (10ml); the resulting mixture was diluted to 100ml with water. The prepared reagent was stable to prolonged storage at 4°C. A protein assay was carried out by mixing 1ml of reagent with 5-50μl of protein solution. The solution was allowed to stand at room temperature for 5 minutes and the absorbance at 595nm was recorded. A standard curve was constructed using bovine serum albumin (BSA; 0-2mg/ml in water, stored at 4°C).

SDS-PAGE Analysis:

Stacking gel:	5% acrylamide-bis (29:1), 0.2% SDS, 250mM Tris-HCl pH6.8
Resolving gel:	12.5% acrylamide-bis (29:1), 0.1% SDS, 375mM Tris-HCl pH8.8
Running buffer:	25mM Tris-base, 250mM glycine, 0.1% SDS
Gel-loading buffer:	100mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 50mM Tris-HCl pH6.8

SDS-PAGE was performed on a BioRad Protean II gel apparatus using 16x16cm gels of 1mm thickness and a resolving gel concentration of 12.5% acrylamide. Samples were made up to 15µl with gel loading buffer and incubated at 100°C for 5 minutes prior to application to gel wells. Broad-range molecular weight markers (New England Biolabs 77015) were applied to the outer wells. The gel was run at an initial voltage of 8 V/cm, increasing to 15 V/cm once the dye-front had moved into the resolving gel. Proteins were visualised by staining with Coomassie Brilliant Blue staining reagent (0.1% w/v Coomassie Blue R250, 40% methanol, 10% acetic acid and 50% water). After staining, the gels were destained by repeated washing with destain reagent (10% methanol, 10% acetic acid and 80% water).

2.3.1. Preparation of MhpC

The *mhpC* gene, originally expressed in *E. coli* DH5α, had previously been cloned into a high-expression plasmid, pIPC, under control of the *lac-lpp* promoter and with ampicillin-dependent selection (Figure 24). The plasmid was constructed and kindly supplied by Dr. J.-L. Garcia of the Centro de Investigaciones Biologicas, Madrid, Spain.

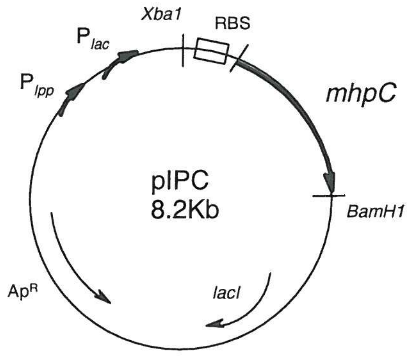


Figure 24: A plasmid map of pIPC containing the gene for MhpC.

Restriction enzyme digest of pIPC: To 1µl of pIPC (900ng/µl) was added in separate tubes 1µl *BamH1,* 1µl *XbaI,* and 1µl of both *BamH1* and *XbaI.* To each solution was also added 1µl MULTI-CORE buffer (Promega) and distilled water to a final volume of 10µl. The restriction enzymes were added last and the solutions incubated at 37°C for 2 hours. The reactions were then analysed by agarose gel electrophoresis on a 1% agarose gel (2ml 50xTAE, 1g agarose in 98ml water, with 3μl ethidium bromide) and visualised by the addition of 1μl of Orange G (10x) to each sample. The 10μl samples were loaded into wells of the gel alongside Gibco BRL 100bp and 1kb DNA ladders (5μl each). The BioRad Mini-Sub Cell GT DNA Electrophoresis apparatus was run at 100mA constant current for 30 minutes and the gel analysed under UV light and photographed.

Transformation of pIPC: The plasmid was transformed into *E. coli* B834(DE3)pLysS competent cells, made using the calcium chloride protocol (Ausubel *et al.*, 1995). A 20ml overnight culture of strain B834(DE3)pLysS was grown in LB containing no antibiotic. 0.2ml of this culture was used to inoculate 20ml of LB media and the cells incubated at 37°C until an OD_{600} of 0.38 was reached. The cells were centrifuged at 1800x*g* for 7 minutes at 4°C and the supernatant discarded. The cells were resuspended in 10ml of ice-cold transformation buffer (50mM calcium chloride and 10mM Tris pH 7.5) and incubated for 30 minutes at 4°C. The supernatant was again discarded following centrifugation at 1100 x*g* for 5 minutes at 4°C. The cells were then resuspended in a further 2ml of ice-cold transformation buffer to give the competent cell solution. To 100μl of the competent cells in pre-chilled tubes was added 10-20ng of plasmid pIPC and the solution incubated for 10 minutes at 4°C. Following this the cells were heat-shocked at 42°C for 2 minutes and incubated for a further 10 minutes at 4°C. Then 1ml of LB medium was added and the cells incubated at 37°C for 30 minutes with shaking. The cells were centrifuged at 13000rpm for 1 minute and resuspended in 200μl LB containing ampicillin (100μg/ml; LB-Amp). The cells were applied to LB-Agar plates containing ampicillin (100μg/ml) and incubated overnight at 37°C. Colonies indicated that the transformation had been successful, as pIPC had conferred ampicillin resistance to the host strain.

Protein expression: To check the over-expression system of the plasmid a 50ml culture of the transformed cells in LB(Amp) was grown overnight at 37°C. MhpC expression was induced by the addition of 1mM isopropyl-*β*-D-thiogalactopyranoside (IPTG) when the cell optical density reached 0.6 at 600nm. The expression of MhpC was examined by the use of a time-course SDS-PAGE experiment, where the cells were left to grow at 37°C and 1ml aliquots taken at hourly intervals.

A large-scale preparation of MhpC was carried out by the fermentation of a 10L culture of B834(DE3)pLysSpIPC cells in LB media containing ampicillin (100μg/ml). The protein expression was induced by the addition of 1mM IPTG at an OD_{600} of 0.6, and the cells incubated overnight at 37°C prior to harvesting.

2.3.2. Purification of MhpC

The cells were harvested by centrifugation in a Sorvall RC-3B centrifuge at 5000rpm for 30 minutes at 4°C. After harvesting, the cells were resuspended in 250ml degassed lysis buffer (50mM potassium phosphate pH7, 0.2mM EDTA, and 10mM β-mercaptoethanol (β-ME)) and lysed by sonication (20 seconds on, 60 seconds off, 12 cycles). The sonicated cells were centrifuged in a Beckman J2 centrifuge at 14000rpm for 30 minutes at 4°C to separate the cell debris. The supernatant thus obtained being the crude extract (250ml).

MhpC was then purified using an already established protocol (Robertson, 1999). Ammonium sulphate precipitation: Powdered ammonium sulphate (36g) was added to the crude extract (250ml) to 25% saturation. The suspension was stirred at 4°C for 30 minutes and precipitated protein separated from the soluble fraction by centrifugation at 14000rpm for 30 minutes at 4°C.

Dialysis: The precipitated protein was resuspended in degassed buffer (50mM potassium phosphate pH7, 0.2mM EDTA, and 10mM β-ME) and dialysis carried out overnight at 4°C against 5L of degassed anion exchange buffer (50mM potassium phosphate pH7.0, 0.2mM EDTA, 10mM β-ME). The dialysis was repeated again under the same conditions and the sample centrifuged at 14000rpm for 30 minutes at 4°C to remove any precipitated protein.

Ion-exchange chromatography: The protein was further purified by elution through a pre-washed, pre-equilibrated Q-Sepharose anion exchange column (bed volume 120ml). The protein was eluted using anion exchange buffer with a sodium chloride gradient of 0 - 0.5M (250ml no salt buffer and 250ml 0.5M NaCl buffer). The flow rate was 5ml/min and fractions of 6ml were collected.

Gel filtration chromatography: The partially purified protein was further purified using a pre-equilibrated Superdex G-75 gel filtration column (column volume 120ml) on a Pharmacia Biotech LCC-501 Plus FPLC system with a 10ml loop. The protein was eluted with elution buffer (100mM Tris pH7.5, 0.2mM EDTA) with a flow rate of 0.5ml/min and 2ml fractions were collected.

The purification procedure was monitored after each step by the analysis of key fractions using SDS-PAGE.

Protein concentration: Purified MhpC was concentrated to 25mg/ml using Amicon and Centricon apparatus. The protein solution was also buffer exchanged from 100mM Tris (pH7.5) to 10mM Tris (pH7.5) whilst carrying out the Centricon concentration in a Sigma laborzentrifugen 3K15 benchtop centrifuge.

MhpC concentration was calculated using the Beer-Lambert relationship. A 1mg/ml solution of MhpC has an absorbance of 0.74 at 280nm, when measured with a path-length of 1cm.

2.3.3. Crystallisation of MhpC

Crystallisation of MhpC was achieved using the hanging drop vapour diffusion method. Two sets of crystallisation conditions for native MhpC crystals had previously been found (Dunn, 1998; Mohammed, 1999). These were:

(1) 0.1M sodium cacodylate (pH6.5), 0.2M calcium acetate, 0.2M ammonium chloride, 44% 2-methyl-2,4-pentanediol (MPD), at room temperature;

(2) 50mM calcium chloride, 7% PEG 2000, 10mM DTT, 25% ethylene glycol, 0.1M HEPES (pH7.7), at 4°C. Addition of a 1:1 mixture of Dow Corning 200/1cS fluid and liquid paraffin gave slower growth and larger crystals.

2.3.3.1. Soaking crystals with ligands

Native crystals of MhpC, grown using either of the two methods mentioned previously, were placed in a 4μl drop of mother liquor containing 1mM phenylmethylsulphonylfluoride (PMSF), an inhibitor of serine proteases. Crystals were left to soak for 40 minutes before mounting and X-ray data collected.

Crystal soaking experiments were also carried out with laevulinic acid, 2,6-diketo-nona-1,9-dioic acid (DKNDA), the methyl ester of the RFP (MeRFP), and the ethyl ester of HPD (EtHPD). For all ligands their final concentrations were 10mM and soaking was carried out for 24 hours prior to crystal freezing and data collection.

2.3.3.2. Co-crystallisation with ligands

An aliquot of 200mM PMSF in methanol was added to MhpC (25mg/ml) to give a 5% v/v solution (10mM final PMSF concentration). The second set of crystallisation conditions were then used with varying PEG 2000 concentrations (7-9%), and drops consisted of 2μl MhpC/PMSF solution:2μl mother liquor.

MhpC was also co-crystallised with 12.5, 25, 50 and 100mM succinate, by the addition of varying quantities of succinate (200mM solution in 10mM Tris pH7.0) to 20μl aliquots of MhpC (25mg/ml). The first set of crystallisation conditions were used with hanging drops of 2μl MhpC/succinate solution:2μl mother liquor.

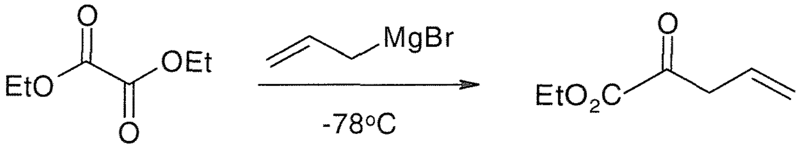
Co-crystallisation of MhpC (25mg/ml) with laevulinic acid pH7.0 (10mM final concentration) was undertaken by incubating together for 6 hours prior to setting up crystallisation trials. Both sets of crystallisation conditions were used with drops consisting of 2μl MhpC/ligand:2μl mother liquor. Further co-crystallisation experiments were undertaken with laevulinic acid (10mM) after incubation with MhpC for 2 hours. This time the precipitant concentrations of the two sets of crystallisation conditions were varied, 40.5-46% MPD for the first conditions and 9-14.5% PEG 2000 for the second set of conditions. Drops were 2.5μl MhpC/ligand:2.5μl mother liquor, and the crystallisation trays were incubated at 4°C. Co-crystallisation was also attempted using 10mM laevulinic acid with 12.5mg/ml MhpC for the first set of conditions, with varying concentrations of MPD (40-47%). Trays were incubated at room temperature. Finally, co-crystallisation of MhpC (25mg/ml) with laevulinic acid (10mM) was also carried out using Molecular Dimensions crystal screening kits MD1-01 and MD1-02.

Co-crystallisation of MhpC (25mg/ml) with azelaic acid, DKNDA, 4-oxosebacic acid (OSA), 4,7-dioxosebacic acid (DOSA), EtHPD, succinic acid, 4,6-dioxoheptanoic acid (DOHA), *trans,trans*-2,4-hexadienoic acid (2,4-HDA), *trans*-3-nonene-2-one (t3N2O) and MeRFP (all 10mM final concentrations) was also carried out separately using Molecular Dimensions crystal screening kits MD1-01 and MD1-02. EtHPD and succinic acid were also co-crystallised jointly with MhpC using the screening kits, with final concentrations of 20mM of each. DKNDA, DOHA, OSA, DOSA and succinic semialdehyde were also co-crystallised with MhpC at final concentrations of 20mM. Another serine protease inhibitor, 4-(2-aminoethyl)benzenesulphonyl fluoride (AEBSF), less toxic and more water soluble than PMSF, was co-crystallised with MhpC (25mg/ml) at a final concentration of 10mM using Molecular Dimensions crystal screening kits. With all co-crystallisation trials done using Molecular Dimensions solutions drop sizes were 2μl MhpC/ligand:2μl mother liquor.

Once initial hits had been obtained using the Molecular Dimensions screening solutions, then further screening was carried out based around these conditions to optimise the crystal quality and additives were used, such as acetone, isopropanol or dioxane, to reduce the likelihood of crystal twinning.

DKNDA and MeRFP were synthesised and kindly donated by Dr. Gang Lin (University of Warwick, UK). OSA and DOSA were synthesised and kindly donated by Reinhard Neier (Uni Neuchatel, Switzerland). EtHPD was synthesised as in section 2.3.4. All other ligands used in co-crystallisation trials with MhpC were commercially available.

2.3.4. Synthesis of Ethyl 2-keto-penta-4-enoate



The ethyl ester of 2-hydroxy-penta-2,4-dienoic acid (HPD), one of the products of the MhpC catalysed reaction, was synthesised so that it could be used in soaking and co-crystallisation trials with MhpC.

The compound was prepared by the method of Rambaud *et. al.* (1988). In a three-necked flask fitted with a dropping funnel was placed anhydrous tetrahydrofuran (40ml), anhydrous diethyl ether (40ml) and diethyl oxalate (1ml, 20mmole) under a nitrogen atmosphere. The reaction was cooled to -78°C and allyl magnesium bromide (20ml of a 1M solution in diethyl ether) was added dropwise over 30 minutes. The temperature was maintained at -78°C during the addition and for a further 30 minutes. After quenching with 1M H_2SO_4 (40ml) and extraction into diethyl ether (2 x 35ml), the product was purified by flash chromatography and concentrated *in vacuo* to give a yellow oil (0.98 g, 8.7mmole, 44%). Positive electrospray mass spectrometry showed the expected product peak [m/z (ES+) 143 $[M+H]^+$].

2.3.5. Sequencing of the *mhpC* gene

2.3.5.1. Plasmid purification

All steps in the purification of pIPC were carried out using the QIAGEN® Plasmid Maxi Kit. A 500ml culture of B834(DE3)pLysSpIPC in LB(Amp) was grown at 37°C for 14 hours with vigorous shaking and the cells harvested by centrifugation at 5000rpm for 30 minutes. The bacterial pellet was resuspended in 10ml Resuspension buffer (50mM Tris-HCl pH8.0, 10mM EDTA, 100μg/ml RNase A) by vortexing. Next 10ml of Lysis Buffer (200mM NaOH, 1% SDS) was added and mixed by inverting, then left to incubate at room temperature for 5 minutes. The lysate appeared viscous. To this was added 10ml of chilled Neutralisation Buffer (3.0M potassium acetate pH5.5), which was mixed immediately by inversion and incubated at 4°C for 20 minutes. The lysate became less viscous and a white precipitate containing genomic DNA, proteins and cell debris was formed. The sample was centrifuged at 20000rpm for 20 minutes at 4°C and the supernatant, containing plasmid DNA, removed and filtered through a 0.2μM filter. A QIAGEN-tip 500 column was equilibrated by applying 10ml Equilibration Buffer (750mM NaCl, 50mM MOPS pH7.0, 15% isopropanol, 0.15% Triton X-100) and allowing the column to empty by gravity flow. The supernatant containing plasmid DNA was applied to the column and allowed to enter the resin by gravity flow. The column was then washed with 2x 30ml Wash Buffer (1.0M NaCl, 50mM MOPS pH7.0, 15% isopropanol) and the DNA eluted with 15ml Elution Buffer (1.25M NaCl, 50mM Tris-HCl pH8.5, 15% isopropanol) and stored at 4°C. The DNA was precipitated by the addition of 0.7 volumes room temperature isopropanol to the eluted DNA. After mixing and centrifugation at 20000rpm for 30 minutes at 4°C the supernatant was carefully decanted. The DNA pellet was then washed with 5ml room temperature ethanol (70%) and centrifuged at 20000rpm for 20 minutes. The supernatant was carefully decanted, the pellet air-dried and resuspended in 10mM Tris pH7.5, 2mM EDTA (1ml). The purity of the plasmid DNA was checked by measuring the A_{260}. The plasmid DNA was purified further by adding phenol to the plasmid solution (1:1), vortexing and centifuging at 13000rpm for 3 minutes. The top layer was removed and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to this layer and vortexed. The mixture was centrifuged at 13000rpm for 3 minutes and the step repeated. To the top layer was added an equal volume of chloroform and the mixture vortexed once more. Following centrifugation at 13000rpm for 3 minutes the top layer was again removed and this extra-pure plasmid DNA solution stored at 4°C. The purity of the plasmid DNA was checked by measuring the A_{260} or by using the Invitrogen DNA Dipstick Kit.

2.3.5.2. Primer design

The primers used for sequencing were F71027 (5'-CAGCCACAAACCGAAGCC-3') for sequencing of the DNA encompassing the triplets encoding S110, D235 and H263, F73338 (5'-CCATGCAGCAGGACAACG-3') for sequencing the N-terminus, and F71648 (5'-GCGCGCCTGAATAATATGC-3') for sequencing the C-terminus. The sequencing reaction products were analysed with an Applied Biosystems Inc. automatic sequencer by Oswel DNA Service, University of Southampton, UK.

2.3.5.3. Analyses of amino acid sequences

Sequences were obtained from the SWISS-PROT/TrEMBL protein sequence database (Bairoch & Apweiler, 2000). Multiple amino acid sequence alignments were performed with the MALIGN program (Johnson, 1990) and colour presentation was generated by CINEMA version 2.1 (Attwood *et. al.*, 1997).

2.3.6. Tryptic digest of MhpC and analysis by Mass Spectrometry

To 2mg of MhpC (80μl of a 25mg/ml solution) was added 1/100th the amount of trypsin (20μl of a 1mg/ml solution). This was incubated at 37°C for 24 hours in 400μl 20mM Tris.HCl pH8.0 containing 2mM calcium chloride. An equal volume of 0.2% TFA in water was then added.

The MhpC tryptic digest analysis was carried out on a Varian Prostar HPLC system using a Varian C18 Reverse Phase Column (150 x 4.6mm). Elution solutions were 0.1% TFA in water (A) and MeCN containing 0.1% TFA (B) with an initial gradient of 1%/min going from 100% A to 90% B. 200μl samples of digest solution were injected on to the column. The gradient was changed to run from 100% A to

40% B over 90 minutes following preliminary inspection of peak data. Peak data were obtained at 280nm and 214nm. Samples were collected manually and analysed on a Fisons Instruments VG Quattro II triple quadropole electrospray mass spectrometer. Running solvent for mass spectrometry was 50:50 H_2O:CH_3CN. The capillary voltage was 4V, HV lens voltage 0.18V and cone voltage between 20 and 50V. Formic acid (0.1% v/v) was added to 10µl samples prior to injection into the mass spectrometer. Mass spectrometry data was analysed using MassLynx version 2.1 (Fisons Instruments plc) and predicted peptide fragments generated *in silico* using Biolynx (Micromass UK Ltd.).

On-line LC-MS (Liquid Chromatography Mass Spectrometry) of a tryptic digest of MhpC was carried out using the same HPLC conditions as described previously with a Varian C18 Reverse Phase HPLC column (150 x 2.1mm) mounted between the mass spectrometer injector and probe. The flow rate was 300µl/min and 200µl samples of digest were injected directly onto the column. Tryptic digestion was carried out as described previously. Following separation by HPLC, the flow-through was split in a 10:1 ratio, with the smaller portion going to the mass spectrometer probe. Mass spectrometry data was analysed as described previously.

2.3.7. ^{31}P NMR Spectroscopy of MhpC

Samples for NMR were a 1mM solution of MhpC (in water) and a 1mM KH_2PO_4 reference solution. ^{31}P-spectra were recorded on a Bruker DPX400 spectrometer (corresponding to a ^{31}P Larmor frequency of 162MHz) equipped with a 10mm broadband probehead. Experiments were run at 300K with a spectral width of 16kHz, 64k time domain points and a relaxation delay of 2.0s. Prior to Fourier transformation, exponential line broadening of 5Hz was applied to all datasets to reduce spectral noise. Chemical shifts were measured downfield from external 85% H_3PO_4.

CHAPTER 3

PREPARATION, PURIFICATION AND CRYSTALLISATION OF MhpC

3.1. Preparation and purification of MhpC

3.1.1. Preparation of MhpC from plasmid construct pIPC

The construct pIPC, a modified pINIII expression vector (Masui *et. al.*, 1983), containing the gene for MhpC, had previously been produced by Dr. J.-L. Garcia, of the Consejo Superior de Investigaciones Cientificos (Madrid, Spain). This plasmid contains the gene *mhpC* under the control of the *lac-lpp* promoter and is IPTG inducible, with ampicillin dependent selection.

The plasmid was first checked to see if it contained a gene fragment of the correct size for *mhpC* by the use of a restriction enzyme digest. There are two restriction sites flanking the gene encoding MhpC, *Xba*I and *BamH*I (see Figure 24 in Chapter 2.3.1). Therefore, restriction enzymes were used to cut at these sites and the fragment excised examined by agarose gel electrophoresis. This showed a band of approximately 900 base pairs in size, corresponding to a gene which would encode for a protein of about 300 amino acid residues. Hence, it was presumed that this fragment contained the gene for MhpC (Figure 25).

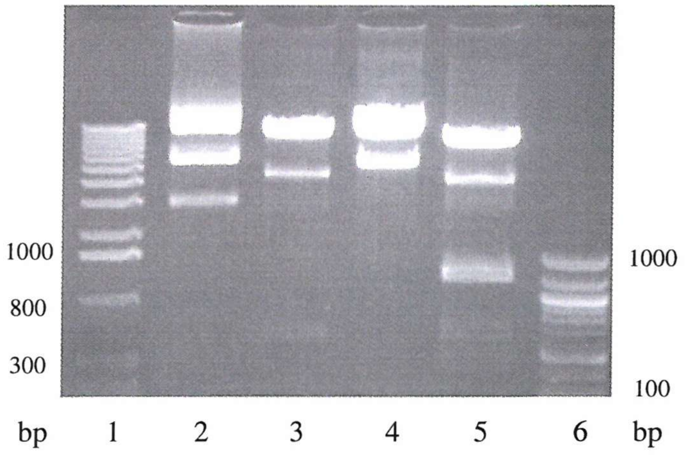


Figure 25: Agarose gel electrophoresis of a restriction enzyme digest of pIPC. Lane 1: 1kb DNA ladder; Lane 2: uncut pIPC; Lane 3: pIPC cut with *BamH*1; Lane 4: pIPC cut with *Xba*I; Lane 5: pIPC cut with *BamH*1 and *Xba*I; Lane 6: 100bp DNA ladder.

The plasmid was then transformed into *E. coli* B834(DE3)pLysS cells and the over-expression system was checked by the use of a time-course SDS-PAGE experiment. After induction with IPTG, aliquots of cells were taken at hourly intervals and analysed. The resulting gel showed that the over-expression system was working, as there was a massive band present at 29kDa corresponding to the size of an MhpC monomer, which got progressively bigger with time (Figure 26).

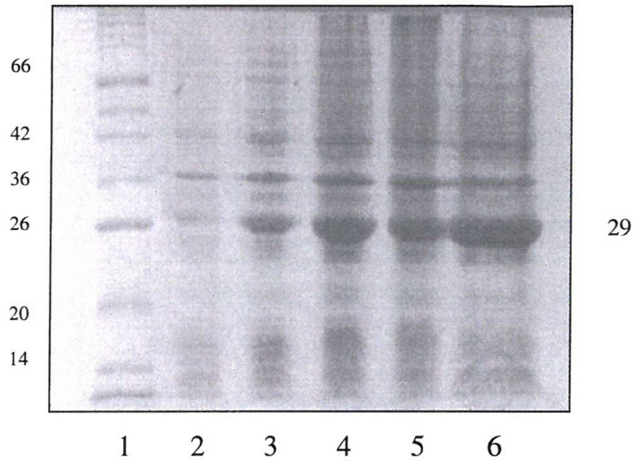


Figure 26: Time-course SDS-PAGE analysis of MhpC over-expression. Lane 1: Molecular weight markers (kDa); Lane 2: pre-induction; Lane 3: 1 hour post-induction; Lane 4: 2 hours post-induction; Lane 5: 3 hours post-induction; Lane 6: 4 hours post-induction.

3.1.2. Purification of MhpC to homogeneity

MhpC was produced on a larger scale by fermentation of a 10L culture of *E. coli* B834(DE3)pLysSpIPC cells. The over-expression system was so successful that it was common to obtain an MhpC yield of 100mg/L. The protein was purified from the crude extract by ammonium sulphate precipitation, anion-exchange chromatography and gel filtration chromatography following an already established protocol (Robertson, 1999). The success of the purification procedure was checked after each step by the use of SDS-PAGE (Figures 27 and 28). Following purification the enzyme was deemed homogenous and therefore suitable for crystallisation. The enzyme was concentrated to 25mg/ml in 10mM Tris.HCl (pH7.5) prior to crystallisation trials.

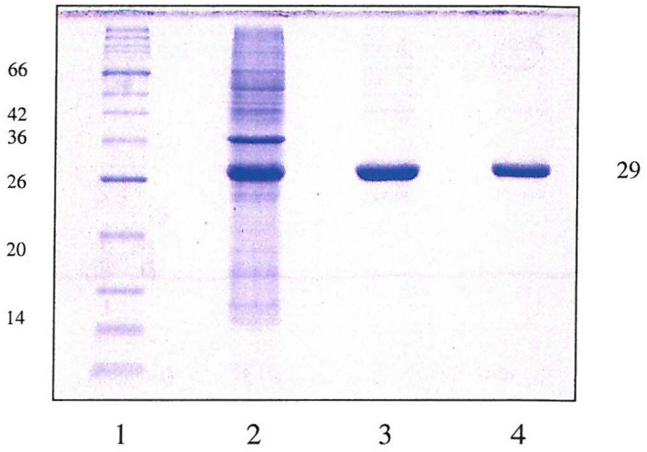


Figure 27: SDS-PAGE analysis of MhpC purification after ammonium sulphate precipitation, dialysis, and anion exchange chromatography. Lane 1: Molecular weight markers (kDa); Lane 2: Crude extract showing massive over-expression of MhpC at 29kDa; Lanes 3 and 4: MhpC after ammonium sulphate precipitation, dialysis and anion exchange chromatography (differing dilutions).

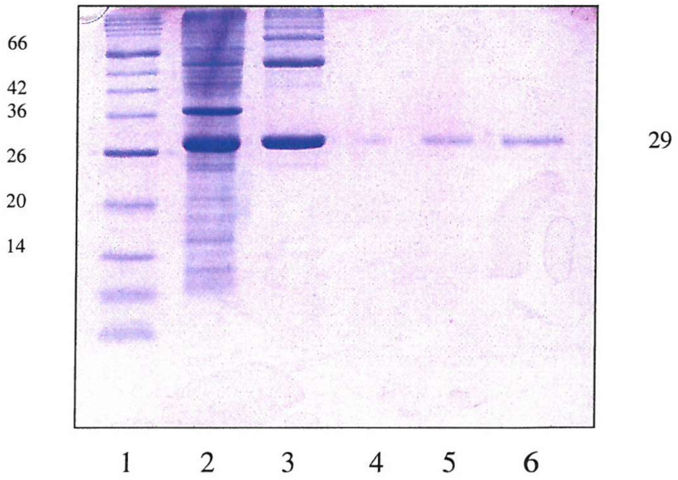


Figure 28: SDS-PAGE analysis of MhpC purification including gel filtration chromatography. Lane 1: Molecular weight markers (kDa); Lane 2: Crude extract; Lane 3: Anion exchange chromatography; Lanes 4,5,6: Gel filtration chromatography (different dilutions).

3.1.3. Determination of K_m for MhpC

A number of methods have been adopted for kinetic analysis of MhpC. These include monitoring the decrease in absorbance at 394nm due to the consumption of the RFP or the use of a succinate release assay (Lam & Bugg, 1997). Alternatively, when stopped-flow kinetic experiments were conducted then MhpC was assayed by monitoring the disappearance of substrate at 317nm or appearance of product at 270nm (Henderson & Bugg, 1997).

In this instance MhpC was assayed by the first approach in potassium phosphate buffer (pH 8.0) and with varying concentrations of RFP. The decrease in absorbance at 394nm was recorded and activity calculated using $\varepsilon = 15600\ M^{-1}\ cm^{-1}$. It was found that Michaelis-Menten kinetics were followed and a K_m of 2.4 μM was measured, which compares favourably with the literature value of 2.1 μM (Lam & Bugg, 1997).

3.2. Crystallisation of MhpC

Crystallisation conditions for MhpC had previously been determined (Dunn, 1998; Mohammed, 1999). One set of conditions contained 10% polyethylene glycol (PEG-2000) as precipitant, with 0.1M HEPES buffer (pH7.7), 25% ethylene glycol and 50mM calcium chloride, and produced crystals after incubation at 4°C. The other crystallisation conditions had 44% 2-methyl-2,4-pentanediol (MPD) as precipitant, with 0.1M sodium cacodylate (pH6.5), 0.2M ammonium chloride and 0.2M calcium acetate, and produced crystals after 72 hours when grown at room temperature (Figure 29). Native crystals were obtained by the hanging drop vapour diffusion method using both sets of conditions.

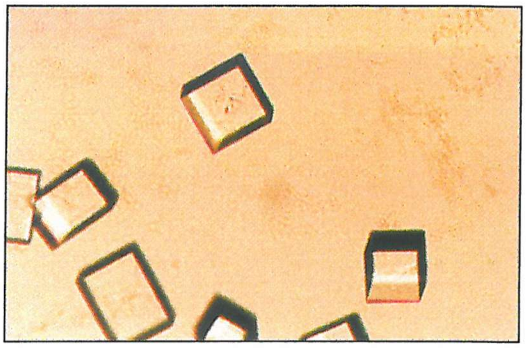


Figure 29: Large crystals of MhpC grown using the hanging drop vapour diffusion method at room temperature in 44% MPD, 0.1M sodium cacodylate, 0.2M calcium acetate and 0.2M ammonium chloride.

3.2.1. Soaking ligands into crystals

As serine proteases are known to be inhibited by certain sulphonyl halide compounds (Gold & Fahrney, 1964), and structures exist of not only chymotrypsin with the active site serine covalently modified by PMSF (Botos *et. al.*, 2000), but also α/β-hydrolases such as the carboxylesterase from *P. fluorescens* (Kim *et. al.*, 1997) and dienelactone hydrolase (Robinson *et. al.*, 2000), it was attempted to undertake similar experiments with MhpC. However, attempts to soak PMSF into native MhpC crystals were unsuccessful as most crystals were destroyed when this was done, independent of the length of soaking time.

To further investigate the catalytic machinery of MhpC, ligands that mimic certain functionalities of the natural substrate, 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid (RFP), were obtained or synthesised. Laevulinic acid, succinic acid, azelaic acid, 2,6-diketo-nona-1,9-dioic acid (DKNDA) and 2-hydroxy-penta-2,4-dienoic acid ethyl ester (EtHPD) were all used in crystal soaking experiments with native MhpC (Figure 30). In all cases the crystals were soaked for a period of 24 hours in mother liquor containing ligand (10mM final concentration) prior to crystal freezing and data collection.

Attempts to soak these ligands into MhpC crystals were also unsuccessful despite, in some cases, the crystals surviving and data collection being possible. This was due to the lack of any difference density being evident, for a bound ligand, within the active site cavity, after data collection and structure determination. Therefore, it was decided that due to the difficulties in soaking ligands into MhpC another approach was required.

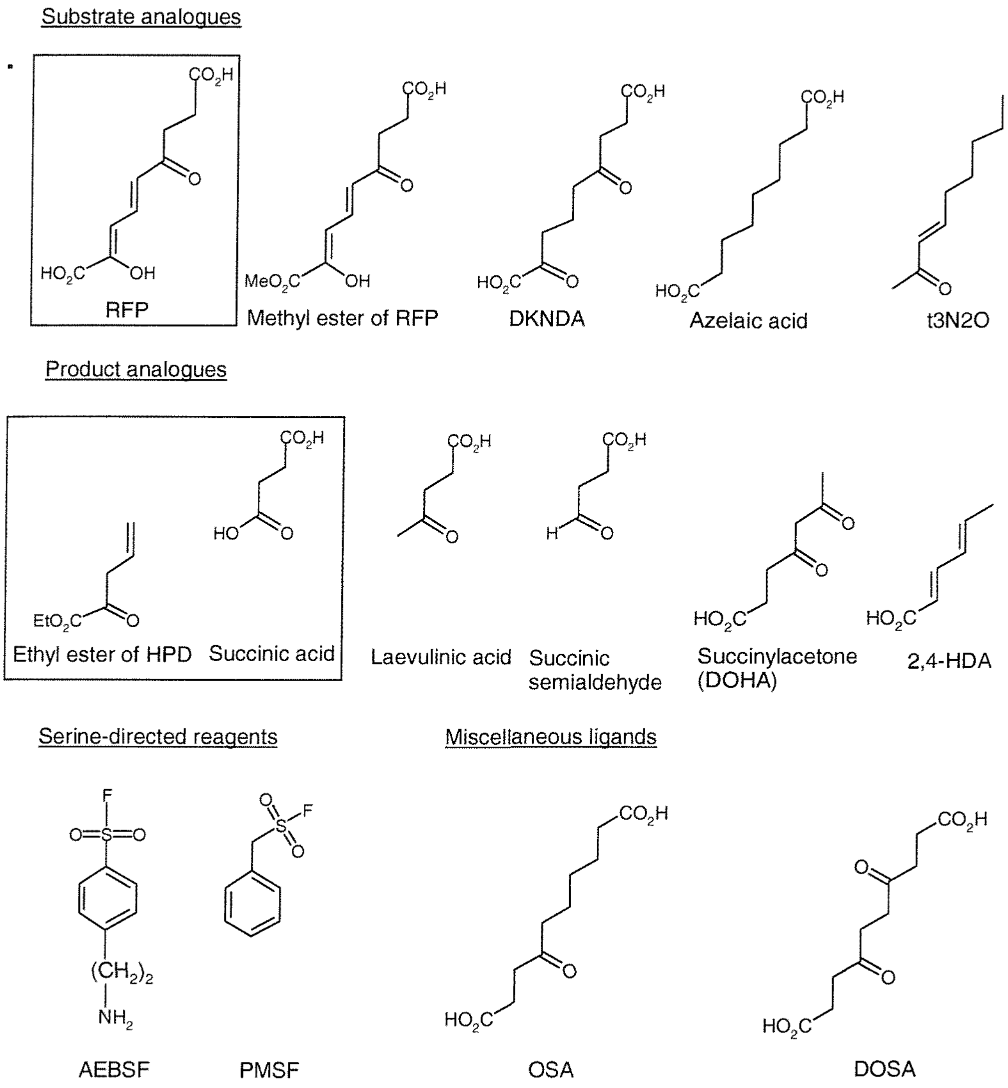


Figure 30: Compounds used in crystal soaking and co-crystallisation experiments with MhpC that mimic functionalities of the natural substrate (RFP) or products of the MhpC catalysed reaction (succinic acid and HPD). Also shown are the structures of the two serine-directed reagents used in studies with homologous enzymes.

3.2.2. Co-crystallisation of MhpC with substrate and product analogues

Due to the difficulty of soaking ligands into native crystals of MhpC, the technique of co-crystallisation was used instead. In this case, the protein and ligand(s) (Figure 30) were incubated together prior to crystallisation. However, the previously established crystallisation conditions for native MhpC would rarely produce crystals when a ligand was added to the protein, so crystal screening kits were used (Molecular Dimensions Inc.) to obtain new crystallisation conditions. Each co-crystal was grown by the hanging drop vapour diffusion method at either room temperature or 4°C, with varying crystallisation conditions required for most compounds (final concentration 10mM, except where stated) when incubated with MhpC (25mg/ml, except where stated). Drop sizes consisted of 2μl protein/ligand mixed with 2μl mother liquor.

3.2.2.1. Co-crystallisation of MhpC with serine-directed reagents

Co-crystallisation of MhpC with PMSF was carried out using the protocol established for the co-crystallisation of human chymase, a serine protease, with PMSF (McGrath *et. al.*, 1997). Unfortunately, co-crystallisation of MhpC and PMSF was never successfully conducted using the original crystallisation conditions, as the protein would normally precipitate once PMSF was added.

Therefore, AEBSF, a water-soluble and less toxic analogue of PMSF, was co-crystallised with MhpC using the Molecular Dimensions kits. The best crystals were obtained with Kit MD-01 Solution 47: 30% w/v PEG-1500. This produced small, plate-like crystals, which only diffracted to ~5Å resolution and so were not suitable for full data collection.

3.2.2.2. Co-crystallisation of MhpC with substrate analogues

The co-crystallisation of MhpC with azelaic acid was undertaken using a commercially available screen and initial hits were obtained under the following conditions:

Kit MD-02 Solution 3: 2% w/v dioxane, 0.1M bicine (pH9.0) and 10% w/v PEG-20000;

Kit MD-02 Solution 8: 0.2M $(NH_4)_2PO_4$ (monobasic), 0.1M Tris (pH8.5) and 50% v/v MPD;
Kit MD-02 Solution 18: 0.1M Hepes (pH7.5) and 4.3M NaCl.
A crystal was taken directly from the screen condition MD-02 Sol 3 and gave good diffraction without the need for further refinement of the crystallisation conditions. Therefore, a full data set was collected with this crystal.

The methyl ester of the RFP (MeRFP) was also co-crystallised with MhpC using the same kits and initial hits obtained with:
Kit MD-01 Solution 23: 0.2M $CaCl_2.2H_2O$, 0.1M NaHepes (pH7.5) and 28% w/v PEG-400;
Kit MD-02 Solution 8: as above;
Kit MD-02 Solution 22: 0.1M Mes (pH6.5) and 12% w/v PEG-20000;
Kit MD-02 Solution 42: 25% v/v ethylene glycol; and
Kit MD-02 Solution 43: 35% dioxane.
Grid screens, of varying pH and precipitant concentration, were done with solutions based on MD01 Sol 23 and MD02 Sol 42 as these gave the best quality crystals. Eventually, small crystals ($0.1mm^3$) were obtained that diffracted to a reasonable resolution at ESRF, Grenoble, and so a full data set was collected.

The co-crystallisation of MhpC and DKNDA was also successfully undertaken using Molecular Dimensions screening kits. Preliminary hits were obtained with the following solutions:
Kit MD-01 Solution 6: 0.1M Na acetate trihydrate (pH4.6) and 8% w/v PEG-4000;
Kit MD-01 Solution 9: 0.1M trisodium citrate dihydrate (pH5.6), 20% w/v 2-propanol and 20% w/v PEG-4000;
Kit MD-02 Solution 8: as above; and
Kit MD-02 Solution 18: as above.
Screening was also carried out with DKNDA having a final concentration of 20mM. The best crystals were obtained with MD-02 Sol 8, which is a particularly useful condition as it enables the crystal to be flash-frozen without the need for a cryoprotectant. Attempts to replicate these conditions were unsuccessful and so a larger volume of the commercially available solution was obtained and serial dilutions made in attempts to both reduce the number of crystals per drop and

increase their relative size. The addition of 0.5% dioxane, 0.5% acetone and 0.5% isopropanol was also investigated to see if they would reduce crystal twinning. Crystals large enough for both 'in-house' and synchrotron data collection were obtained.

The compound *trans*-3-nonen-2-one (t3N2O), which is similar in some respects to the keto-intermediate of the RFP but lacking the carboxylate groups, has also been co-crystallised with MhpC. The best crystals were obtained by dilution of MD-02 Sol 8 with water (such that MPD was between 45 and 46%), and also containing 0.5% acetone, which has been found to reduce the twinning problem associated with this enzyme and crystallisation condition. The crystals were quite small (<0.1mm) and unfortunately the majority only diffracted to ~6Å 'in-house'. One crystal gave slightly improved diffraction, to 3.2Å, but has not yet been utilised for collection of a full data set.

Co-crystallisation of MhpC with OSA and DOSA (20mM final concentrations) was conducted using screens as with other ligands. No crystals were obtained with OSA as ligand, but DOSA gave small blocky crystals, which had an unusual yellow colour, with MD-02 Sol 8. It was noted that on addition of MhpC/DOSA to drops of a mother liquor that contained ammonium salts, after a period of incubation (normally up to two weeks), the drops would turn a yellow colour, the colouration being deeper the more concentrated the ammonium ion. This was also found without MhpC being present, and so is proposed to be a reaction of the colourless ligand with ammonium ions to form a yellow compound. In any case, currently the crystals of MhpC with this 4,7-diketodeca-1,10-dioic acid are too small for data collection even after prolonged incubation and so further grid and factorial screening is required.

3.2.2.3. Co-crystallisation of MhpC with product analogues

The co-crystallisation of MhpC with laevulinic acid (pH 7.0) was attempted initially by varying the already established crystallisation conditions by the use of grid screening. However, this remained unsuccessful despite varying quite widely the two sets of conditions. In particular, by either decreasing the MPD or protein concentration (12.5 mg/ml) to prevent precipitation for one set of conditions or by increasing the PEG-2000 concentration to obtain crystals for the other set of conditions. The only way to obtain crystals was by the use of a Molecular Dimensions Inc. commercial screen containing 100 widely varying conditions. Preliminary hits were obtained with the following conditions:

Kit MD-01 Solution 45: 4.0M sodium formate;

Kit MD-01 Solution 46: 0.05M KH_2PO_4 and 20% w/v PEG-8000; and

Kit MD-01 Solution 47: 30% w/v PEG-1500.

Grid screens were constructed for all of the above conditions. However, the best crystals, thick long needles, were obtained with 4.0M sodium formate, and one of these was used for full data collection on beamline ID14-2 at the ESRF, Grenoble.

The screen of MhpC with DOHA (20mM final concentration) gave hits as follows:

Kit MD-01 Solution 7: 0.2M ammonium acetate, 0.1M trisodium citrate dihydrate (pH5.6) and 30% w/v PEG-4000;

Kit MD-01 Solution 45: as above;

Kit MD-02 Solution 18: as above;

Kit MD-02 Solution 19: 0.1M Hepes (pH7.5), 10% w/v PEG-8000 and 8% v/v ethylene glycol; and

Kit MD-02 Solution 22: as above.

Grid screening was done around the two best hits (MD-01 Sol 7 and MD-02 Sol 18). As good quality crystals were obtained with MD-02 Sol 8 for the MhpC/DKNDA co-crystal then a semi-random factorial screen was also conducted with these conditions for MhpC and DOHA (10mM). It was determined that crystals grew best in 0.1M $(NH_4)_2PO_4$, they were larger at slightly lower concentrations of MPD (48%) and the pH of the Tris did not affect crystal growth (varied from 8.3 to 8.7). The presence of DTT (10mM) did not affect the quality of the crystals or prevent the formation of a 'skin' on the drops. This 'skin' makes it particularly difficult to pick the crystals up prior to freezing. However, the large amount of MPD in the crystallisation conditions means that no cryoprotectant is required for the crystal to survive rapid freezing. Nonetheless, the best crystals were obtained, as in the case of MhpC/DKNDA, by the serial dilution of MD02 Sol 8 to give ~45% MPD, and with the addition of 0.5%

acetone. A full diffraction data set was collected for one of these crystals on beamline 9.2 at Daresbury Laboratory, UK.

A compound that mimics the dienoic acid moiety of the HPD product, 2,4-HDA, was screened as above and gave the following hits:
Kit MD-01 Solution 7: as above;
Kit MD-01 Solution 9: as above;
Kit MD-01 Solution 38: 0.1M Tris (pH8.5) and 8% PEG-8000; and
Kit MD-02 Solution 18: as above.
Grid screens were set up around hits from all the above solutions as these all gave fairly good crystals. Factorial screens were also conducted around MD-01 Sol 38 and MD-02 Sol 8 as these sets of conditions have consistently given the best crystals for MhpC with ligands. The conditions that eventually gave very good quality crystals, after incubation for only ~24 hrs, were 0.1M Tris (pH8), 25% MPD, 0.2M $CaCl_2$ and 6% PEG-8000. Thus, these crystals were taken to the synchrotron for diffraction data to be collected.

Attempts to co-crystallise MhpC with succinate (12.5, 25, 50 and 100mM final concentrations) were carried out using one of the previously determined crystallisation conditions, that which contained MPD, calcium acetate, sodium cacodylate and ammonium chloride in the mother liquor. The Molecular Dimensions screen of MhpC with its natural products, succinic acid and HPD (as its ethyl ester, EtHPD), was carried out with them separately and in unison (20mM of each in all cases). Another compound mimicking the succinyl moiety, succinyl semialdehyde (20mM final concentration), was also screened with MhpC using Molecular Dimensions kits. However, none of these co-crystallisation trials gave crystals of sufficient quality for data collection 'in-house' or at a synchrotron.

3.2.2.4. Summary

All of the co-crystals of MhpC and ligand which grew to a size and quality suitable for full data collection, at the European Synchrotron Radiation Facility (ESRF), Grenoble, at the synchrotron at Daresbury Laboratory, UK, or 'in-house', will be discussed in greater depth in the following chapters.

CHAPTER 4

STRUCTURAL FEATURES OF NATIVE MhpC

4.1. Data collection and processing of native MhpC crystals

Due to the numerous attempts to soak ligands into crystals of MhpC, a large number of diffraction data sets have been collected. However, as it was eventually discovered that soaking was an impractical technique then all they have lead to, ultimately, is many native structures of the protein. Initially, the quality of the crystals varied widely and so therefore did the maximum resolution of the diffraction data collected, mostly in the region of 2.5 to 3.5Å. Further improvements in crystal quality, plus the increased power of synchrotron radiation, improved the resolution obtained. Crystals soaked with DKNDA and laevulinic acid, one from each of the different initial crystal forms, gave data to 2.3Å resolution. Another dataset was obtained to 2.5Å resolution from the failed crystal soak with PMSF. Conversely, the highest resolution structure for native MhpC, so far, has been determined to 2.1 Å resolution by co-crystallisation with 2,4-HDA. Therefore, the remaining sections of this chapter will deal almost exclusively with the use of this crystal to determine the native structure of MhpC. Nevertheless, some reference will also be made to several of the lower resolution structures as these influenced numerous investigations into the structure and mechanism of MhpC prior to the collection of the 2.1Å data set.

Synchrotron diffraction data were collected with a single MhpC/2,4-HDA crystal using an ADSC CCD detector on station 9.6 (λ=0.87Å) at the synchrotron at Daresbury Laboratory, UK. Low temperature data (100K) was collected using an Oxford Instruments cryojet. The crystal initially diffracted X-rays to a resolution of 1.8Å (Figure 31) but towards the end of data collection diffraction to 2.2Å was observed, possibly as a result of crystal damage caused by radiation.

A total of 190 1°-oscillation frames were measured, with an exposure time of 30s per frame and a crystal to detector distance of 230.0 mm. The raw data were processed with MOSFLM (Leslie, 1992) and scaled and merged using programs of the CCP4 suite (CCP4, 1994). The merged intensities were subsequently converted to structure factor amplitudes using TRUNCATE (CCP4, 1994). The relevant data processing statistics for the MhpC/2,4-HDA crystal are shown in Table 5.

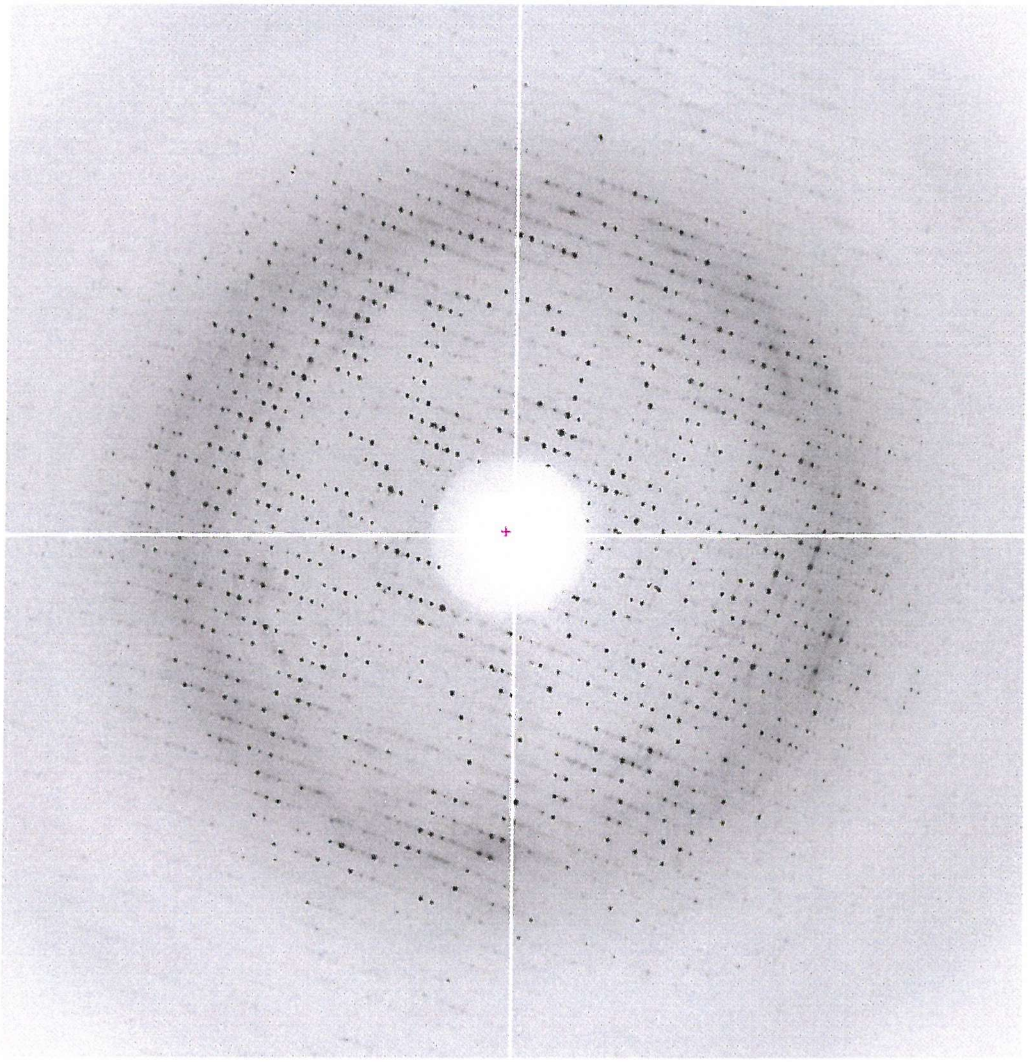


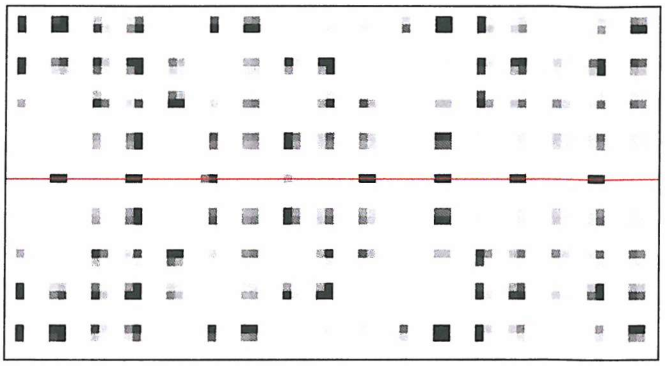
Figure 31: Typical diffraction pattern for the MhpC/2,4-HDA crystal. The crystal diffracted well to about 2.1Å. Over 180 1°-oscillation frames were collected on beamline 14.1 at the synchrotron at Daresbury Laboratory, UK.

The unit cell dimensions for MhpC/2,4-HDA (a=62.1Å, b=144.2Å, c=144.9Å, α=β=γ=90°) were determined using MOSFLM. This indicated that the crystal belonged to an 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 *k* and *l*, consistent with the presence of two 2_1 screw axes (Figure 32). A crystal form of MhpC had previously been found that belonged to orthorhombic space group $P2_12_12$, with unit cell dimensions of a=144.99Å, b=144.43Å and c=62.35Å (Mohammed, 2001). These observations indicated that the MhpC/2,4-HDA crystal was almost definitely identical to this crystal form and so the data were reindexed to give the same order of unit cell dimensions (a=144.9Å, b=144.2Å, c=62.1Å), and space group $P2_12_12$. This assumption would be proven correct by the attainment of an acceptable rotation and translation function solution during molecular replacement.

Total number of reflections	319359
Number of unique reflections	71450
Resolution (Å)	2.1
Completeness (%)	93.7
R_{merge} (%) (outer resolution shell)	7.5 (36.2)
Multiplicity	4.6
Average $I/\sigma(I)$ (outer resolution shell)	6.3 (1.9)

Table 5: Data processing statistics for the crystal of MhpC/2,4-HDA.


(a)

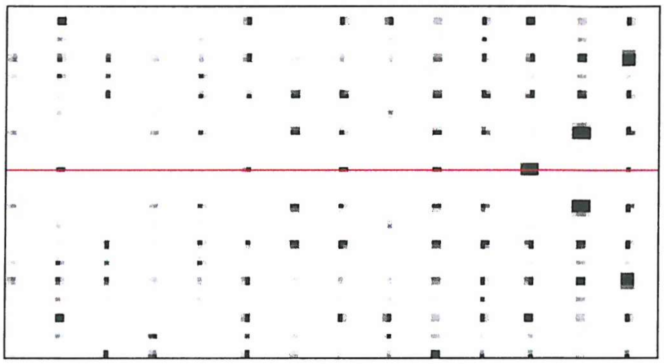

(b)

Figure 32: Reflections along axis *k* (**a**) and *l* (**b**) showing the pattern observed when a 2_1 screw axis is present (ie. every other reflection along the axis is missing).

4.2. Structure determination and refinement

4.2.1. Determination of structure by molecular replacement

The structure of MhpC had previously been solved using the MAD method. MhpC was 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=62.22Å, b=154.72Å, c=78.94Å, β=89.7Å) with 4 MhpC molecules per asymmetric unit. MAD data were collected at the ESRF beamline BM14 (Grenoble, France). The seleno-methionine crystal diffracted X-rays to a resolution of 2.3Å and phases were determined by locating 32 selenium atoms by direct methods using the 'Shake and Bake' program (Weeks & Miller, 1999).

4.2.1.1. Cross Rotation Search

A solvent content of 53% (V_m=2.59 Å^3/Da) was calculated (Matthews, 1968) assuming there are four molecules of MhpC (31,000 Da) in the asymmetric unit of the *E. coli* MhpC/2,4-HDA crystal. This is similar to that of the seleno-methionine crystal form. Initial phases for the MhpC/2,4-HDA 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 20 to 3.0Å resolution range. The rotation function yielded one significant peak at 14.75 σ, with a radius of integration of 25Å (Table 6).

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	155.45	90.00	182.91	14.75
2	160.58	48.50	98.08	4.77
3	156.88	74.85	149.72	4.50
4	149.00	48.05	107.70	4.33

Table 6: The 4 highest peaks following cross rotation calculations performed with MOLREP using data between 20 and 3Å.

4.2.1.2. Translation search

Since only one rotation function solution was determined (Peak 1 from Table 6), it seemed likely that either the estimated number of subunits was incorrect or that the four molecules of MhpC in the asymmetric unit all had the same orientation. The translation search, calculated in space group $P2_12_12$, confirmed the latter by giving four strong translation function peaks for the first rotation function solution. The position of the first molecule corresponded to fractional co-ordinate shifts of X=0.830, Y=0.753 and Z=0.265. The position of the first MhpC monomer was fixed (Table 7; 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 8, 9 and 10).

Peak Number	X	Y	Z	Dens/σ
1	0.830	0.753	0.265	10.43
2	0.342	0.271	0.271	8.43
3	0.356	0.252	0.284	7.93
4	0.233	0.271	0.274	7.76
5	0.328	0.271	0.270	7.63

Table 7: The 5 highest translation function peaks calculated in space group $P2_12_12$, using the first rotation function solution (α=155.45°, β=90.00° and γ=182.91°).

Peak Number	X	Y	Z	Dens/σ
1	0.340	0.529	0.261	27.02
2	0.856	0.252	0.285	26.93
3	0.344	0.014	0.287	24.35
4	0.340	0.982	0.261	14.39
5	0.316	0.493	0.276	12.12

Table 8: 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	0.856	0.252	0.282	45.10
2	0.344	0.015	0.287	39.89
3	0.316	0.029	0.246	20.31
4	0.340	0.980	0.261	17.53
5	0.830	0.303	0.263	16.95

Table 9: The 5 highest non-crystallographic translation function peaks calculated after fixing the position of two MhpC molecules.

Peak Number	X	Y	Z	Dens/σ
1	0.344	0.015	0.287	78.85
2	0.340	0.981	0.261	35.09
3	0.316	0.028	0.247	34.48
4	0.369	0.029	0.271	29.72
5	0.342	0.000	0.252	24.25

Table 10: The 5 highest non-crystallographic translation function peaks calculated after fixing the position of three MhpC molecules.

The crystal packing of the solution was visualised using MOLPACK (Wang *et. al.*, 1991) and shown to possess sensible crystal contacts (Figure 33).

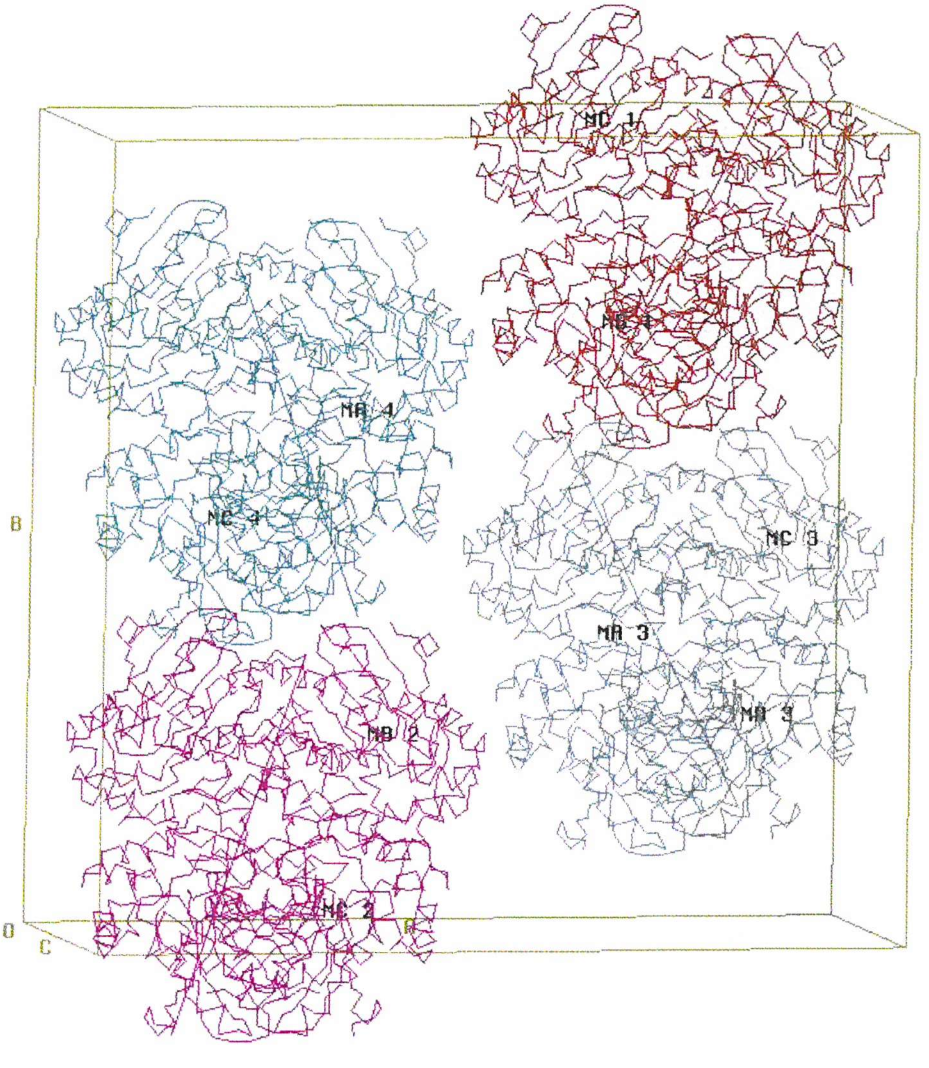


Figure 33: Crystal packing of MhpC in the MhpC/2,4-HDA crystal, generated using the MOLPACK program of the CCP4 suite. There are 4 molecules per asymmetric unit, and so therefore 16 MhpC molecules in the unit cell due to the orthorhombic spacegroup $P2_12_12$. It can be clearly seen that there are four tetrameric arrangements of MhpC per unit cell with no steric clashes between oligomers.

4.2.2. Refinement of the native MhpC structure

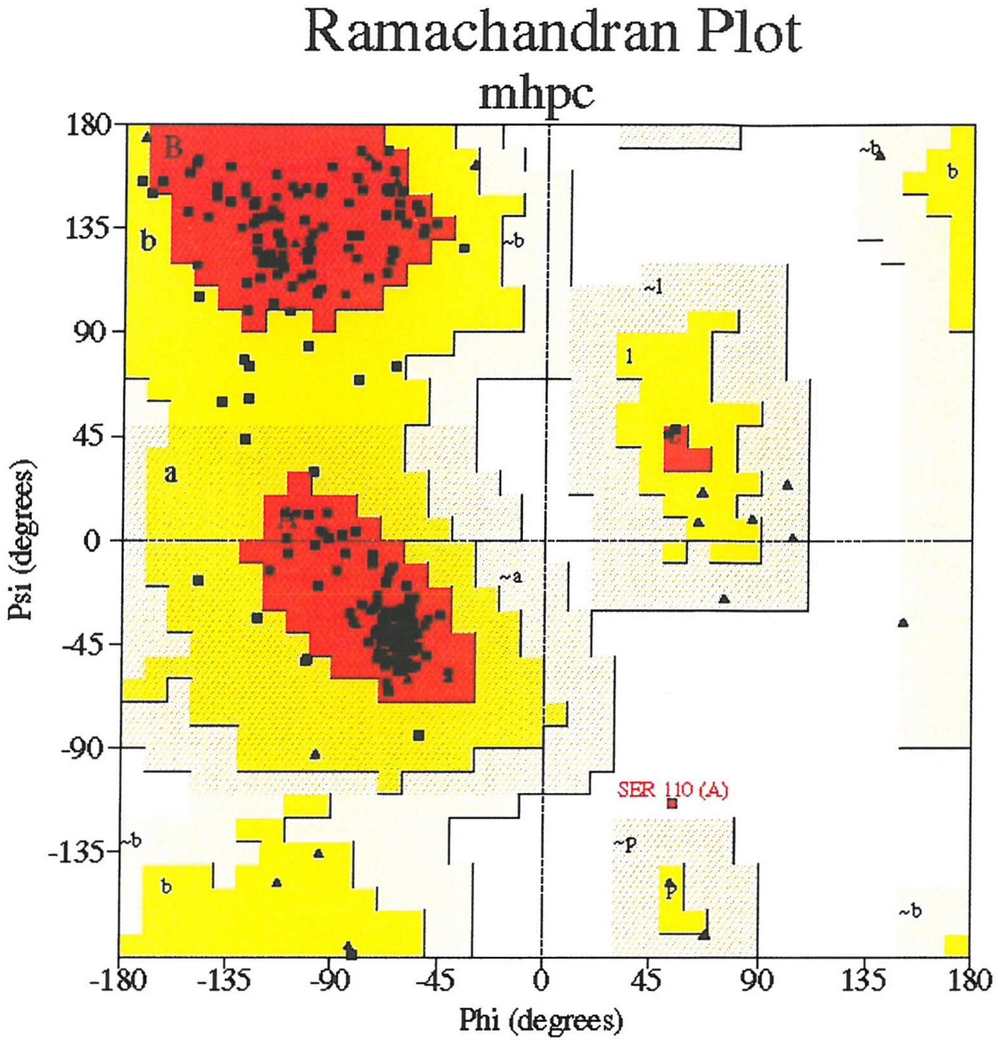
The model was refined with programs of CNS using 70355 reflections in the resolution range 20-2.1 Å. A test set of reflections (5% of data) was used for R_{free} calculations. During the early stages of refinement, excepting rigid body refinement, NCS restraints were maintained between the four monomers. The model was subjected to an initial round of refinement consisting of rigid body refinement, positional refinement and grouped B-factor refinement. This reduced the R_{factor} from 47.0% (R_{free} = 48.7%) to 33.1% (R_{free} = 36.4%). The resulting model was used to calculate sigmaA weighted electron density maps that were subjected to four-fold NCS averaging. The maps were calculated with CNS and converted into X-PLOR format using MAPMAN (Kleywegt & Jones, 1994).

All map inspections and model building were performed using QUANTA (Molecular Simulations Inc., Burlington, Mass., USA). Initial examination of the averaged maps revealed no well defined *Fo-Fc* density for the 2,4-HDA ligand. However, owing to the better resolution of the crystal, compared to previous crystals, the maps were of very good quality. The model was improved by several rounds of manual rebuilding and refinement consisting of rigid body refinement, positional refinement and grouped and individual B-factor refinement. During rebuilding, regions of the model in poor agreement with the electron density were adjusted. Furthermore, residues with side chains that lacked electron density were replaced with alanine.

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 3σ, reappeared in subsequent *2Fo-Fc* maps, formed hydrogen bonds with chemically reasonable groups and had B-factors less than 60Å^2. These water molecules were added in successive steps and were included in the subsequent refinement cycles. The refinement statistics after round 4 of model building and refinement are listed in Table 11. Inspection of the Ramachandran plot using PROCHECK (Laskowski *et. al.*, 1993) demonstrated acceptable geometry for most main chain and side chain parameters, except Ser110 which occurs in a disallowed region (Figure 34). This is not unexpected as this is also seen for a number of active site serine residues of related proteins, and is believed to be due to a sterically strained conformation (Ollis *et. al.*, 1992).

Resolution range (Å)	20.0 – 2.1
R_{factor} (%)	23.5
R_{free} (%)	27.4
Number of reflections in working set	66792
Number of reflections in test set	3563
Number of protein atoms	8764
Number of solvent atoms	529
RMS bond length deviation (Å)	0.0066
RMS bond angle deviation (°)	1.25

Table 11: Current refinement statistics for the native *E. coli* MhpC structure.



Plot statistics

```
Residues in most favoured regions  [A,B,L]                224    92.2%
Residues in additional allowed regions  [a,b,l,p]          18     7.4%
Residues in generously allowed regions  [~a,~b,~l,~p]       0     0.0%
Residues in disallowed regions                              1     0.4%
                                                          ---    -----
Number of non-glycine and non-proline residues            243   100.0%

Number of end-residues (excl. Gly and Pro)                  0

Number of glycine residues (shown as triangles)            28
Number of proline residues                                 13
                                                          ---
Total number of residues                                  284
```

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 34: Ramachandran plot of the current model for an MhpC protomer. Plot output from the CCP4 structure validation program PROCHECK (CCP4, 1994).

4.3. MhpC is a member of the α/β-hydrolase fold superfamily

4.3.1. Description of a single subunit of MhpC

MhpC is a globular protein (47Å x 45Å x 51Å) of the α/β-hydrolase fold (Ollis *et. al.*, 1992), containing a central β-sheet of eight β-strands with strand β2 running anti-parallel to the rest. The β-sheet has a half-twist of 180° from end to end, and on the concave face 3 α-helices pack, whilst on the convex face 2 α-helices pack (Figure 35). The subunit of the MhpC enzyme can be divided into two domains, the largest being the core domain (residues 1-133 and residues 217-284).

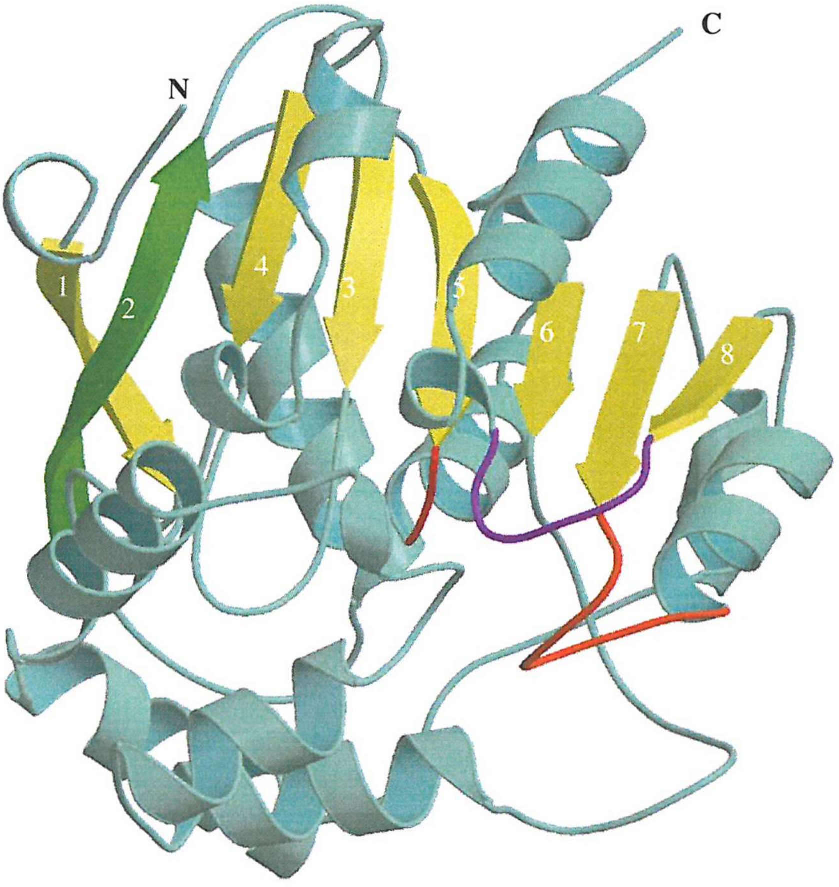


Figure 35: Structure of an MhpC subunit illustrating the α/β-hydrolase fold. The nucleophile (Ser110) is located on the red loop following β5, and the acid (Asp235) and histidine (His263) residues are located on the orange and purple loops, respectively. The parallel β-strands are depicted as yellow arrows, whilst the anti-parallel β2 is depicted as a green arrow.

The enzyme also possesses a 4 helix capping domain (residues 149-209), as shown in Figure 36, similar to that observed in lipases such as *Rhizomucor miehei* lipase (Brzozowski *et. al.*, 1991) and hydrolases such as epoxide hydrolase from *Agrobacterium radiobacter* AD1 (Nardini *et. al.*, 1999). However, these lid domains are not absolutely conserved amongst α/β-hydrolases. In those enzymes with capping domains, the helical lid must move to allow access for substrate to the active site, as well as normally contributing several residues important for interaction with substrates. Therefore, a similar mechanism could also be involved in allowing access of the large substrate of MhpC to its buried active site. The crystal structure (at 3Å resolution) of a complex of *R. miehei* lipase with *n*-hexylphosphonate ethyl ester, a substrate analogue, has shown that the enzyme active site is exposed by the movement of the helical lid (Brzozowski *et. al.*, 1991).

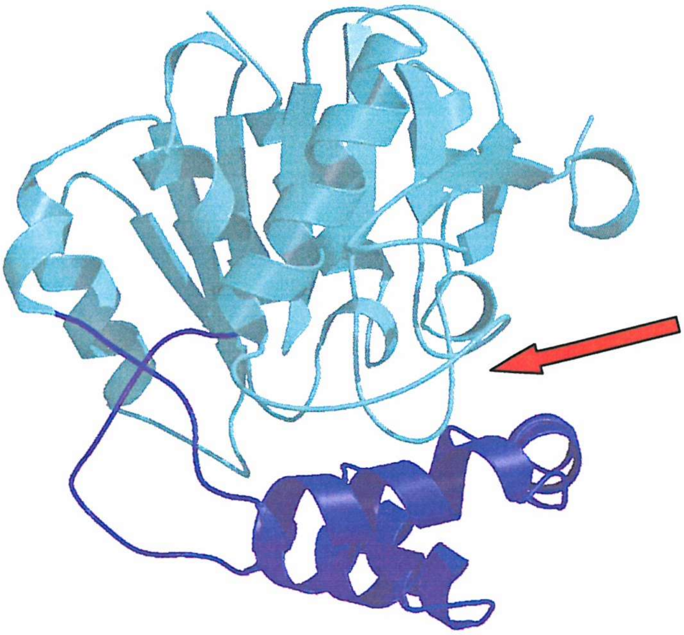


Figure 36: The core (cyan) and lid (blue) domains of the MhpC subunit. The core domain contains the central β-sheet of the canonical α/β-hydrolase fold. The lid domain consists of an α-helical bundle. The active site is sandwiched between these two domains (arrowed).

The loop regions connecting the two domains (residues 134-148 and residues 210-216) are quite flexible, indicated by the higher temperature factors for these regions compared to other parts of the molecule, and could therefore be the 'hinge' regions. Furthermore, the capping domain itself has higher B-factors than the core domain, as well as the *N*-terminal section comprising β-strands 1 and 2, implying that these regions are less well-ordered than the remainder of the protein (β-strands 3-8), as shown in Figure 37. This is not unexpected as this 6 strand β-sheet is thought to represent the minimal α/β-hydrolase fold and most members of the family are highly conserved over this region (Heikinheimo *et. al.*, 1999).

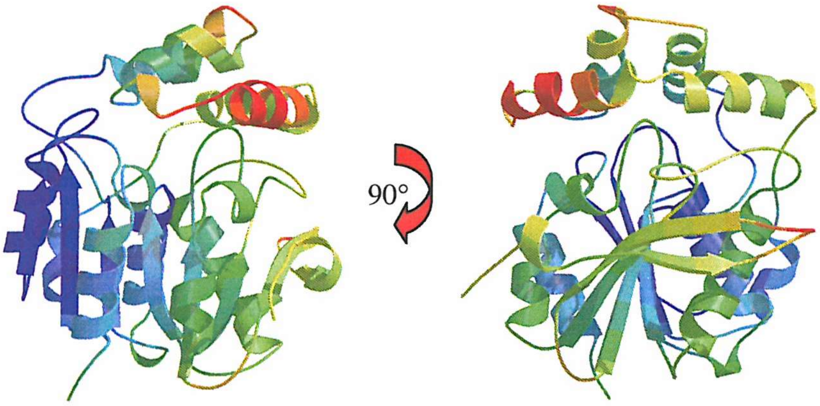


Figure 37: Representation of an MhpC monomer by the temperature factors of its amino acid residues. The lower temperature factors are in blue and range through cyan, green and yellow to red for the higher temperature factors. This illustrates that the helical domain, as well as some loop regions, have greater flexibility and that this could be important for access of substrate to the active site cavity (see text).

The capping domain of MhpC forms one boundary to the active site, whilst the other boundary is formed by loop regions of the β-sheet of the core domain. The topology of the MhpC structure is shown in Figure 38, and contains the canonical α/β-hydrolase fold. Helices αD-αG form the capping domain, whilst helices αA and αI pack onto the convex face of the β-sheet and helices αB, αC and αH pack onto the concave face of the β-sheet. The strand order for MhpC is 12435678, with the capping domain being formed by a large insertion following strand β6. This is a general feature of α/β-hydrolases, where the substrate-binding domains are often made of several insertions and form a cap on top of the active site (Heikinheimo *et. al.*, 1999).

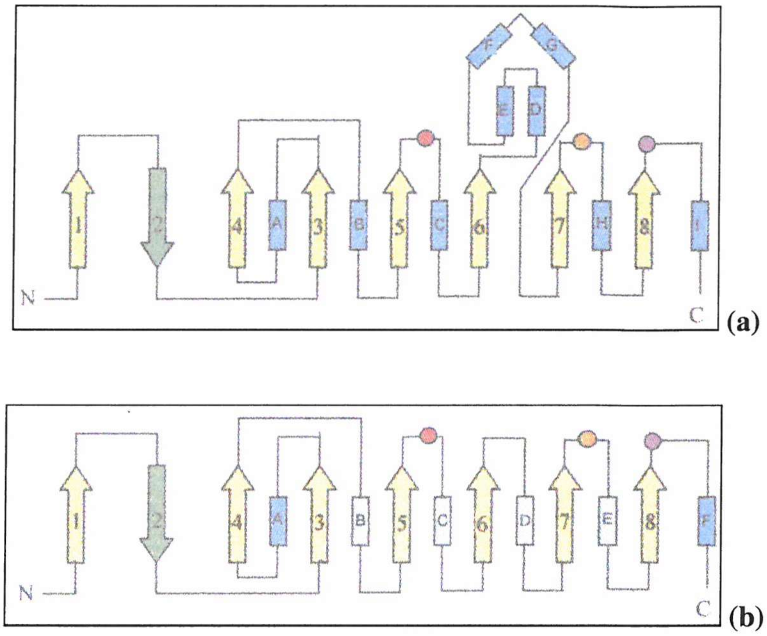


Figure 38: Topology diagram of MhpC **(a)** and comparison with the canonical α/β-hydrolase fold **(b)**. Strands are represented as arrows and helices as rectangles. The catalytic residues are represented by a red dot (nucleophile), orange dot (acid) and purple dot (histidine). In the canonical fold, the dark blue rectangles are helices on the convex face of the β-sheet and the light blue rectangles are helices on the concave face of the β-sheet. Colour scheme as in Figure 35.

MhpC is most probably a cousin of both the epoxide hydrolases (Nardini *et. al.*, 1999) and haloalkane dehalogenase (Verschueren *et. al.*, 1993), as although it possesses less than 30% sequence identity with either of them, it does share the feature of having an α-helical insertion between strands $\beta6$ and $\beta7$ which partially caps the active site cavity and contributes to the binding pocket. Primary sequence alignment studies of MhpC have identified three highly conserved residues that together may form a catalytic triad: Ser110, Asp235 and His263 (Ferrandez *et. al.*, 1997). Preliminary analysis of the active site reveals that Ser110 is located on the 'nucleophile elbow' following strand β5, whereas Asp235 and His263 are located on loop regions following strands β7 and β8. Thus, MhpC contains all of the requirements to be an α/β hydrolase fold enzyme: the sequence order of the triad is nucleophile-acid-histidine, with the nucleophile on canonical strand β5; there is a nucleophile elbow at the top of strand β5, with the sequence pattern Gly-X-Nuc-X-Gly; and, there is a long loop at the end of strand β7 that allows the side-chains of the triad residues to hydrogen-bond.

The residue numbering system has been altered from that obtained from sequence databases, as in all structures of MhpC solved, so far, the protein chain can only reliably be observed in *2Fo-Fc* electron density maps from Pro6 to Pro284. Therefore, the numbering system used is from Pro1 to Pro284, and consequently the active site triad comprises Ser110, Asp235 and His263. This coincides neatly with the residue range for C-C hydrolase homologue BphD which has the catalytic residues numbered identically to that of MhpC. Nevertheless, some residues of the flexible *N*-terminal pentapeptide MMSYQ can be seen in electron density maps for some structures of MhpC, but this is not generally the case. The *C*-terminus is not as flexible as the *N*-terminus as it is held by a salt-bridge formed between Arg283 and Glu254. The secondary structural features of a single MhpC protomer are tabulated overleaf (Table 12).

Residue & Number	**Structural Feature**
Pro1	N-terminus
Ser8 – Val13	Strand β1
Leu20 – Cys27	Strand β2
Glu32 – Gly39	Strand β3
Ser51 – Ala60	Helix αA (concave face)
Tyr62 – Cys69	Strand β4
Ser83 – Gln98	Helix αB (convex face)
Ile104 – Asn109	Strand β5
Asn109-<u>**Ser110**</u>-Met111	Nucleophile Elbow
Met111 – Glu125	Helix αC (convex face)
Gly128 – Gly134	Strand β6
Thr148 – Arg159	*Helix αD*
Thr162 – Val174	*Helix αE*
Thr181 – Arg195	*Helix αF*
Arg196 – Ala208	*Helix αG*
Ala225 – Gly232	Strand β7
Gly232 – Asp241	Acid Turn
Asp241 – Gly249	Helix αH (convex face)
Gly252 – Asp260	Strand β8
Asp260 – Ala265	Histidine loop
Ala270 – Ala282	Helix αI (concave face)
Pro284	C-terminus

Table 12: Secondary structural elements of MhpC by sequence number. Residues at the beginning and end of structural features are shown. Important catalytic features of the α/β-hydrolase fold are also shown. The putative catalytic serine residue is in bold and underlined. Helices of the lid domain are in italics.

4.3.2. The quaternary structures of MhpC

MhpC exists as a homodimer in solution (62 ± 2 kDa), as shown by gel filtration studies (Lam & Bugg, 1997), but in all crystal forms it exists as a homotetramer, so leading to the possibility of three different dimer associations. At present it is unclear which of these three possible oligomers is the physiological dimer. Most α/β-hydrolase fold enzymes are either monomers or dimers (Ollis *et. al.*, 1992). The related C-C hydrolase XylF has a native molecular mass of 65 kDa and is also believed to exist as a homodimer (Duggleby & Williams, 1986). However, other α/β-hydrolases have different oligomerisations, such as BphD, where the crystal structure revealed an octameric assembly (Nandhagopal *et. al.*, 2001).

4.3.2.1. Dimer arrangement 1: Formation of an elongated β-sheet

One possible dimer arrangement could be made by the alignment of the 8-strand β-sheets from two protomers to form a 16-strand β-sheet (Figure 39).

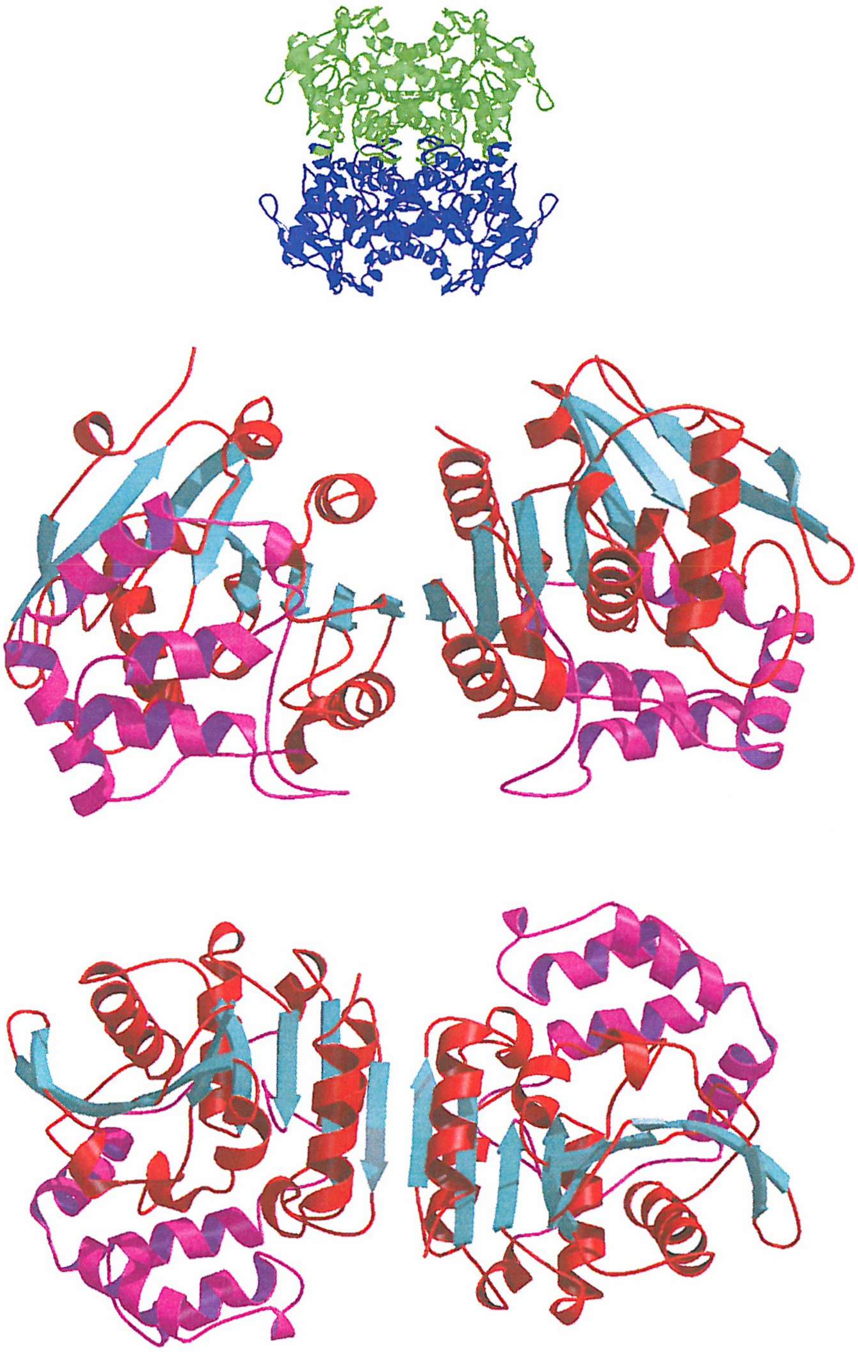


Figure 39: First possible dimer arrangement for MhpC. The dimer (blue) is depicted from the tetrameric arrangement (green). The dimer is formed by the alignment of the β-sheets from two protomers to form an extended 16-strand β-sheet. Helices are coloured red, strands are coloured cyan and the lid domains are coloured violet. Two views are shown and are related by a 90° rotation.

The β8 strand from one monomer runs anti-parallel with the β8 strand from another monomer forming the dimer interface (Figure 40). This interface is stabilised by six hydrogen bonding interactions between alternate backbone amide nitrogens of one monomer and backbone carbonyl oxygens of the other monomer (Table 13). The dimer interface is also stabilised by further hydrogen bonds between side-chains of Glu254 and His256 from opposite monomers and also interactions between Leu255 and Ile257 side-chains forming an hydrophobic pocket. The Glu254 side-chain also forms a salt-bridge with the Arg283 residue of its own subunit. A further hydrophobic stacking interaction occurs between the Phe258 residue and His269 of the same subunit. The side-chains of the residues involved in the dimer interface alternately point downwards and upwards, so that the Glu254, His256 and Phe258 side-chains point down and the Leu255 and Ile257 side-chains point up. The leucine and isoleucine side-chains from both subunits pointing upwards creates a very hydrophobic pocket of a cubic nature with the leucine and isoleucine residues at the bottom corners of the cube and with Leu224 and Met240 side-chains forming the top four corners. The methionine residue side-chains are positioned above the isoleucine side-chains and the leucine side-chains above the leucine side-chains. The dimer interface has a surface contact area of 2236Å^2, 10.6% of the total surface, as determined using the molecular surface program GRASP (Nicholls *et. al.*, 1991).

Subunit A residue	Subunit B residue	Distance (Å)
Ser253 O	Arg259 N	3.11
Leu255 N	Ile257 O	2.90
Leu255 O	Ile257 N	2.71
Ile257 N	Leu255 O	2.78
Ile257 O	Leu255 N	2.96
Arg259 N	Ser253 O	3.26
His256 NE2	Glu254 OE1	2.93
Glu254 OE1	His256 NE2	2.78
His256 ND1	His256 ND1	2.75

Table 13: Hydrogen-bonds between subunits at the dimer interface.

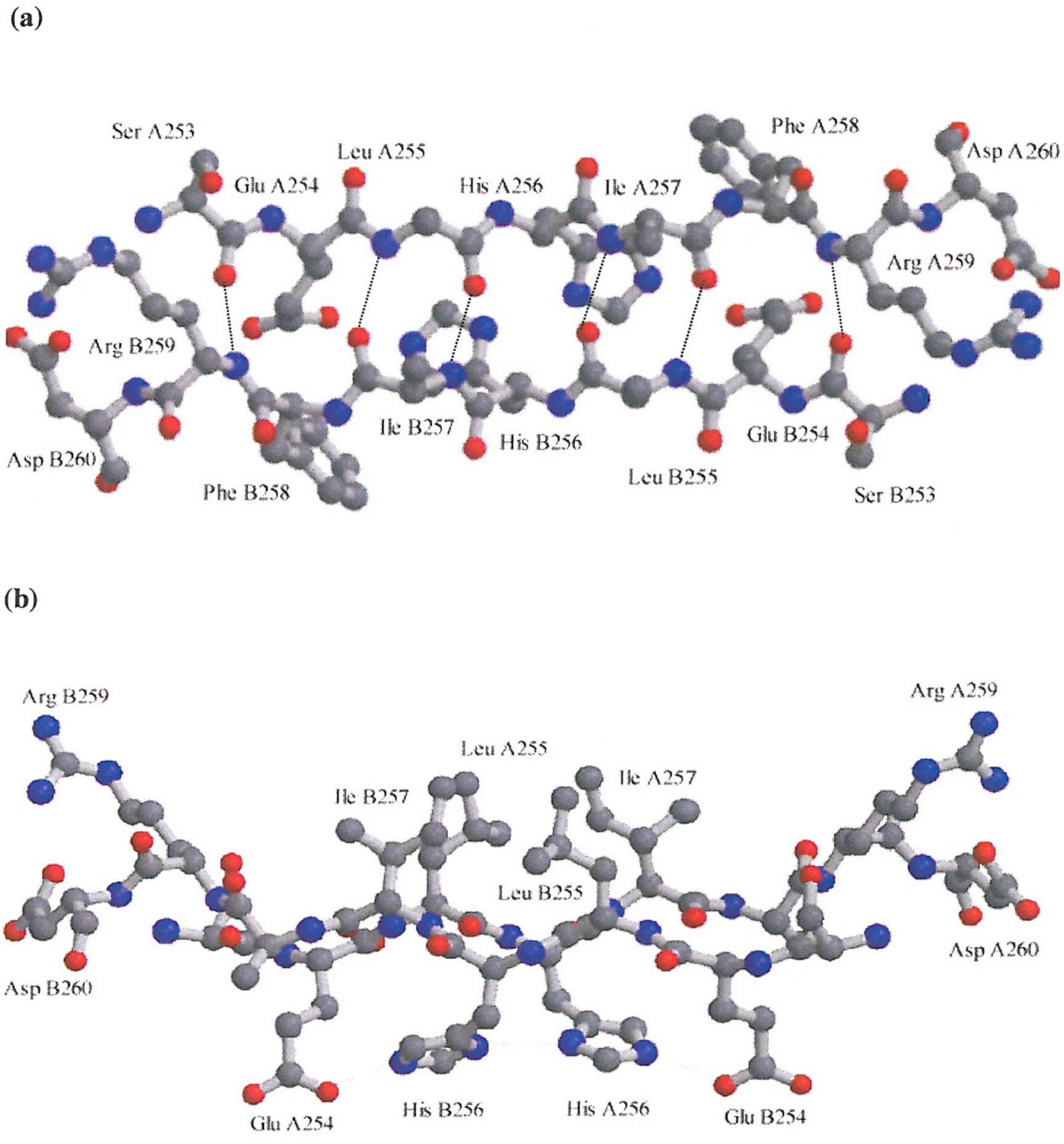


Figure 40: (a) Close-up of the dimer interface illustrating hydrogen-bonding interactions (dotted lines) between main-chain N and O groups of the β8 strand from one subunit running antiparallel with the β8 strand from another subunit; **(b)** Dimer interface viewed from the side showing the creation of a hydrophobic pocket by the interactions of Leu and Ile residues and the hydrogen-bonding of the His and Glu residue side-chains. Some residues have been changed to alanine to aid clarity.

This dimer interface is particularly stable as shown by the temperature factors of the residues involved (Figure 41). The β-strands close to the dimer interface have much lower B-factors than those strands at the beginning of the β-sheet (β-strands 1-3) as this interaction would appear to make the central portion of the extended β-sheet more rigid. The helical lid regions still have much higher B-factors compared to the core domain as they appear to be uninvolved in any inter-subunit association and are thought to be more flexible for substrate accessibility to the active site.

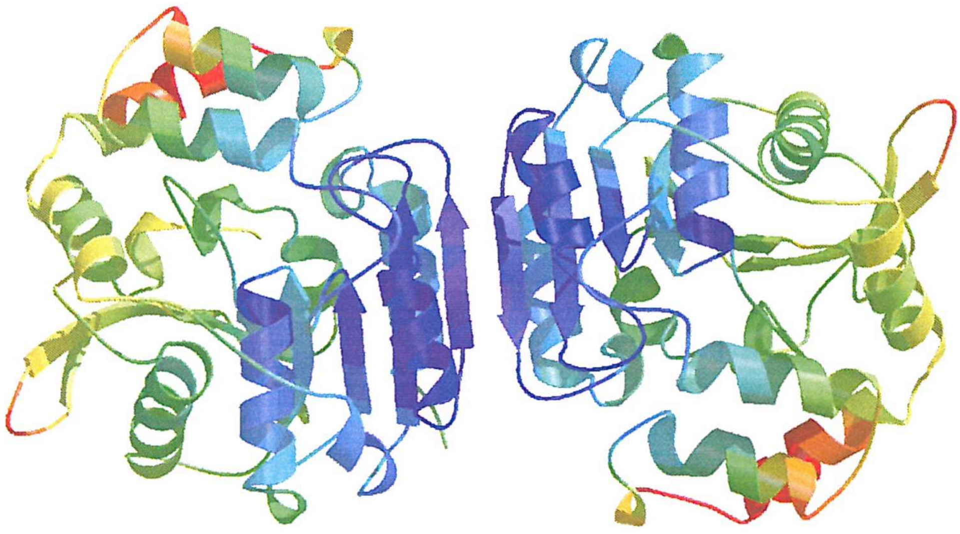


Figure 41: Temperature factors of the residues of MhpC in the extended β-sheet dimer formation. Residues are coloured as described in Figure 37. It suggests that this dimer formation is extremely stable as the central portion of the extended β-sheet has very low B-factors whilst the periphery β-strands as well as the lid domains have higher B-factors.

4.3.2.2. Dimer arrangement 2: Loop-helix hydrophobic interaction

This dimer arrangement is related by two-fold rotational symmetry, as shown in Figure 42. The area of contact between the subunits is smaller than the previous dimer arrangement and so therefore this dimer is not likely to be that observed physiologically.

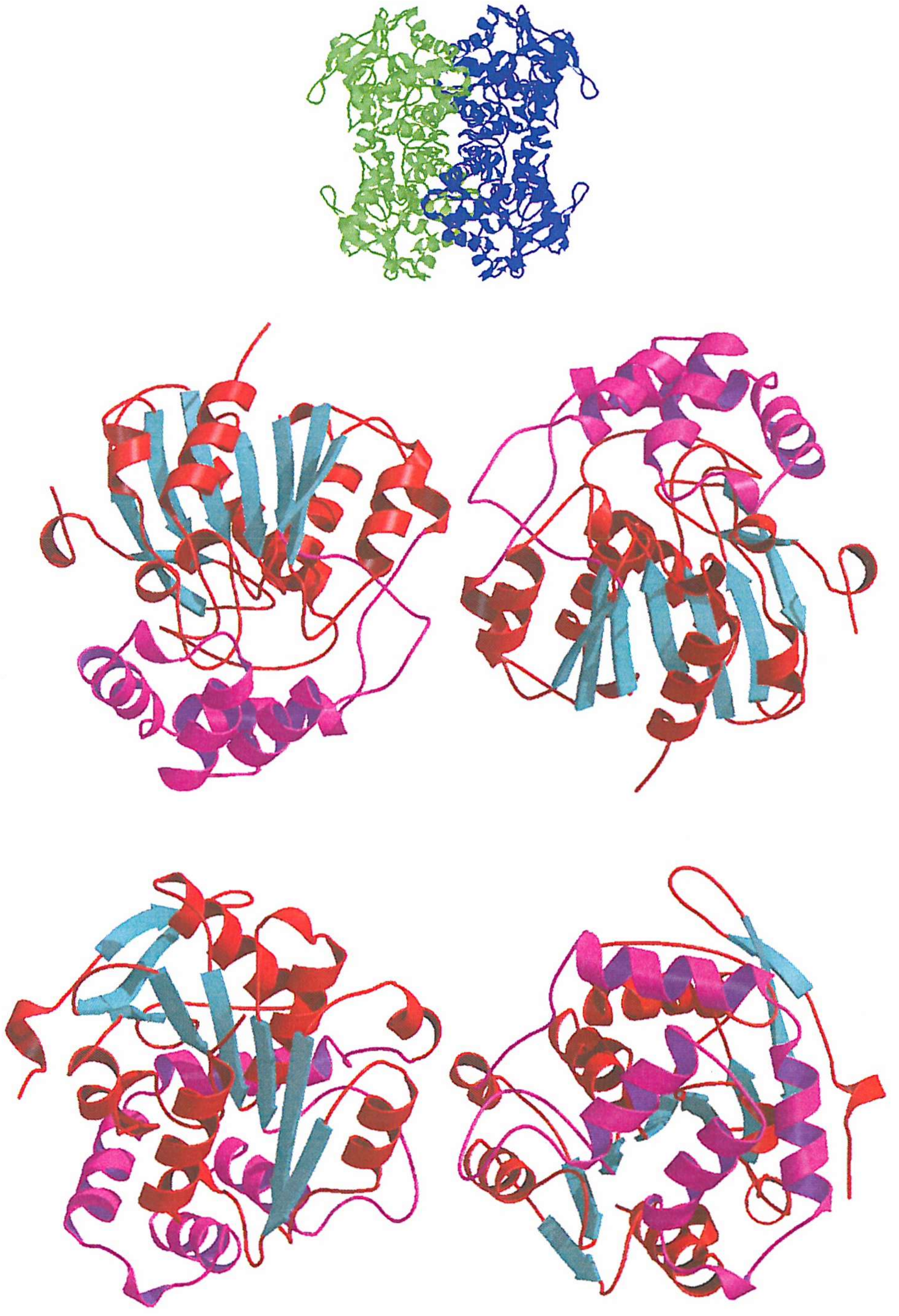


Figure 42: Second possible dimer arrangement for MhpC. The dimer (blue) is depicted from the tetrameric arrangement (green). This dimer is formed by the interactions of one loop from each subunit with a helix from the other subunit. Colour scheme is as described in Figure 39. Two views are shown and are related by a 90° rotation.

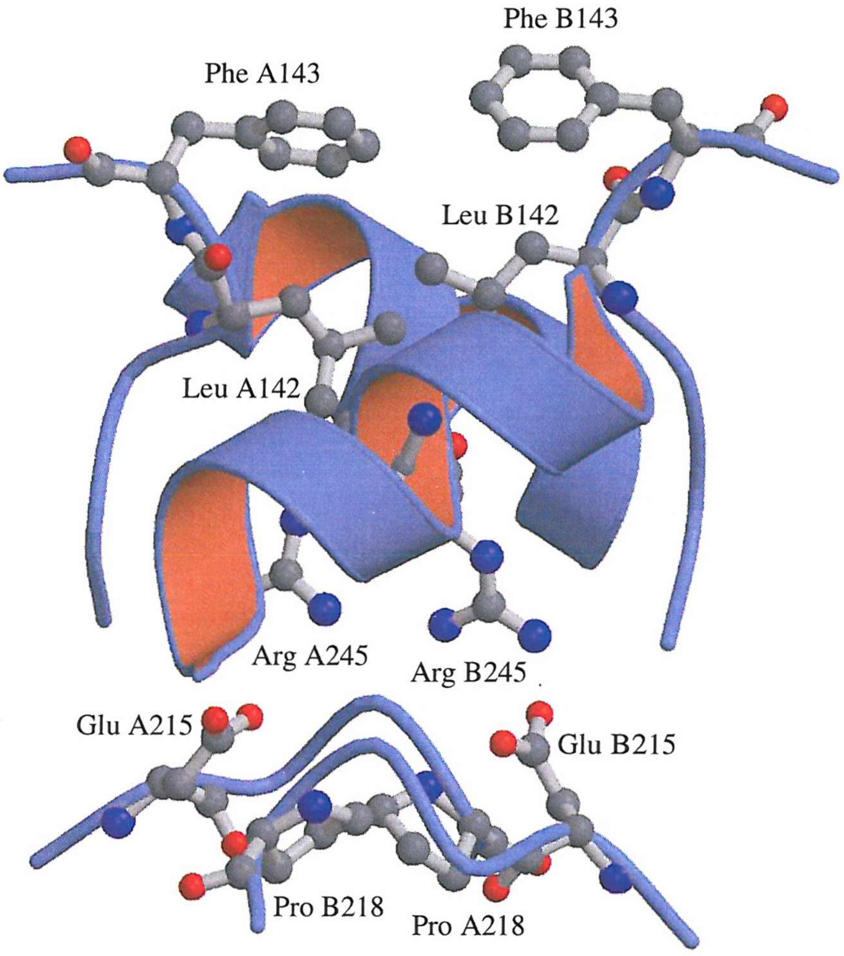


Figure 43: Close-up of the second possible MhpC dimer interface. The interactions are predominantly van der Waals contacts between Phe, Leu and Pro residues from each subunit. These residues are all borne on loops at the interface, and the presence of two buried salt-bridges between Arg and Glu residues also strengthens this interface. The Arg residues are on helices at the interface, whilst the Glu residues are present on the same loops as the Pro residues.

The dimer interface is created by the extended loops following strand β6, from each protomer, coming within van der Waals contact of one another. This loop contains a phenylalanine residue (Phe143) and both of the Phe side-chains align end to end with one another (Figure 43). The long loop following strand β6 is one which connects the core domain to the lid domain and so in solution may be more flexible than in the constrained environment of the crystal. The interaction between the protomers is of a mostly hydrophobic nature, and also involves two proline residues (Pro218 from each subunit) and two leucine residues (Leu142 from each subunit). Leu142, of the 'hinge' loop from one protomer, interacts with Leu142 of the other protomer. On the helix following strand β7 is Arg245 which forms a salt-bridge with Glu215 of the same protomer. This Glu residue resides on the loop before strand β7 which also contains Pro218. The Arg245-containing helices also pack against the long loops with the Phe143 and Leu142 residues so that all three of the structural elements at this interface are in some way interconnected. The dimer interface buries ~1469 Å^2 of surface area (6.9%), as found using the GRASP program (Figure 44), and suggests that, compared to the first possible dimer, this subunit association may be implausible outside of the crystal.

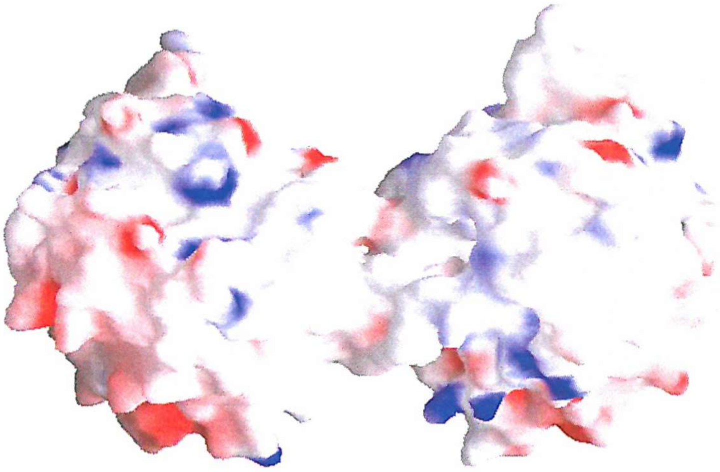


Figure 44: GRASP representation of the dimer interface of MhpC formed by interaction of hydrophobic residues of the loop following strand β7. The electrostatic surface potential is shown with positive regions in blue and negative regions in red.

4.3.2.3. Dimer arrangement 3: Phenylalanine stacking interaction

In this dimer arrangement the two subunits are also related by a non-crystallographic two-fold axis, as shown in Figure 45. This interface is formed by four loops coming into close proximity with one another, two loops from the lid domain and two loops from the core domain. In effect, one lid domain loop and one core domain loop are provided by each subunit.

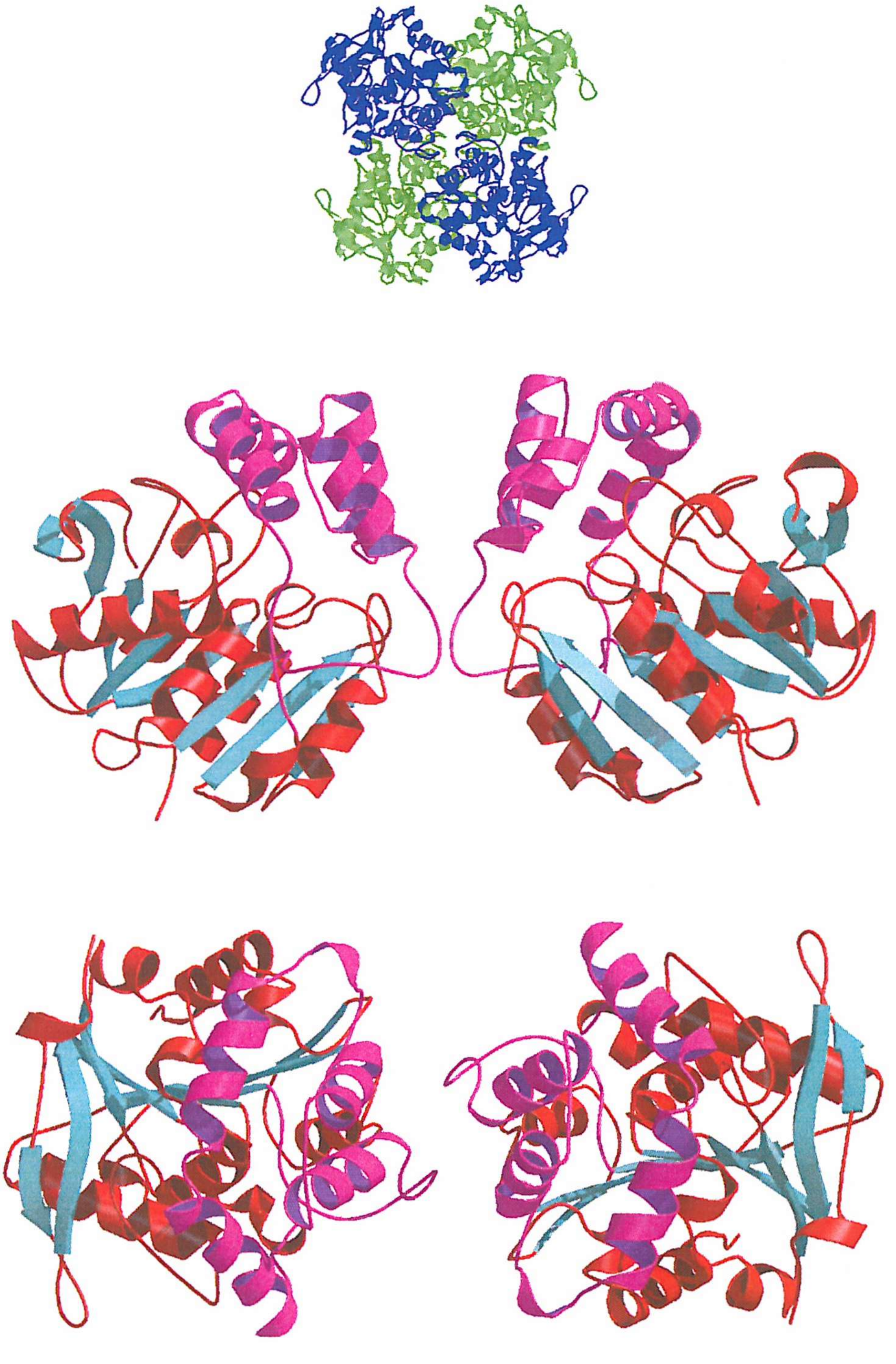


Figure 45: Third possible dimer arrangement for MhpC. The dimer (blue) is depicted from the tetrameric arrangement (green). This dimer is formed by two loop-loop interactions which creates both hydrophobic and hydrophilic regions at the interface. Colour scheme as in Figure 39. Two views are shown related by a 90° rotation.

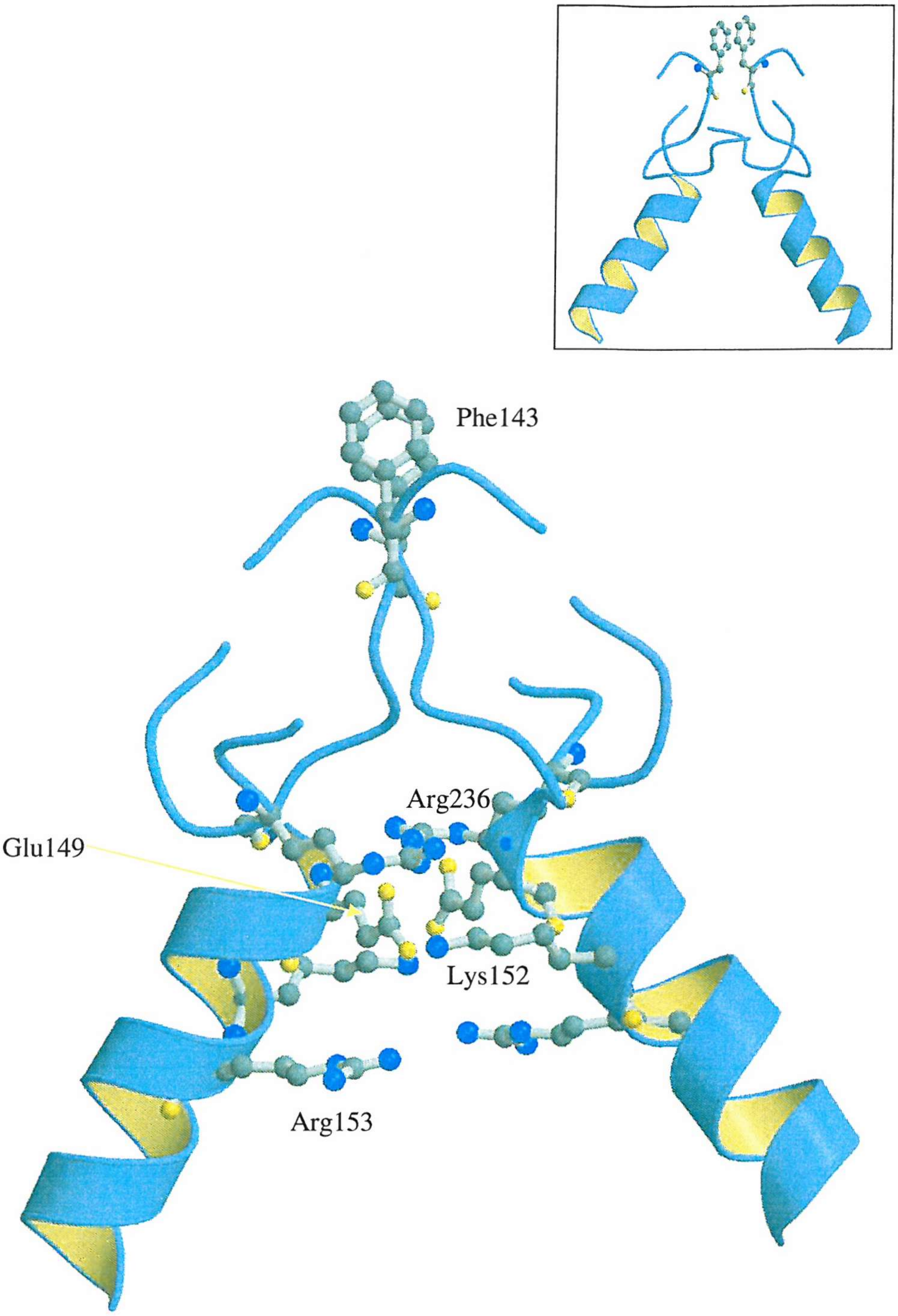


Figure 46: Close-up of the third possible dimer association. The numerous electrostatic interactions between Arg, Lys and Glu residues are shown in the main picture. The numbering of the residues is identical for both subunits. The hydrophobic stacking of the Phe143 side-chains is also shown, but is clearer in the inset picture.

The loops of the 'hinge' connecting strand β6 to the first helix (αD) of the lid domain interact with one another through hydrophobic stacking of their Phe143 side-chains. This forms one region of the dimer interface, with the adjacent region being formed through interactions of Glu149 from one protomer and Lys152 of the other protomer (Figure 46). Thus, the interface has a hydrophobic section and a hydrophilic region and this would therefore make the dimer less stable as the electrostatic interactions on one edge of the interface could easily be disrupted. The salt-bridge formed between Glu149 and Lys152 is of course repeated twice as the same interaction occurs in the opposite direction. Both of these residues are located on the same helix (αD), the first one of the lid domain. A further salt-bridge is formed to Glu149 by Arg236, which occurs on the loop following strand β7 of the same subunit. The presence of a further arginine residue (Arg154) on the same helix as the lysine, along with Arg236, causes this part of the molecule to be mainly positively charged. The surface contact area between the two subunits is 1814 Å^2, which is 8.6% of the total surface, and this smaller area when compared to the first possible dimer interface, indicates that this dimer may not be very likely outside of the crystal environment (Figure 47).

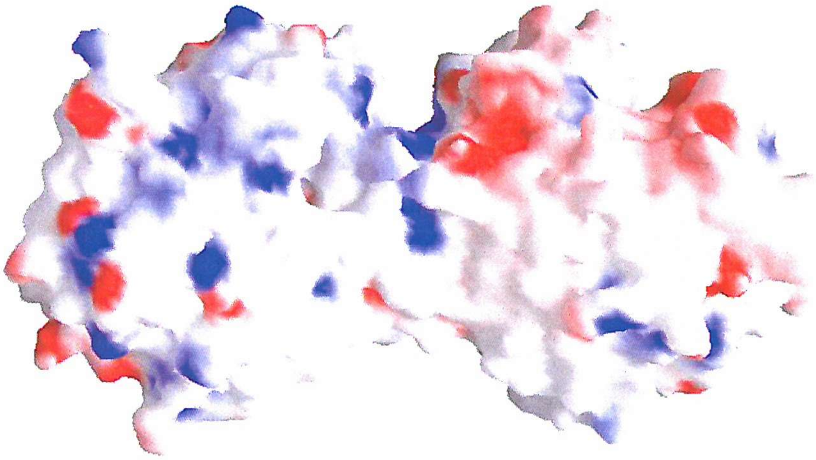


Figure 47: GRASP representation of a possible dimer interface of MhpC formed by a phenylalanine stacking interaction and salt-bridge interactions. The electrostatic surface potential is shown (coloured as in Figure 44) and illustrates that one portion of the interface is mostly hydrophobic whilst the other region of the interface is quite hydrophilic (mostly positively charged).

4.3.2.4. Tetrameric arrangement of MhpC observed in the crystal form

In the crystalline form MhpC exists as a tetramer, by the interaction of two dimers through hydrophobic stacking of the side-chains of Phe143 residues (Figure 48). The phenylalanine residue is located on a loop region that brings both dimers into contact with one another (Figure 49). This is a fairly weak interaction and so tetramer formation is almost certainly only possible due to the rigidity of crystal contact interactions. A tetramer for MhpC has not been observed in solution and so therefore it is unlikely that this oligomer is physiological. The majority of the inter-subunit contacts were discussed in greater detail in the preceding sections on the arrangement of the three possible dimers. The solvent accessible area excluded by this tetrameric interaction, as determined using GRASP, is in the region of 7332 Å^2, which is ~17% of the total surface area.

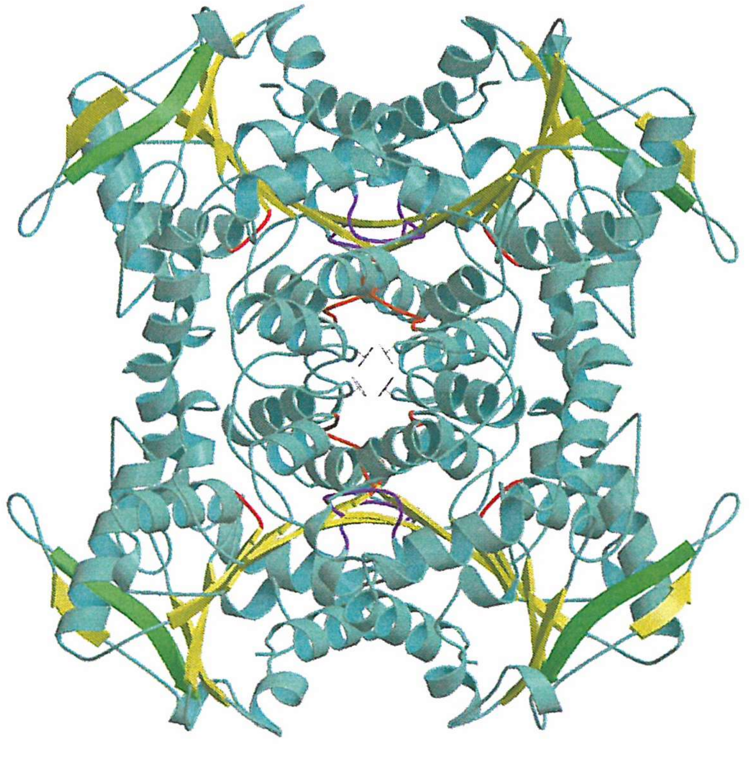


Figure 48: Tetrameric arrangement of MhpC observed in the crystal environment. This is not likely to exist outside of the crystal, and in fact only a dimer has been shown to exist in solution. Colour scheme as in Figure 35.

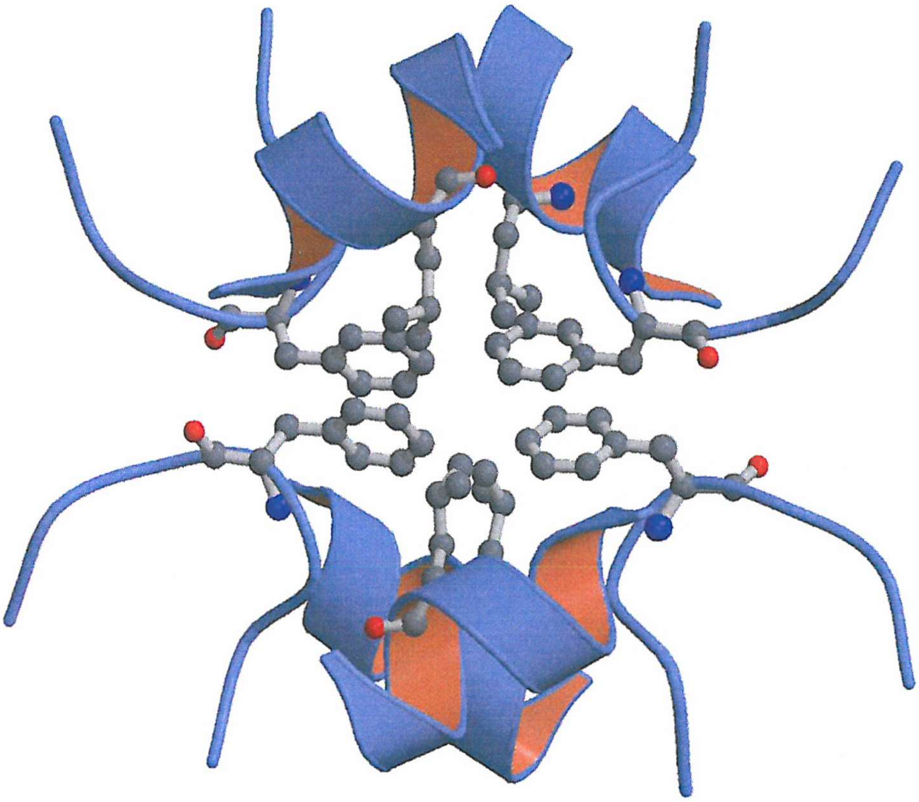


Figure 49: Close-up of the tetrameric association of MhpC (a dimer of dimers). The interaction is a mostly hydrophobic one formed by stacking of the side-chains of the Phe143 residues and also the close proximity of several leucine (ie. Leu244) and methionine residues (not shown).

4.3.2.5. Comparison of MhpC oligomerisation with related enzymes

The crystal structures of two C-C hydrolases of the α/β-hydrolase fold have thus far been determined: *Rhodococcus* sp. RHA1 BphD (Nandhagopal *et. al.*, 2001) and *Pseudomonas fluorescens* IP01 CumD (Fushinobu *et. al.*, 2002). The overall structure of the BphD enzyme is composed of eight monomers related by *422* point-group symmetry, which is identical with the crystallographic symmetry. The octameric structure can be regarded as a stack of two planar rings, each of which consists of four subunits related by a 4-fold rotational symmetry. Across the 2-fold axis, two β-sheets, each consisting of eight β-strands from the adjoining subunits, form a single large β-sheet consisting of 16 β-strands (Nandhagopal *et. al.*, 2001). The two subunits in the asymmetric unit of type-I CumD crystals were related by a non-crystallographic two-fold axis corresponding to the 1-to-5 interaction of RHA1 BphD. The β8 strands of both of the subunits form an anti-parallel β-sheet, and this tight interaction seems to be responsible for the dimeric structure of CumD in solution (Saku *et. al.*, 2002). The monomer in the asymmetric unit of the type-II CumD crystal structures showed the same dimeric interaction through a crystallographic two-fold axis (Fushinobu *et. al.*, 2002). Therefore, the subunit interactions of these two related proteins point to the first dimer association for MhpC being the physiological one, as indicated by the greater stability and extensiveness of its interface contacts. However, the alternative dimers that MhpC could form are not to be discounted as their surface contact areas are between 1450-1800 Å^2, and in comparison the V_H/V_L interface of a Fab typically buries 1000-1400Å^2 (Stanfield *et. al.*, 1993), thus indicating that they too could be relatively stable.

Nonetheless, the first dimeric arrangement is not only observed in C-C hydrolases with the α/β-hydrolase fold but has also been found for other members of this large super-family. The elongated β-sheet formation has been observed in the crystal structures of numerous esterases of the α/β-hydrolase fold, such as Brefeldin A esterase from *Bacillus subtilis* (Wei *et. al.*, 1999) and a hyperthermophilic carboxylesterase from *Archaeoglobus fulgidus* (De Simone *et. al.*, 2001).

4.3.3. The active site architecture of MhpC

4.3.3.1. General features of the active site

The active site cavity of MhpC is bordered on one side by the lid domain and on the other by the core domain (Figure 36). A long mostly hydrophobic loop following strand β3 (the fourth strand from the *N*-terminus of the β-sheet) extends down from the core domain and interacts with the lid domain. This loop consists of residues 37 to 45 and contains Ser40 and a Pro turn which are highly conserved features of this family of enzymes, as shown in the boxed section of Figure 50.

The size of the active site cavity appears to be rather large when the structure is viewed as a ribbon diagram, as illustrated in Figure 36. However, when a space-filling model is used it shows that the active site is of a funnel-like nature, broader at one end and narrowing towards the other. It is not an enclosed funnel though as it is possible to see 'light at the end of the tunnel', as it were, by viewing the space-filling model from the perspective of looking along the β-sheet (Figure 51).

Figure 50 (overleaf): Amino acid sequence alignment of MhpC with related enzymes. Sequences used: **1**. *E. coli* MhpC (2-hydroxy-6-oxonona-2,4-diene-1,9-dioic acid 5,6-hydrolase)(Ferrandez *et. al.*,1997); **2**. *Burkholderia cepacia* BphD (2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase)(Hofer *et. al.*, 1993); **3**. *P. pseudoalcaligenes* BphD (Watanabe *et. al.*, 2000); **4**. *Pseudomonas* sp. strain KKS102 BphD (Kimbara *et. al.*, 1989); **5**. *P. putida* F1 TodF (2-hydroxy-6-oxo-2,4-heptadienoate hydrolase)(Menn *et.al.*, 1991); **6**. *P. fluorescens* CumD (2-hydroxy-6-oxo-7-methylocta-2,4-dienoate hydrolase)(Habe *et. al.*, 1996); **7**. *Pseudomonas* sp. strain CF600 DmpD (2-hydroxymuconic semialdehyde hydrolase)(Nordlund & Shingler, 1990); **8**. *P. putida* XylF (2-hydroxymuconic semialdehyde hydrolase)(Horn *et. al.*, 1991); **9**. *Rhodococcus* sp. strain RHA1 BphD (Masai *et. al.*, 1997); **10**. *R. globerulus* HppC (2-hydroxy-6-oxonona-2,4-dienoate hydrolase)(Barnes *et. al.*, 1997); **11**. *P. stutzeri* 2-hydroxy-6-oxo-6-(2'-aminophenyl)hexa-2,4-dienoate hydrolase (Ouchiyama *et. al.*, 1998); **12**. *Xanthobacter autotrophicus* haloalkane dehalogenase (Janssen *et. al.*, 1989); **13**. *Psychrobacter immobilis* Lip1 (triacylglycerol lipase)(Arpigny *et. al.*, 1993); **14**. *Acinetobacter calcoaceticus* ELH2 (enol-lactone hydrolase II)(Shanley *et. al.*, 1994); **15**. *P. oleovorans* PHAB (poly-(3-hydroxyalkanoate) depolymerase)(Huisman *et. al.*, 1991). Key: Residue types – blue=polar positive; red=polar negative; green=polar neutral; purple=non-polar aromatic; white=non-polar aliphatic; brown=proline/glycine; yellow=cysteine. * = putative catalytic triad residues: Ser110, Asp235 and His263 (for MhpC). > = Start point of mature MhpC sequence, althougth only residues Pro6 to Pro284 are observed structurally. The conserved region that is boxed is an area that may be involved in oxyanion hole formation.

```
                                 >
MhpC         MDCRIWLVKRQNRELNMQEKMMSYQPQTEAATSRFL-------NVEEA--
burce_bphd   ---------------------MT--ALTESSTSKFV-------KINEK--
pseps_bphd   ---------------------MT--ALTESSTSKFV-------KINEK--
psesp_bphd   ---------------------MS--ELNESSTSKFV-------TINEK--
psepu_todf   ---------------------MT--NVNAEIGRMVL--------------
psefl_cumd   ---------------------MA--NL--EIGKSIL--------------
psesp_dmpd   -------------------MNAP--QNSPEIGREII--------------
psepu_xylf   -------------------MNAP--QQSPEIGREIL--------------
rhosp_bphd   ---------------------MA--KTVEIIEKRFP--------------
rglob_hppc   -------------------MTTT--EIPAYITDDFF-------GLEDKWI
psest_hoad   ---------------------ML--NKAEQISEKSE-------SAYVERF
xanau_halo   ----------------MINAIRTPDQRFSNLDQYPF------SPNYLDDL
psyim_lip1   -MLLKRLCFAALFSLSMVGCTNAPNALAVNTTQKIIQYERNKSDLEIKSL
acica_elh2   --------------------------PVFHFKDTLT--------------
pseol_phab   -------------------------MPQPYIFRTVE--------------
             .........10........20........30........40........50

MhpC         -G-KTLRIHFNDC-GQGD-ETVVL-LHGSGPGATGWANFSRNIDPLVEAG
burce_bphd   -GFSDFNIHYNEA-GNG--ETVIM-LHGGGPGAGGWSNYYRNVGPFVDAG
pseps_bphd   -GFSDFNIHYNEA-GNG--ETVIM-LHGGGPGAGGWSNYYRNVGPFVDAG
psesp_bphd   -GLSNFRIHLNDA-GQG--ERVIM-LHGGGPGAGGWSNYYRNIGPFVEAG
psepu_todf   --AGGIETNLHDV-GAGN--PVVL-VHGSGPGVTAWANW-RTVMPELSRH
psefl_cumd   --AAGVLTNYHDV-GEGQ--PVIL-IHGSGPGVSAYANW-RLTIPALSKF
psesp_dmpd   --AAGIRTNLHDS-GAGF--PLMM-IHGSGPGVTAWANW-RLVMPELAKS
psepu_xylf   --AAGYRTNLHDQ-GEGF--PALL-IHGSGPASPPGPTG-AGSFRS-SQT
rhosp_bphd   --SGTLASHALVA-GDPQSPAVVL-LHGAGPGAHAASNW-RPIIPDLAEN
rglob_hppc   ETADGELTHYHEL-GEGT--PILF-LHGSGTGVTAAANW-WLNLPVLSEQ
psest_hoad   VNAGGVETRYLEA-GKGQ--PVIL-IHGGGAGAESEGNW-RNVIPILARH
xanau_halo   PGYPGLRAHYLDE-GNSDAEDVFLCLHGEPTWSYLYRKM---IPVFAESG
psyim_lip1   TLASGDKMVYAEN-GNVAGEPLLL-IHGFGGNKDNFTR-----IARQLEG
acica_elh2   --AQDVALNYATF-GQADRPALIF-SNSLGTNLSMWQQQ----IAYFQDK
pseol_phab   --LDNQSIRTAVRPGKPHLTPLLI-FNGIGANLELVFPF----IEALDPD
             .........60........70........80........90.......100

MhpC         YRVILLDCPGWGKSDSVVNSGSRSDLN----ARILKSVVDQLDIA-KIHL
burce_bphd   YRVILKDSPGFNKSDAVVMDEQRGLVN----ARAVKGLMDALDID-RAHL
pseps_bphd   YRVILKDSPGFNKSDAVVMDEQRGLVN----ARAVKGLMDALGID-RAHL
psesp_bphd   YRVLLPDAPGFNKSDTVVMDEQRGLVN----ARSVKGMMDVLGIE-KAHL
psepu_todf   RRVIAPDMVGFGFTQRPHGIHYGVESW----VAHLAGILDALELD-RVDL
psefl_cumd   YRVIAPDMVGFGFTDRPENYNYSKDSW----VDHIIGIMDALEIE-KAHI
psesp_dmpd   RRVIAPDMLGFGYSERPADAQYNRDVW----VDHAVGVLDALEIE-QADL
psepu_xylf   RRVIAPDMLGFGYSERPADGKYSQARW----VEHAIGVLDALGIQ-QGDI
rhosp_bphd   FFVVAPDLIGFGQSEYPETYPGHIMSWVGMRVEQILGLMNHFGIE-KSHI
rglob_hppc   GRCIAIDSIGYGQSVVAPNTEYGIKEW----VRHAVRVLDALGIE-KTWI
psest_hoad   YRVIAMDMLGFGKTAKP-DIEYTQDRR----IRHLHDFIKAMNFDGKVSI
xanau_halo   ARVIAPDFFGFGKSDKPVDEEDYTFEF---HRNFLLALIERLDLR-NITL
psyim_lip1   YHLIIPDLLGFGESSKPMSADYRSEAQ----RTRLHELLQAKGLASNIHV
acica_elh2   YFVICYDTRGHGASSTPVG-PYRIDQL----GTDVIALLDHLQIP-QATF
pseol_phab   LEVIAFDVPGVGGSSTPRH-PYRFPGL----AKLTARMLDYLDYG-QVNV
             ........110.......120.......130.......140.......150
                *
MhpC         LGNSMGGHSSVAFTLKWPERVGKLVLMGG-------GTGGMSLFTPMPTE
burce_bphd   VGNSMGGATALNFALEYPDRIGKLILMGP-------GGLGPSMFAPMPME
pseps_bphd   VGNSMGGATALNFAIEYPERIGKLILMGP-------GGPGPSMFAPMPME
psesp_bphd   VGNSMGGAGALNFALEYPERTGKLILMGP-------GGLGNSLFTAMPME
psepu_todf   VGNSFGGALSLAFAIRFPHRVRRLVLMGA-------VG----VSFELT-D
psefl_cumd   VGNSFGGGLAIATALRYSERVDRMVLMGA-------AG----TRFDVT-E
psesp_dmpd   VGNSFGGGIALALAIRHPERVRRLVLMGS-------AG----VSFPIT-E
psepu_xylf   VGNSFGGGLALALAIRHPERVRRLVLMGS-------VG----VSFPIT-A
rhosp_bphd   VGNSMGGAVTLQLVVEAPERFDKVALMGS-------VG----APMNARPP
rglob_hppc   VGNSLGGWLAFQFAIDFPERLLGIVSMGT-------GG----AK--LT-G
psest_hoad   VGNSMGGATGLGVSVLHSELVNALVLMGS-------AG----LVVEIH-E
xanau_halo   VVQDWGGFLGLTLPMADPSRFKRLIIMNACLMTDPVTQPAFSAFVTQPAD
psyim_lip1   GGNSMGGAISVAYAAKYPKDVKSLWLVDS-------AGF-WSAGIPKSLE
acica_elh2   CGISMGGLTGQWLAIHFPERFNQVIVANT-------AA--KI----GEAQ
pseol_phab   IGVSWGGALAQQFAHDYPERCKKLVLAAT-------AA--GAVMVPGKPK
             ........160.......170.......180.......190.......200
```

```
MhpC            GIKRLNQLYRQPTIENLKLMMDIFVFDTSDLT--DALFEARLNNMLS---
burce_bphd      GIKLLFKLYAEPSYETLKQMLQVFLYDQSLIT--EELLQGRWEAIQR---
pseps_bphd      GIKLLFKLYAEPSYENLKQMIQVFLYDQSLIT--EELLQGRWEAIQR---
psesp_bphd      GIKLLFKLYAEPSLETLKQMLNVFLFDQSVIT--DELLQGRWANIQR---
psepu_todf      GLDAVWGY--EPSVPNMRKVMDYFAYDRSLVS--DELAELRYKASTRPGF
psefl_cumd      GLNAVWGY--TPSIENMRNLLDIFAYDRSLVT--DELARLRYEASIQPGF
psesp_dmpd      GLDAVWGY--NPSFAEMRRLLDIFAFDRNLVN--DELAELRYQASIRPGF
psepu_xylf      GLETAWGY--TPSLANMRRLLDLFAHDRTLVN--DELAELRYQASIRPGF
rhosp_bphd      ELARLLAFYADPRLTPYRELIHSFVYDPENFPGMEEIVKSRFEVANDPEV
rglob_hppc      ALAGHSNP--NLTEAGIRKTLELFVVDKSLVT--DELVSLRYQSALNDTA
psest_hoad      DLRPIINY--DFTREGMVHLVKALTNDGYKID--DAMINSRYTYATDEAT
xanau_halo      GFTAWKYDLVTPSDLRLDQFMKRWAPTLTEAE--ASAYAAPFPDTSYQAG
psyim_lip1      GATLENNPLLIKSNEDFYKMYDFVMYKPPYLP--KSVKAVFAQERIK---
acica_elh2      AWQARAQLVREQGLTPIAQTAATRWFTPGFIEDSPEIVE-KLSHDLAQGS
pseol_phab      VLWMMASPRRYVQPSHVIRIAPT-IYGGGFRRD-PELAM-QHASKVRSGG
                .......210.......220.......230.......240.......250
                                                         *
MhpC            RRDHLENFVKSLEA-NPKQFPDFGPRLAEIKAQTLIVWGRNDRFVPMDAG
burce_bphd      QPEHLKNFLISAQK-APLSTWDVTARLGEIKAKTFITWGRDDRFVPLDHG
pseps_bphd      QPEHLKNFLISAQK-APLSTWDVTARLGEIKAKTFITWGRDDRFVPLDHG
psesp_bphd      NPEHLKNFILSAQK-VPLSAWDVSARLGEIKAKTLVTWGRDDRFVPLDHG
psepu_todf      QEAFASMFPAPRQR-WVDALASSDQDIRDIRHETLILHGRDDRVIPLETS
psefl_cumd      QESFSSMFPEPRQR-WIDALASSDEDIKTLPNETLIIHGREDQVVPLSSS
psesp_dmpd      HESFAAMFPAPRQR-WVDGLASAEAAIRALPHETLVIHGREDQIIPLQTS
psepu_xylf      QESFAAMFPPPRQN-GVDDLASNETDIRALPNETLVIHGREDRIIPLQAS
rhosp_bphd      RRIQEVMF-ESMKA-GMESLVIPPATLGRLPHDVLVFHGRQDRIVPLDTS
rglob_hppc      SDRLAEVV-AARDR-DRTELPLDFDVLSRLDVPVLLIHGVQDVVIPVSRT
psest_hoad      RKAYVATMQWIREQ-G--GLFYDPEFIRKVPVPTLVVHGKDDKVVPVETA
xanau_halo      VRKFPKMVAQRDQACIDISTEAISFWQNDWNGQTFMAIGMKDKLLGPDVM
psyim_lip1      NKELDAKILE--QI-VTDNVEERAKIIAQYKIPTLVVWGDKDQIIKPETV
acica_elh2      AEGYASCCEALAE-------ADVRPQLQRISIPVLVIAGAQDPVTTVADG
pseol_phab      KMGYYWQLFAGLG-------WTSIHWLHKIQQPTLVLAGDDDPLIPLINM
                .......260.......270.......280.......290.......300
                                    *
MhpC            LRLLSGIAG-SELHIFRDCGHWAQWEHADAFNQLV---------LNFLAR
burce_bphd      LKLLWNIDD-ARLHVFSKCGHWAQWEHADEFNRLV---------IDFLRH
pseps_bphd      LKLLWNIDD-ARLHVFSKCGHWAQWEHADEFNRLA---------IDFLRQ
psesp_bphd      LKLIANMQD-AHVHVFPRCAIGRSGSTRTPSTG
psepu_todf      LRLNQLIEP-SQLHVFGRCGHWVQIEQNRGFIRLV---------NDFLAA
psefl_cumd      LRLGELIDR-AQLHVFGRCGHWTQIEQTDRFNRLV---------VEFFNE
psesp_dmpd      LTLADWIAR-AQLHVFGQCGHWTQIEHAARFASLV---------GDFLAE
psepu_xylf      LTLAQWIPN-AQLHVFGQCGHWTQIEHAERFARLV---------ENFLAE
rhosp_bphd      LYLTKHLKH-AELVVLDRCGHWAQLERWDAMGPML---------MEHFRA
rglob_hppc      WELLNVIPN-ADVHIFSQCGHWSQVERAEEFNTVI---------TQYLSA
psest_hoad      YKFLDLIDD-SWGYIIPHCGHWAMIEHPEDFANAT---------LSFLSR
xanau_halo      YPMKALINGCPEPLEIADAGHFVQ-EFGEQVAREA---------LKHFAE
psyim_lip1      NLIKKIIPQ-AQVIMMEDVGHVPMVEALDETADNY---------KAFRSI
acica_elh2      QFLCEHIVH-STLEVLEA-SHISNVEQPQAFNHAV---------EAVMKR
pseol_phab      RLLAWRIPN-AQLHIIDD-GHLFLITRAEAVAPIIMKFLQQERQRAVMHP
                .......310.......320.......330.......340.......350

MhpC            P
burce_bphd      A
pseps_bphd      A
psesp_bphd
psepu_todf      ED
psefl_cumd      ANTPKLVGRP
psesp_dmpd      ADAAAIS
psepu_xylf      ADALHS
rhosp_bphd      A
rglob_hppc      RGVSRS
psest_hoad      RADITRAAA
xanau_halo      TE
psyim_lip1      LEAQR
acica_elh2      FN
pseol_phab      RPASGG
                .......360.......370.......380.......390.......400
```

By viewing the active site in this way it can be seen that it would be a disadvantage to the enzyme for access of the substrate to be made without the protein in some way opening. The RFP substrate of MhpC is a fairly long chain with a rigid dienoic acid moiety and so simply 'feeding in' to the active site through the larger opening of the funnel would be a highly ineffective way of accessing the catalytic apparatus of the enzyme. From Figure 51 it is possible to clearly see both Ser110 and Ser40 when looking into the active site cavity from the larger opening. It is also possible to see Arg188 and Asn49 at the far end of the funnel where it has narrowed greatly, and both of these residues are fairly conservative among members of this family of hydrolase. The larger opening (~ 15Å in diameter) of the funnel is formed by a ring of hydrophobic residues, except for Asp215 and Lys211, with His114 also being in the vicinity of the entrance to the funnel.

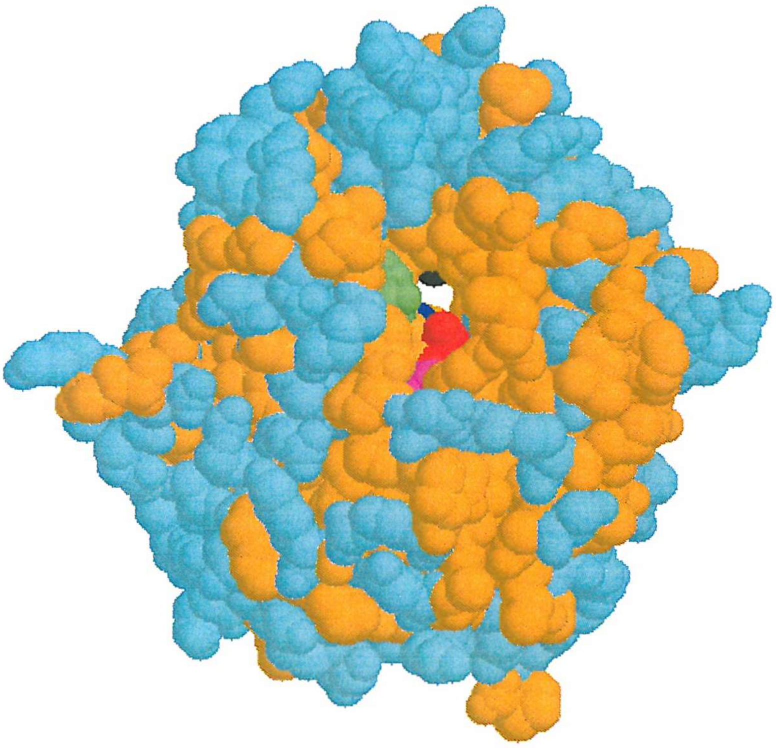


Figure 51: Space-filling model of an MhpC protomer showing the tunnel traversing the whole subunit. The view from the wider opening showing the positions of Ser110 (red), Ser40 (green), Arg188 (black), Asn49 (blue) and His114 (magenta). Apart from these residues the hydrophobic residues are coloured brown and hydrophilic residues are depicted in cyan.

The smaller opening of this tapering tunnel (~ 8Å in diameter) is also surrounded by mostly hydrophobic residues, and is much too small to allow access to substrate or exit of either product of the MhpC-catalysed reaction. It is also possible to see Ser110 in a space-filling model from this opening to the active site cavity (Figure 52).

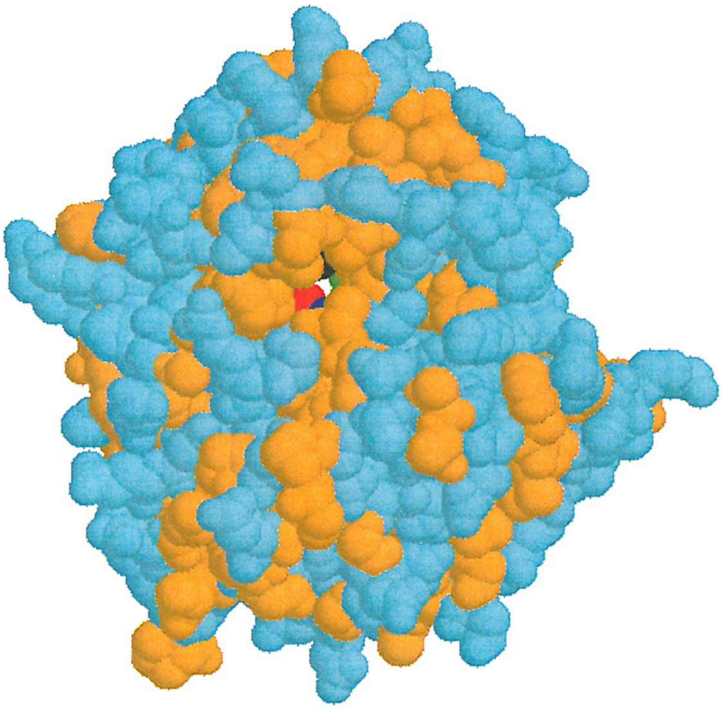


Figure 52: Space-filling model of an MhpC protomer viewed from the smaller opening of the tunnel. The residues are coloured as in Figure 51.

This feature of having a 'tunnel' traversing the whole subunit of the enzyme is not unique to MhpC, as it has also been observed in another enzyme, the macrocycle-forming thioesterase (TE) domain from the 6-deoxyerythronolide B synthase (DEBS), which is a related member of the α/β-hydrolase fold family (Tsai *et. al.*, 2001). The tertiary architecture of this protein belongs to the α/β-hydrolase family with two unusual features unprecedented in the family: a leucine-rich dimer interface and a substrate channel that passes through the entire protein. As the active site triad is situated in the middle of the substrate channel it is anticipated that the substrate passes through the protein. This is unlikely to be the case for MhpC as the substrate is much too cumbersome to pass completely through the protein, especially considering the narrowing of the active site cavity.

4.3.3.2. MhpC possesses an active site catalytic triad

From primary sequence alignment with members of the α/β-hydrolase fold family, three highly conserved residues have been identified that may putatively form the catalytic triad: Ser110, Asp235 and His263 (Figure 50). In common with many members of this family, Ser110 is located on the 'nucleophile elbow' following strand β5, and Asp235 and His263 are located on loop regions following strands β7 and β8, respectively.

It has been noted that many C-C bond hydrolases that work on substrates from single phenol compounds have a phenylalanine residue after the catalytic serine, such as for TodF, CumD, DmpD and XylF, whereas hydrolases of biphenyl compounds, like BphD, have a methionine residue after the serine (Diaz & Timmis, 1995). However, it can be seen from the sequence alignment shown in Figure 50 that MhpC also has a methionine after the triad serine residue, and so places this single phenol degradative enzyme apart from the others, and more in common with BphD and some lipases, such as *Psychrobacter immobilis* triacylglycerol lipase and *Acinetobacter calcoaceticus* enol-lactone hydrolase II, which also have a methionine residue in this position. The substitution of Phe by Met in position 108 of XylF (the residue following active site Ser107) created an enzyme with increased thermostability and altered substrate specificity, thereby indicating that the nature of this residues side-chain may have some relevance in the case of MhpC and for this family in general.

4.3.3.3. Structural and mechanistic details of the catalytic triad

From the structure of native MhpC around the active site, it can be seen that these residues do in fact form a triad reminiscent of that seen in many serine hydrolases (Figure 53). The Φ, Ψ angles of Ser110 occur in an unfavourable region of the Ramachandran plot, calculated with PROCHECK (Laskowski *et. al.*, 1993), and are due to its strained conformation. It is known that the strained turn structure, the 'nucleophile elbow', is well conserved in all the α/β-hydrolases (Heikinheimo *et. al.*, 1999; Arand *et. al.*, 1994) and that other members of the family also have this residue in the unfavourable region of the Ramachandran plot.

Prior to structural determination of the 2.1Å dataset the highest resolution structures of MhpC had been determined to ~2.3-2.5Å from failed crystal soaking experiments. In these structures the catalytic triad Asp235 was within H-bonding distance (~2.7Å) of His263. However, His263 and Ser110 were not observed to be within H-bonding distance (~4Å). This would lend weight to the proposed alternative mechanism for MhpC, instead of the more usual serine hydrolase mechanism in which the serine acts as a nucleophile to attack the carbonyl of the substrate. However, it cannot be discounted that the serine or histidine side-chains could undergo a conformational shift on substrate binding to bring them to within hydrogen-bonding distance.

Nonetheless, a number of related enzymes also have some interesting differences from the normal catalytic triad of serine hydrolases. Acetylxylan esterase from *Penicillum purpurogenum*, an α/β-hydrolase enzyme similar to cutinase from *F. solani*, has recently had its structure solved to 0.9Å at two different temperatures (Ghosh *et. al.*, 2001). In this enzyme, the serine and histidine residues hydrogen bond in the 295K structure, but in the 85K structure the serine points away from the histidine and instead hydrogen bonds to a sulphate group. Of more interest is the structure of human acyl protein thioesterase I, an α/β-hydrolase enzyme similar to *P. fluoroscens* carboxylesterase. From the 1.5Å structure, it was revealed that although this enzyme does contain a Ser114-Asp169-His203 triad, the His-Asp pair does not bind to the serine but instead the site typically occupied by the hydroxyl is taken by a well-ordered water molecule (Devedjiev *et. al.*, 2000). This could be analogous to the mechanism proposed for MhpC based on biochemical evidence, that of base-catalysed attack of water on the substrate, as seen in the aspartic proteases.

Therefore, the catalytic triad of MhpC may in fact just be a dyad of Asp235 and His263, with Ser110 acting in a non-nucleophilic capacity. More direct evidence for this proposal comes from another α/β-hydrolase fold enzyme, epoxide hydrolase from *Agrobacterium radiobacter* AD1. The X-ray structure of this enzyme has been determined to 2.1Å resolution and the catalytic residues found to be Asp107, His275 and a water molecule (Nardini *et. al.*, 1999). Asp107 occurs on the 'nucleophile elbow' of this enzyme, and the hydrolytic water is within hydrogen bonding distance of the NE2 atom of His275.

This epoxide hydrolase enzyme also contains a tetrapeptide motif HGXP around residue 40 which is conserved amongst α/β-hydrolase fold enzymes. A similar motif of HG(G/S)GP is also observed in MhpC and related enzymes, as shown in Figure 50. As noted previously, the loop bearing Ser40 extends from the core domain of the enzyme. This brings the Ser40 side-chain close to the 'nucleophile elbow' bearing Ser110 at the active site of the enzyme. In fact, the distance between the OG atoms of both these serine residues is in the region of 4Å. Therefore, it could be involved in the reaction mechanism, as the tetrapeptide motif in α/β-hydrolases is thought to be involved in stabilising the charge developing on the OD2 atom of the nucleophile during hydrolysis of the intermediate and also by stabilising the position of the hydrolytic water molecule.

Thus, it could be proposed that in MhpC there is a catalytic triad of Asp-His-Water, and the two serine residues at the active site, Ser110 and Ser40, function in stabilising the formation of a *gem*-diol intermediate and also in positioning the water molecule for hydrolysis.

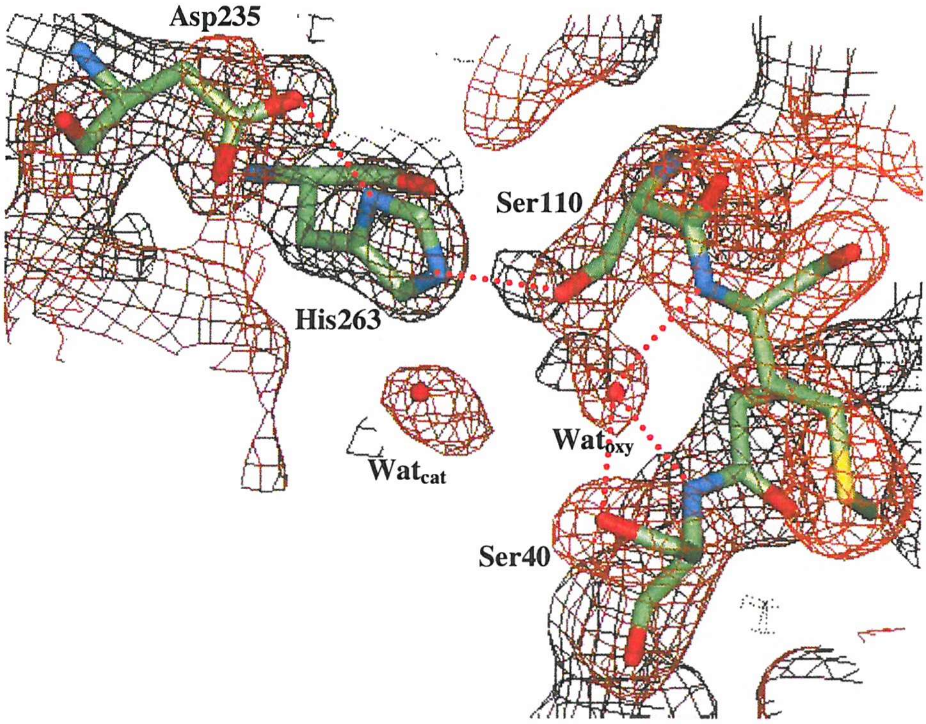


Figure 53: Catalytic triad of MhpC, from the 2.1Å structure, showing that the Asp235-His263-Ser110 residues are within hydrogen-bonding distance (see text). The two water molecules in the active site could be important mechanistically, one water seems to occupy the position known as the oxyanion hole and the other water molecule could be in a position to hydrolyse the substrate. The oxyanion hole is formed by two backbone amide groups and the side-chain of Ser40. The *2Fo-Fc* electron density is contoured at 1.3σ.

More recently, the higher resolution structure obtained for MhpC has thrown even further light on this interesting facet of the two different mechanisms that may be involved. This 2.1Å structure has revealed some features of the catalytic apparatus that in the lower resolution structures were ambiguous. The hydrogen-bonding distance from Asp235 to His263 is still of the same order of magnitude (2.5Å), but the distance from His263 to Ser110 is reduced significantly (3.2Å), although still at the upper limits for a good hydrogen bond interaction (Figure 53). However, the most interesting feature obtained from this structure was the location of two water molecules in the active site of the enzyme. One water (Wat$_{oxy}$) is thought to occupy the position known as the oxyanion hole, and is observed in a number of serine hydrolase structures, such as the 1.2Å structure of *Salmonella typhimurium* aspartyl dipeptidase (peptidase E) which has a unique serine protease fold and catalytic triad of Ser-His-Glu (Håkansson *et. al.*, 2000). In the case of MhpC, this water is within hydrogen-bonding distance (2.81Å) of the backbone amide nitrogen of the residue following the catalytic serine (Met111 N) and also within hydrogen-bonding distance of the backbone amide nitrogen (2.90Å) and side-chain oxygen atom (2.67Å) of Ser40, which although fairly unusual has been implicated as being an oxyanion hole participant in related enzymes. More unusual is that this water molecule is also very close to the side-chain oxygen of the catalytic serine, being only 2.3Å distant, and hence in extremely close hydrogen-bond contact, as normally hydrogen bonds are in the range of 2.5-3.5Å. The close proximity of this water could be due in some respect to the highly unusual character of this residue as compared to a 'normal' serine residue. The other water molecule of interest (Wat$_{cat}$) could be analogous to the water thought to be catalytically important for this class of enzyme, and seen in related structures such as the epoxide hydrolase and thioesterase mentioned previously. This water is within hydrogen-bonding distance of both His263 NE2 (3.12Å) and Ser110 OG (3.50Å), and is in a position where it could be involved in hydrolysis of the acyl-enzyme intermediate. Alternatively, it could be the water which results in formation of the *gem*-diol intermediate.

4.3.3.4. Unusual electron density at the active site of MhpC

In several structures of MhpC there is unusual extraneous electron density at the active site, which is contiguous with that of the side-chain of Ser110 (Figure 54). In the case of MhpC, this has only normally been observed in structures where soaking ligands into the crystals was attempted or from native crystals, but is not evident in many structures obtained from co-crystallization of enzyme with ligand. Hence, in the 2.1Å structure, described in detail previously, no extra density was seen, but in other structures not only is this density seen but it also appears to vary in size. Previous work had revealed that this density was unlikely to be due to a water molecule or a metal ion co-ordinating to the serine residue (Dunn, 2000; Mohammed, 2001).

This extraneous density has also been observed in a number of other members of the α/β-hydrolase fold family of enzymes and has, in most cases, not been sufficiently accounted for. It has been proposed to be a highly-disordered water molecule, a metal ion, or possibly a small molecule from the crystallisation conditions covalently bound to the serine residue.

The structure of bromoperoxidase from *S. aureofaciens* (BPO-A2) has been resolved to 2Å and shown to have the α/β-hydrolase fold, with a catalytic triad of Ser98-Asp228-His257 (Hecht *et. al.*, 1994). Residual electron density at Ser98 was observed, which was taken to be due to a disordered acetate molecule, and not a water molecule, mimicking a tetrahedral or acyl-enzyme intermediate. Also, the recently solved structure of related C-C hydrolase BphD has unaccounted for electron density at the active site serine (Nandhagopal *et. al.*, 2001). The structure of this protein was determined to 2.4Å resolution and reveals that it too has the α/β-hydrolase fold with an identical catalytic triad to MhpC of Ser110-Asp235-His263. However, BphD has 285 amino acid residues per subunit. The extra density around Ser110 is claimed to be neither a water molecule nor a metal ion, but seems to represent a chemical modification to the residue of unknown origin. The structure of CumD that has recently been solved was that of the S103A mutant of the enzyme (Fushinobu *et. al.*, 2002) and so no evidence has been obtained to ascertain if this C-C hydrolase also has a covalent modification to its active site serine side-chain.

The crystal structures of carboxylesterases from *Alicyclobacillus acidocaldarius* (De Simone *et. al.*, 2000) and *Archaeoglobus fulgidus* (De Simone *et. al.*, 2001) have both revealed that these enzymes also have their catalytic serine residues covalently modified. In this case, the adduct is known to be a HEPES molecule derived from the crystallisation solutions. The sulphonyl group of the HEPES molecule is covalently bound to the serine, the arrangement around the sulphur atom being tetrahedral. One oxygen atom of the sulphonyl group is hydrogen-bonded to two water molecules whilst the other oxygen is bound in the oxyanion hole, formed by the backbone amide groups of Gly88, Gly89 and Ala161, the residue following the catalytic serine.

There is also the possibility that this density may be due to a post-translational modification of the serine residue. It has been suggested that the serine could have become phosphorylated, as this is one of the most common protein modifications, and serine kinases are the largest group of kinases. Yet, the putative phosphorylation sites for MhpC do not contain Ser110 in their motifs and occur in other regions of the polypeptide chain. The only modification site that includes this serine residue is the region GNSMGG which is a putative myristoylation site, as determined by the ScanProsite tool of the ExPASy database (Bairoch *et. al.*, 2003). Another group of enzymes with modified active site residues are the sulfatases. These are a group of enzymes with a highly conserved active site region including a Cα-formylglycine that is post-translationally generated by the oxidation of a conserved cysteine or serine residue (Boltes *et. al.*, 2001).

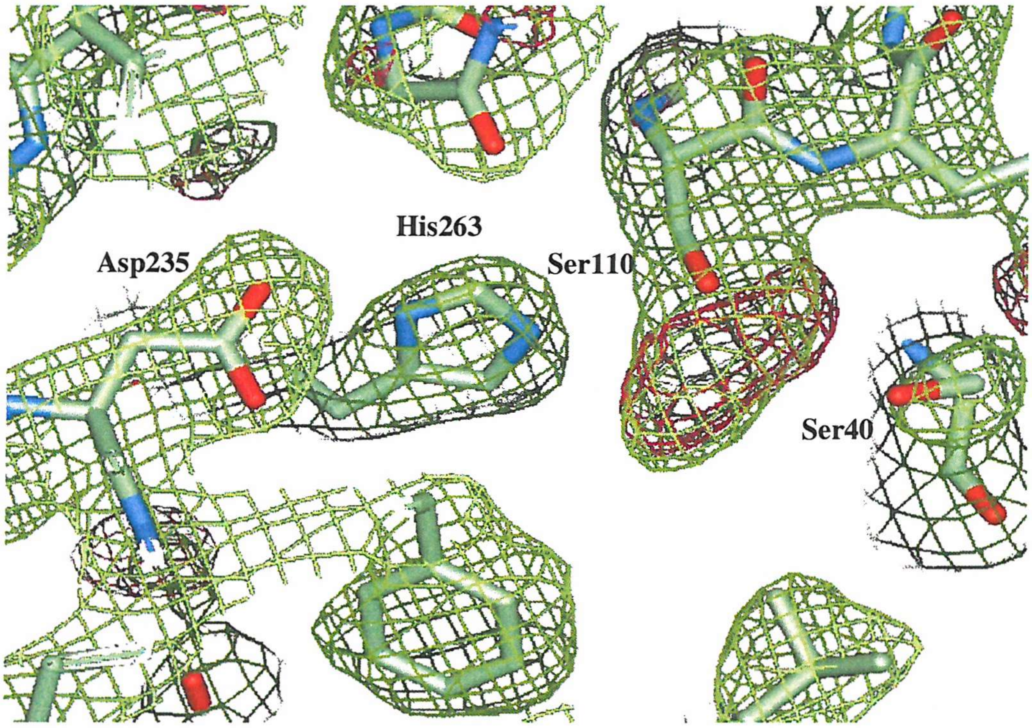


Figure 54: Active site of MhpC showing the catalytic triad of Ser110-His263-Asp235 and the unaccounted for electron density contiguous with the serine residue. The close proximity of Ser40 is also shown. *2Fo-Fc* density is contoured at the 1σ level (green) and *Fo-Fc* density is contoured at the 2σ level (red).

4.3.4. Inhibition of MhpC by thiol-specific reagents

It has been shown that the activity of MhpC is affected by thiol directed reagents (Lam & Bugg, 1997), and this has also been observed for a number of related enzymes, such as a diketoacid hydrolase from beef liver (Hsiang *et. al.*, 1972), fumarylacetoacetate hydrolase (Nagainis *et. al.*, 1981) and 2-hydroxymuconic semialdehyde hydrolase (XylF; Diaz & Timmis, 1995). In the case of MhpC only bulky reagents, such as *p*-hydroxymercuribenzoate, affect the activity, whilst smaller compounds, such as iodoacetate, hardly affect activity at all, with percentage activities falling to 32% and 92%, respectively, after incubation for one hour (Lam & Bugg, 1997). This pattern is the same as that seen for the diketoacid hydrolase from beef liver, as bulky sulfhydryl-specific inhibitors inactivate the enzyme, but iodoacetate does not. Also, the serine-specific inhibitors DFP and PMSF have no effect on this enzyme, much the same as has been observed for MhpC. However, in the case of XylF, the enzyme is inactivated by the serine-specific reagents PMSF and DFP, and also inactivated by the bulky thiol-modifying reagent *N*-ethylmaleimide, but the smaller iodoacetamide molecule does not significantly inhibit the enzyme activity (Diaz & Timmis, 1995). Therefore, as MhpC only contains three cysteine residues, and one of these, Cys261, is close to the active site triad histidine, this may account for the observed loss of activity. For instance, the bulky reagents could cause a conformational change that would move the catalytic His263 so that it was no longer within hydrogen-bonding distance of other residues of the catalytic triad or, alternatively, they might block access of the substrate to the active site. In the case of smaller reagents, such as iodoacetate, these factors may not be of such influence on the reactivity of the enzyme. Thus, this may also explain why some of the other serine hydrolases are also affected by sulfhydryl directed reagents, as this cysteine residue is very well conserved amongst many α/β-hydrolases (Figure 50). Further evidence that this residue is reactive has come from the structure obtained from an MhpC crystal following a thiomersal (a mercuryl compound) soak, in which Cys261 is covalently modified by this thiol specific reagent.

4.4. Attempts to identify the unexplained density at Ser110

4.4.1. Sequencing of the *mhpC* gene

The gene encoding MhpC was sequenced to determine if Ser110 of the catalytic triad had in someway become mutated. It was suggested that this may be the cause of the unexplained extra density on the side-chain of this residue.

The sequencing was done in three stages as the *mhpC* gene is quite large (~900bp), and after only a few hundred bases are sequenced then errors are likely to occur, such as those caused by frame-shifting of the polymerase. Therefore, the terminal regions were sequenced, as well as the central portion of the gene, to confirm not only the whole sequence of the *mhpC* gene, but also the nucleotide regions between the MhpC encoding section and the *BamH1* and *Xba1* restriction sites.

DNA sequencing was performed by the dideoxy chain termination method by Oswel Research Ltd. (Southampton, UK) using plasmid pIPC and three oligonucleotide primers: one to sequence the C-terminus, one to sequence the N-terminus, and the other to sequence the remainder of the gene between the terminal regions, with a degree of overlap between all three regions.

The full sequence for *mhpC* was determined and it was shown that the codon for the residue corresponding to Ser110 does indeed encode a serine residue (codon TCG), and that therefore the unexplained density is not due to a mutation in this position. The sequencing also showed that Ser40, the putative fourth residue of the catalytic tetrad, is also encoded by a primordial serine codon (codon TCC; Krem & Di Cera, 2001), a feature described in greater depth in Chapter 6.2.3.4.

4.4.2. Tryptic digest of MhpC and analysis by Mass Spectrometry

As sequencing had revealed that the active site serine had not become mutated, then it was decided to determine if the extra density for this residue could be due to a covalent modification. This could be either a post-translational modification or a modification caused by the reaction of the serine residue with a small molecule from the crystallisation conditions. It had been suggested that the covalent modification could be a phosphate group, sulphate group, or even an ethylene glycol molecule from the crystallisation conditions. Structures of related enzymes have been solved which also have their active site serine modified, and this modification has been assigned to many molecules, as described in section 4.3.3.4.

A tryptic digest was carried out on MhpC to generate a number of peptide fragments which were then separated by reverse-phase high pressure liquid chromatography (RP-HPLC) (Figure 55). The molecular weight of these peptides could then be determined by mass spectrometry and compared with those generated *in silico* (Table 14). The serine of interest occurs in a peptide that is quite large

(~2000 Da.) and does not contain a tyrosine or tryptophan residue. Therefore, it seemed reasonable to assume that peptides that absorbed well at 280nm and that eluted earlier, ie. where small in size, would probably not be of interest. The residue numbering is that of the mature MhpC protein (289 residues) and not that of the structural protein (284 residues), in which the residue numbering was altered due to disorder at the *N*-terminus. If the peptide fragment corresponding to that containing Ser110 had a mass quite different from its predicted mass, then the size of the covalent modification could be determined. Hence, this could lead to the possible identification of the molecule responsible for the extra density observed in the native structure of MhpC. However, this proved to be difficult by collecting each individual peak from the HPLC separation and determining the molecular weight of each peptide by mass spectrometry.

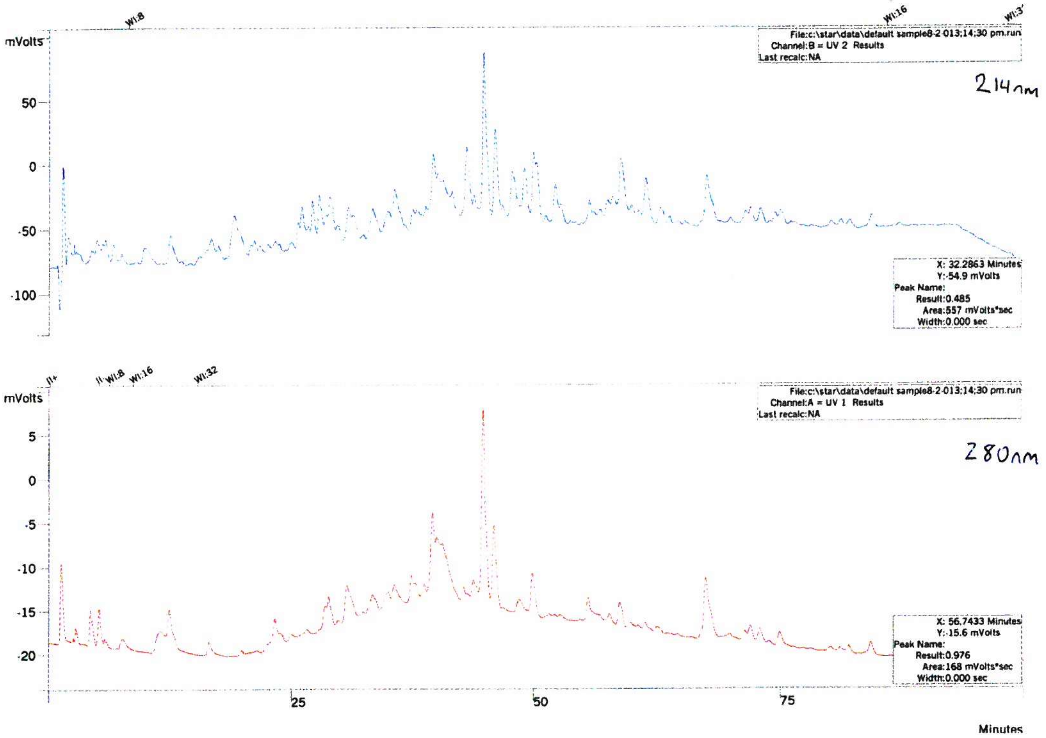


Figure 55: HPLC chromatogram of the tryptic digest of MhpC recorded at 214nm and 280nm.

Position	Peptide sequence	Mass (Da.)
1-4	MQEK	535.254
5-18	MMS**Y**QPQTEAATSR	1600.709
19-27	FLNVEEAGK	1006.520
31-61	IHFNDCGQGDETVVLLHGSGPGATG**W**ANFSR	3242.507
62-72	NIDPLVEAG**Y**R	1246.643
73-83	VILLDCPG**W**GK	1200.644
84-93	SDSVVNSGSR	1007.475
94-99	SDLNAR	675.342
103-112	SVVDQLDIAK	1087.599
113-131	IHLLGN**S**MGGHSSVAFTLK	1969.032
132-135	**W**PER	587.294
139-161	LVLMGGGTGGMSLFTPMPTEGIK	2294.159
163-168	LNQL**Y**R	806.452
169-176	QPTIENLK	942.525
177-197	LMMDIFVFDTSDLTDALFEAR	2450.162
198-204	LNNMLSR	847.445
206-213	DHLENFVK	1001.505
214-220	SLEANPK	758.404
221-228	QFPDFGPR	963.468
229-233	LAEIK	573.361
234-242	AQTLIV**W**GR	1043.600
246-254	FVPMDAGLR	1005.519
255-268	LLSGIAGSELHIFR	1512.853
269-293	DCGH**W**AQ**W**EHADAFNQLVLNFLARP	2938.384

Table 14: Peptide fragments and masses generated *in silico*. The predicted peptides are those obtained assuming complete digestion by trypsin. Aromatic residues (ie. those that absorb at 280nm) are highlighted in bold. The active site serine (Ser110) is in bold and underlined.

Instead, LCMS analysis of the tryptic digest was carried out, which combines both HPLC separation of the peptides and then mass spectrometry in one automated process (Figure 56). It was hoped that this would enable more complete and precise molecular mass measurements of all the peptide fragments generated. However, despite being able to identify a large number of the peptides generated, in comparison to those generated *in silico*, it still proved impossible to identify the peptide containing the serine residue of interest, and hence determine if it had been modified (Table 15). This difficulty in interpretation of the data could be due to both incomplete digestion of the protein with trypsin, and the quandary of separating the smaller fragments generated by HPLC, due to their similarities in molecular weight. There was a general trend that larger peptides were eluted last, but it is difficult to say with any surety that the peptides determined are absolutely correct due to many more peaks in the mass spectrums being present. Of course, the peptide containing the serine residue could have even remained in the insoluble fraction following the digest, and so may not even have been injected onto the HPLC column.

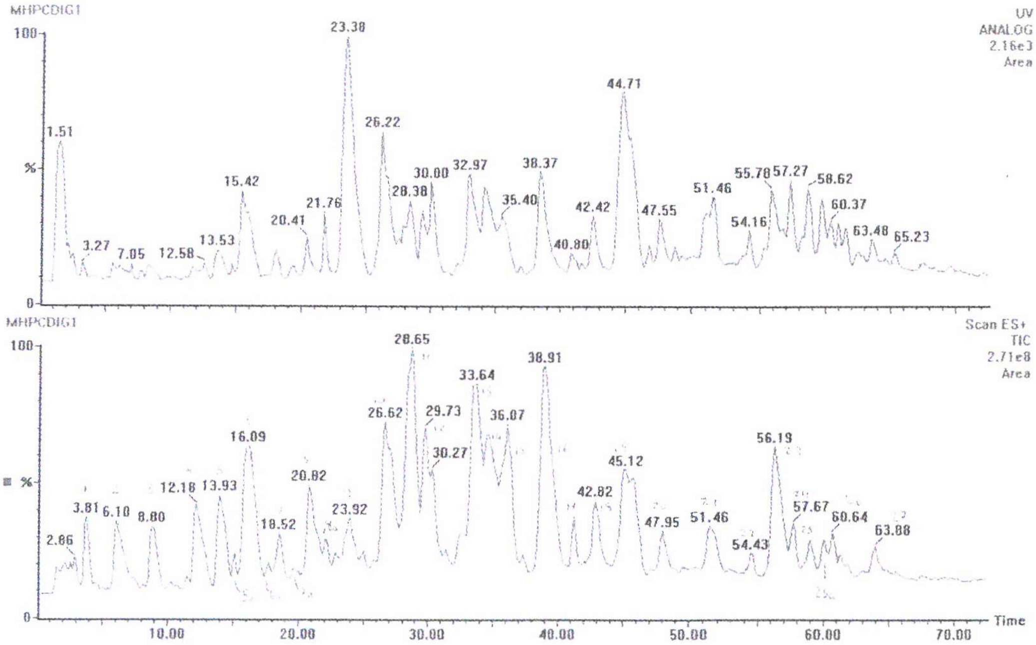


Figure 56: LCMS analysis of an MhpC tryptic digest. The top graph shows the elution profile of peptides separated by HPLC. The bottom graph is the total ion count (TIC) for each of the corresponding HPLC peaks after entering the mass spectrometer.

Peak number	Major mass(es) (Da.)	Minor mass(es) (Da.)	Probable residues	Sequence
1	303.38 [M+H]	ND	136-138	VGK
2	361.47 [M+4H]	721.55 [M+2H]	100-112	ILKSVVDQLDIAK
3	389.48 [M+H]	ND	28-30	TLR
4	373.54 [M+H]	ND	100-102	ILK
5	675.54 [M+H]	338.40 [M+2H]	94-99	SDLNAR
6	758.62 [M+H]	379.94 [M+2H]	214-220	SLEANPK
7	536.56 [M+H]	ND	1-4	MQEK
8	573.63 [M+H]	ND	229-233	LAEIK
9	587.51 [M+H]	ND	132-135	WPER
10	806.68 [M+H]	404.09 [M+2H]	163-168	LNQLYR
11	424.50 [M+2H]	847.62 [M+H]	198-204	LNNMLSR
12	ND	ND	ND	ND
13	1006.77 [M+H]	503.95 [M+2H]	19-27	FLNVEEAGK
14	501.54 [M+2H]	1001.70 [M+H]	206-213	DHLENFVK
15	579.67 [M+2H] 386.94 [M+3H]	1157.71 [M+H]	205-213	RDHLENFVK
16	544.65 [M+2H]	1087.79 [M+H]	103-112	SVVDQLDIAK
17	549.12 [M+4H]	1096.85 [M+2H]	73-93	VILLDCPGWGKSDSVVNSGSR
18	624.10 [M+2H]	1246.82 [M+H]	62-72	NIDPLVEAGYR
19	522.67 [M+2H]	1044.80 [M+H]	234-242	AQTLIVWGR
20	661.17	802.57	ND	ND
21	814.29	543.32	ND	ND
22	601.28 [M+2H]	1200.94 [M+H]	73-83	VILLDCPGWGK
23	757.29 [M+2H] 505.40 [M+3H]	1514.05 [M+H]	255-268	LLSGIAGSELHIFR
24	769.61	1187.53	ND	ND
25	832.16 [M+2H]	1663.06 [M+H]	84-99	SDSVVNSGSRSDLNAR
26	850.88 [M+2H]	567.71 [M+3H]	214-228	SLEANPKQFPDFGPR
27	1147.93 [M+2H]	ND	139-161	LVLMGGGTGGMSLFTPMPTEGIK

Table 15: Peptide fragments from LCMS analysis of a tryptic digest of MhpC. ND = Not determined.

4.4.3. ^{31}P Nuclear Magnetic Resonance (NMR) spectroscopy of MhpC

Phosphorus, often in the form of phosphate, is covalently bound to a number of naturally occurring phosphoproteins and phosphorylated enzyme intermediates. Amino acids commonly phophorylated include serine, threonine, tyrosine and aspartate *via* an oxygen and histidine *via* nitrogen. ^{31}P nuclei occur at 100% natural abundance and are an NMR active, spin ½ species with relatively high sensitivity. ^{31}P NMR spectroscopy is therefore an excellent tool for studying these proteins (Matheis & Whitaker, 1984).

Since it had been suggested that the unexplained density observed for the catalytic serine side-chain could be due to phosphorylation, NMR was utilised to investigate whether this could be the case. The ^{31}P NMR spectrum of a solution of MhpC, at pH7.0, indicates the presence of a phosphate signal at 2.37 ppm relative to an external 85% H_3PO_4 standard (Figure 57). Under identical conditions the ^{31}P spectrum of a solution of KH_2PO_4 lacking MhpC yielded a peak at 3.18ppm. Differing chemical shifts in NMR spectra is indicative of nuclei in different chemical environments and these results therefore suggest the phosphorus signal observed in the protein sample is not a trace amount of contaminating free phosphate (KH_2PO_4) but another phosphate species. Both ^{31}P signals, however, fall in the range of frequencies observed for free phosphate (and phospho-proteins) which is dependent on pH. In order to eliminate any pH uncertainty, more conclusive evidence for the phosphate signal in the protein sample being that of *protein-bound* phosphate was obtained. This was achieved by looking at the half-height linewidth of the signals after processing both datasets under identical conditions and with the application of 5Hz line broadening to enhance signal to noise. Large molecules tumble more slowly in solution which leads to faster relaxation of the NMR signal and an increase in the linewidth of peaks. Free phosphate would therefore give narrower ^{31}P linewidths than protein bound phosphate. This is found to be the case with half-height linewidths of 5Hz and 12Hz measured for the KH_2PO_4 and MhpC samples respectively.

Of course, it cannot be stated with any certainty that it is Ser110 which is phosphorylated. MhpC has many more accessible residues which could have been modified. However, electron density maps show no sign of modified residue side-chains other than Ser110. Residues not normally observed in electron density maps include those at the N- or C-termini due to their inherent flexibility which leads to the possibility that such a residue in MhpC has been altered. The *N*-terminus pentapeptide, missing in almost all electron density maps, does in fact contain a serine residue so this cannot be excluded as a possible modification site.

At present the results of the ^{31}P NMR experiments are inconclusive as to whether the Ser110 modification is due to phosphorylation as, in some instances, the extra density is not present at all and, in other cases, the size of the density varies. Nevertheless, the presence of a protein phosphate signal does seem to suggest that MhpC may be phosphorylated, if only irregularly and quite possibly not at the site of the Ser110 side-chain.

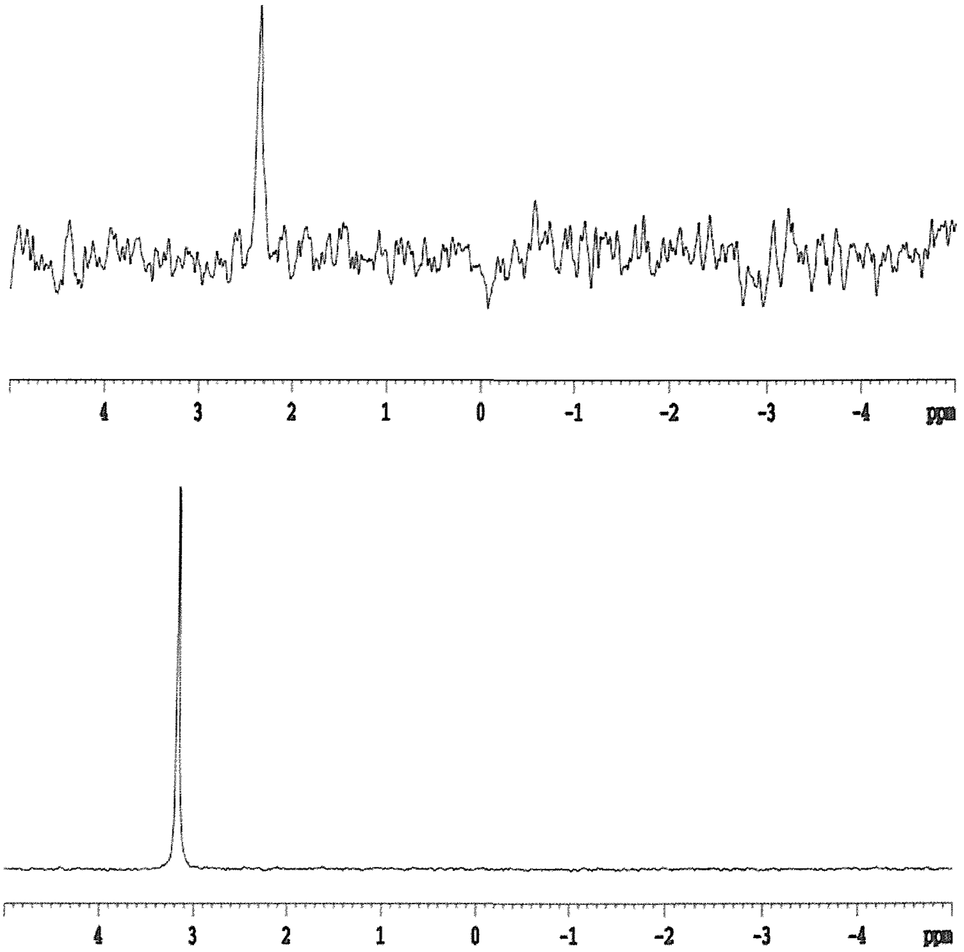


Figure 57: ^{31}P NMR spectra of the MhpC sample (top) and the control (free KH_2PO_4) below. Half height linewidths for the control and experiment were 5Hz and 12Hz respectively. Spectra were processed with 5Hz line broadening applied.

CHAPTER 5

STRUCTURES OF MhpC COMPLEXED WITH PRODUCT ANALOGUES

5.1. MhpC co-crystallised with Laevulinic acid

5.1.1. Data collection and processing of the MhpC/Laevulinic acid co-crystal

Data collection from a single crystal was carried out on ESRF beamline ID14-2 (Grenoble, France) at 100K using a Mar CCD detector. The crystal diffracted to 3.4 Å, and a total of 180 1°-oscillation frames were measured with a crystal to detector distance of 250 mm. The data were processed with MOSFLM and scaled and merged using programs of the CCP4 suite (data processing statistics shown in Table 16). At the start of data processing the poor quality of the diffraction data made determination of space group and cell dimensions difficult. After initially processing the data in space groups P2, P3 and P6 the crystal was found to belong to trigonal space group $P3_221$ with unit cell dimensions of a=b=266.28 Å, c=79.71 Å (γ=120°).

Total number of reflections	84709
Number of unique reflections	38628
Resolution (Å)	3.4
Completeness (%)	86.7
R_{merge} (%) (outer resolution shell)	15.1 (20)
Multiplicity	2.5
Average $I/\sigma I$ (outer resolution shell)	3.0 (2.2)

Table 16: Data processing statistics for the MhpC/Laevulinic acid co-crystal.

5.1.2. Structural determination and refinement of the MhpC/Laevulinic acid co-crystal

5.1.2.1. Determination of structure by molecular replacement

The structure of MhpC co-crystallised with laevulinic acid was determined using the CCP4 program MOLREP with the native MAD structure as the phasing model (cf. chapter 4.2.1.). A solvent content of 72% (V_m = 4.39 Å^3/Da) was calculated assuming there are six molecules of MhpC in the asymmetric unit of the co-crystal. The high solvent content of the crystal could explain why the crystal diffracted so poorly.

The cross rotation calculations were performed, using an MhpC dimer (as described in chapter 4.3.2.1.) as search model, with reflections in the 39 to 3.4 Å resolution range. The rotation function yielded several significant peaks with a radius of integration of 30 Å (Table 17). Using either a single monomer or a tetramer as search model gave five protomers and four protomers respectively. These appeared to be incomplete solutions when the crystal packing was viewed using MOLPACK, whereas the best solution seemed to be that obtained with the dimer as search model.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	88.57	49.24	272.98	7.03
2	31.43	130.76	92.98	7.03
3	32.06	0.00	32.06	6.98
4	0.00	180.00	4.56	6.98
5	56.44	180.00	0.00	4.29
6	3.56	0.00	0.00	4.29

Table 17: The six highest peaks following cross rotation calculations performed with MOLREP using data between 39 and 3.4 Å.

The translation search, calculated in space group $P3_221$, gave three good peaks, one for each dimer in the asymmetric unit (Tables 18, 19 and 20). Thus, this indicated that the correct enantiomorph had been chosen as a translation search using space group $P3_121$ gave no peaks.

Peak Number	X	Y	Z	Dens/σ
1	0.411	0.547	0.322	18.91
2	0.410	0.548	0.352	10.95
3	0.411	0.547	0.293	10.85
4	0.742	0.213	0.184	9.78
5	0.411	0.548	0.066	9.48

Table 18: The five highest translation function peaks calculated in space group $P3_221$ using the second rotation function solution.

Peak Number	X	Y	Z	Dens/σ
1	0.805	0.524	0.181	34.07
2	0.805	0.524	0.124	14.49
3	0.804	0.524	0.240	11.55
4	0.598	0.664	0.584	10.97
5	0.450	0.864	0.462	10.75

Table 19: The five highest non-crystallographic translation function peaks calculated in space group $P3_221$ after fixing the position of the first MhpC dimer.

Peak Number	X	Y	Z	Dens/σ
1	0.195	0.065	0.751	57.19
2	0.195	0.066	0.689	18.78
3	0.195	0.065	0.250	18.37
4	0.194	0.065	0.815	18.08
5	0.196	0.058	0.700	15.10

Table 20: The five highest non-crystallographic translation function peaks calculated after fixing the position of the second MhpC dimer.

When viewed using MOLPACK it can be seen that there are no inter-subunit clashes in a well ordered unit cell which contains six tetramers (as described in chapter 4.3.2.4.) and another six dimers (Figure 58). It seems likely that these dimers are part of tetramers with the other dimer being within the adjacent unit cell.

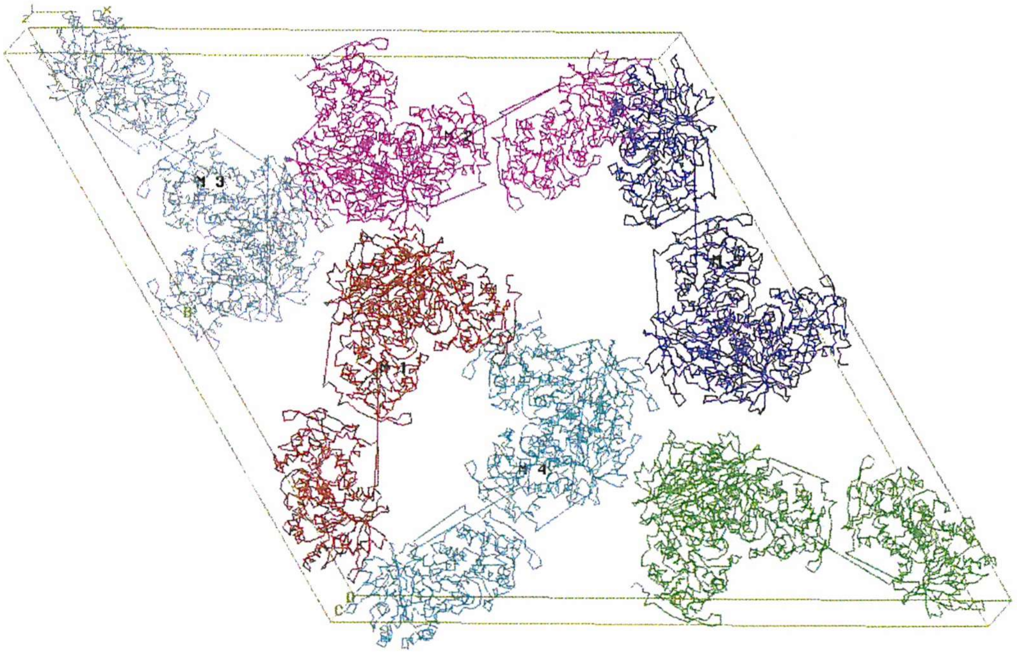


Figure 58: Crystal packing of MhpC subunits in the MhpC/Laevulinic acid co-crystal generated using the MOLPACK program of the CCP4 suite. There are six protomers per asymmetric unit and 36 per unit cell due to trigonal space group $P3_221$. The search model used was a dimer and hence three dimers per asymmetric unit were found by molecular replacement.

5.1.2.2. Refinement of the MhpC/Laevulinic acid co-crystal structure

The model was refined with programs of the CNS suite using 38483 reflections in the resolution range 39 to 3.4 Å. A test set of reflections (5% of data) was used for R_{free} calculations. The model was subject to an initial round of refinement involving rigid body refinement, simulated annealing and positional refinement. This reduced the R_{factor} from 39.5% (R_{free} = 41%) to 29% (R_{free} = 31.5%). The resulting model was used to calculate sigmaA weighted electron density maps. Map inspection and model building was carried out using QUANTA. Following initial examination of the electron density maps *Fo-Fc* density was observed at the active site which was consistent with that for one molecule of laevulinic acid (Figure 59). After building in a model of laevulinic acid to the difference density and following further refinement, involving rigid body refinement, simulated annealing and domain B-factor refinement (each protomer treated as a separate domain), then *2Fo-Fc* density was observed for the ligand. During rebuilding, residues with no electron density for their side-chains were changed to alanine and regions of the model in poor agreement with the density were altered. No solvent atoms were incorporated into the model as there were too few reflections recorded to control the refinement of so many atoms. Current refinement statistics are shown in Table 21.

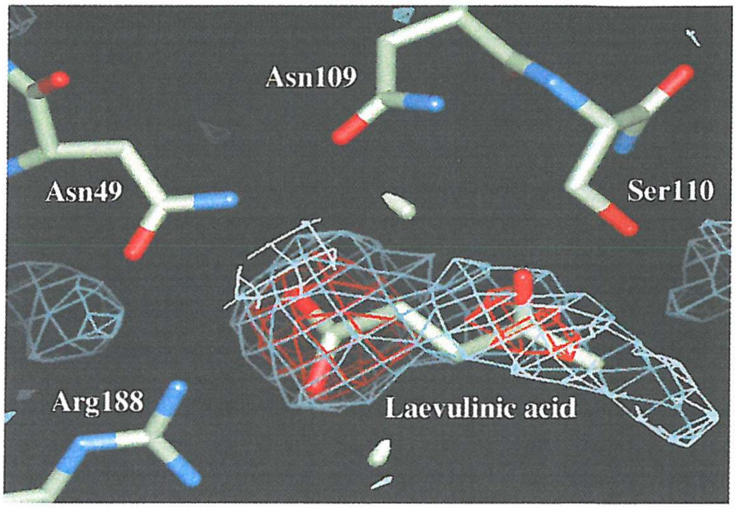


Figure 59: Model of laevulinic acid built into the difference density at the active site of MhpC. The carboxylate group is coordinated to Arg188 and Asn49 with the carbonyl group then being close to Ser110. The *Fo-Fc* density is contoured at 2σ (blue) and 3σ (red).

Resolution range (Å)	39.2 – 3.4
R_{factor} (%)	24.7
R_{free} (%)	28.0
Number of reflections in working set	36553
Number of reflections in test set	1930
Number of protein atoms	13080
RMS bond length deviation (Å)	0.008
RMS bond angle deviation (°)	1.2
Overall average B-factor (Å^2)	37.1

Table 21: Refinement statistics for the MhpC/Laevulinic acid co-crystal structure.

5.1.3. Co-ordination of laevulinic acid within the active site of MhpC

Laevulinic acid, which mimics the succinyl moiety of the RFP, coordinates to Arg188 and Asn49 through its carboxylate group. These two residues are almost completely conserved amongst members of this family of enzyme (see Figure 50), and so could also be involved in coordination of substrates at their active sites. The carbonyl group of the ligand is close to the serine residue of the putative catalytic triad, which may be an important feature for the catalytic process of the enzyme. This ligand occupies one end of the active site cavity, the end which is most deeply buried within the subunit (Figure 60), and so a longer ligand should indicate more fully the full binding features of the active site.

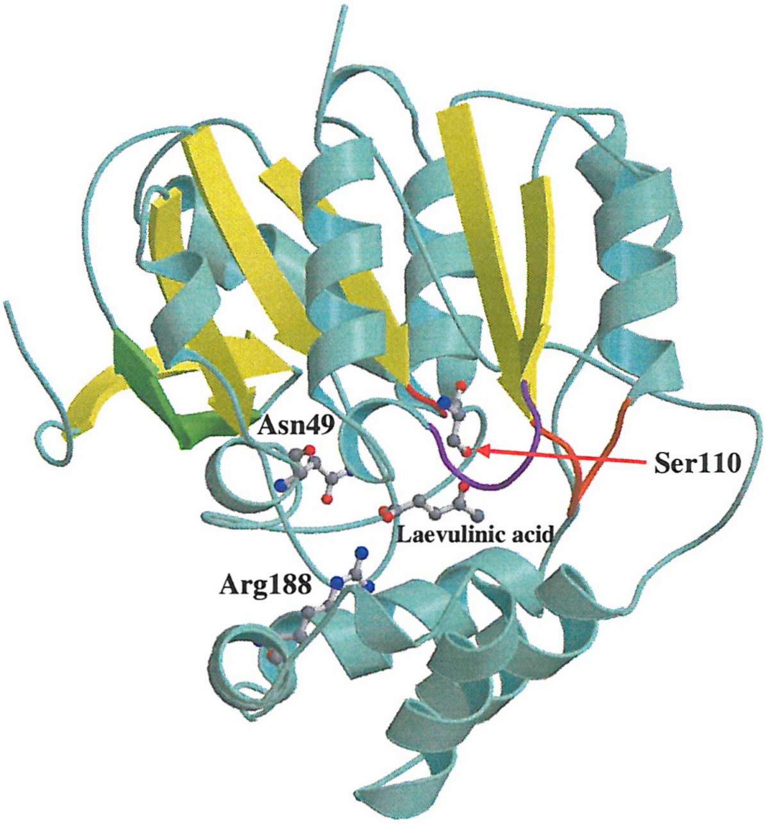


Figure 60: Co-ordination of laevulinic acid at the active site of an MhpC protomer. Residues shown in ball-and-stick are Arg188, Asn49 and Ser110. The laevulinic acid molecule can be seen to coordinate to Arg188 and Asn49 *via* its carboxylate group. The carbonyl group of the ligand is close to the catalytic serine residue. The colouring scheme is the same as that used in Figure 35.

It is interesting to note that the laevulinic acid molecule seems to be directed towards this end of the active site. This molecule mimics one of the end products, succinic acid, of the MhpC catalysed reaction and comprises the propionate side-chain portion of the substrate for MhpB, which is then cleaved to give the RFP. This part of the substrate is that which varies most amongst substrates of this family of C-C hydrolases (Figure 61). The 2-hydroxy-6-oxohexa-2,4-dienoate (HODA) functionality of these *meta*-cleavage compounds remains the same for all of the substrates of the C-C hydrolases from all of the different catabolic pathways. So, in the case of BphD (Furukawa & Miyazaki, 1986), the equivalent product to the succinic acid produced by MhpC would be benzoic acid, whereas for other single phenol degradative pathway enzymes the molecule is normally a short chain acid, such as acetic acid for TodF (Menn *et. al.*, 1991), isobutyric acid for CumD (Eaton *et. al.*, 1998) and propionic acid for EtbD (Yamada *et. al.*, 1998).

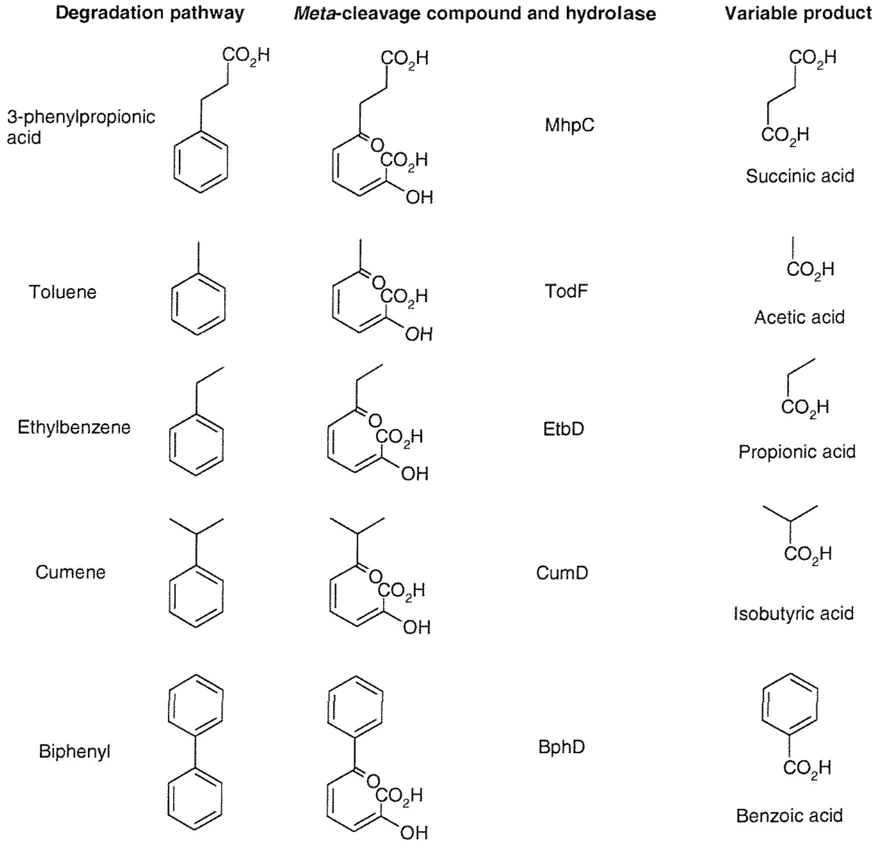


Figure 61: Comparison of various degradative pathways and their products. The MhpC product succinic acid is distinct from the corresponding products of the other hydrolase enzymes in that it has two carboxylate groups.

Succinic acid differs from these in that it contains two carboxylate groups, whereas these other compounds only contain one, with their carboxylate group being that which contains what would have been the C-6 carbonyl of the substrate. Therefore, the active site of these enzymes would require more hydrophobic residues to bind the rest of the molecule, whereas succinic acid would need residues capable of co-ordinating to its other carboxylate group. Thus, there is the ambiguity as to which way around laevulinic acid should be put into the electron density at the active site. As the resolution of this structure is only 3.4Å then the density is not that indicative of atoms absolute positions and so can only give a rough guide to the orientation of the laevulinic acid molecule. Nonetheless, by having the carboxylate group co-ordinated to Arg188 and Asn49, as illustrated in Figure 62A, this would seem to make sense as then the carbonyl group, equivalent to the C-6 carbonyl of the substrate, would be in a position closer to the catalytic triad of the enzyme. Both terminal amine groups of the Arg188 guanidinium are within hydrogen-bonding distance of one of the oxygen atoms of the laevulinic acid carboxylate (Arg188 NH1 – Lev O1 = 2.7Å and Arg188 NH2 – Lev O1 = 3.2Å). The other oxygen atom of the carboxylate is also within H-bonding distance of one of the Arg188 side-chain amine groups (Arg188 NH2 – Lev O2 = 3.4Å). This oxygen atom is also within H-bonding range of both polar atoms of the Asn49 side-chain (Asn49 ND2 – Lev O2 = 2.7Å and Asn49 OD1 – Lev O2 = 3.4Å). The backbone amide groups of residues Gly39 and Gly41 also make interactions with both polar groups of the laevulinic acid carboxylate (Gly39 N – Lev O2 = 3.5Å and Gly41 N – Lev O1 = 3.2Å).

Significantly, it appears that the active site serine residue has swung away somewhat, as the distance between Ser110 OG and His263 NE2 is about 4.4Å, much too great for a hydrogen bond to form. The ligand has forced its way in between these two residues, as shown in Figure 62B. In fact, the carbonyl group of the ligand is close to Ser110 OG (2.9Å) and extremely close to His263 NE2 (2.2Å), although given the resolution of the data then there is some likelihood of variability in these hydrogen-bonding distances. The distance between Asp235 and His263 of the catalytic triad remains in very good hydrogen-bonding range (2.6Å). The putative fourth residue of the catalytic tetrad, Ser40, remains close to the active site serine residue, the distance between their side-chain hydroxyl groups being in the region of 3.6Å.

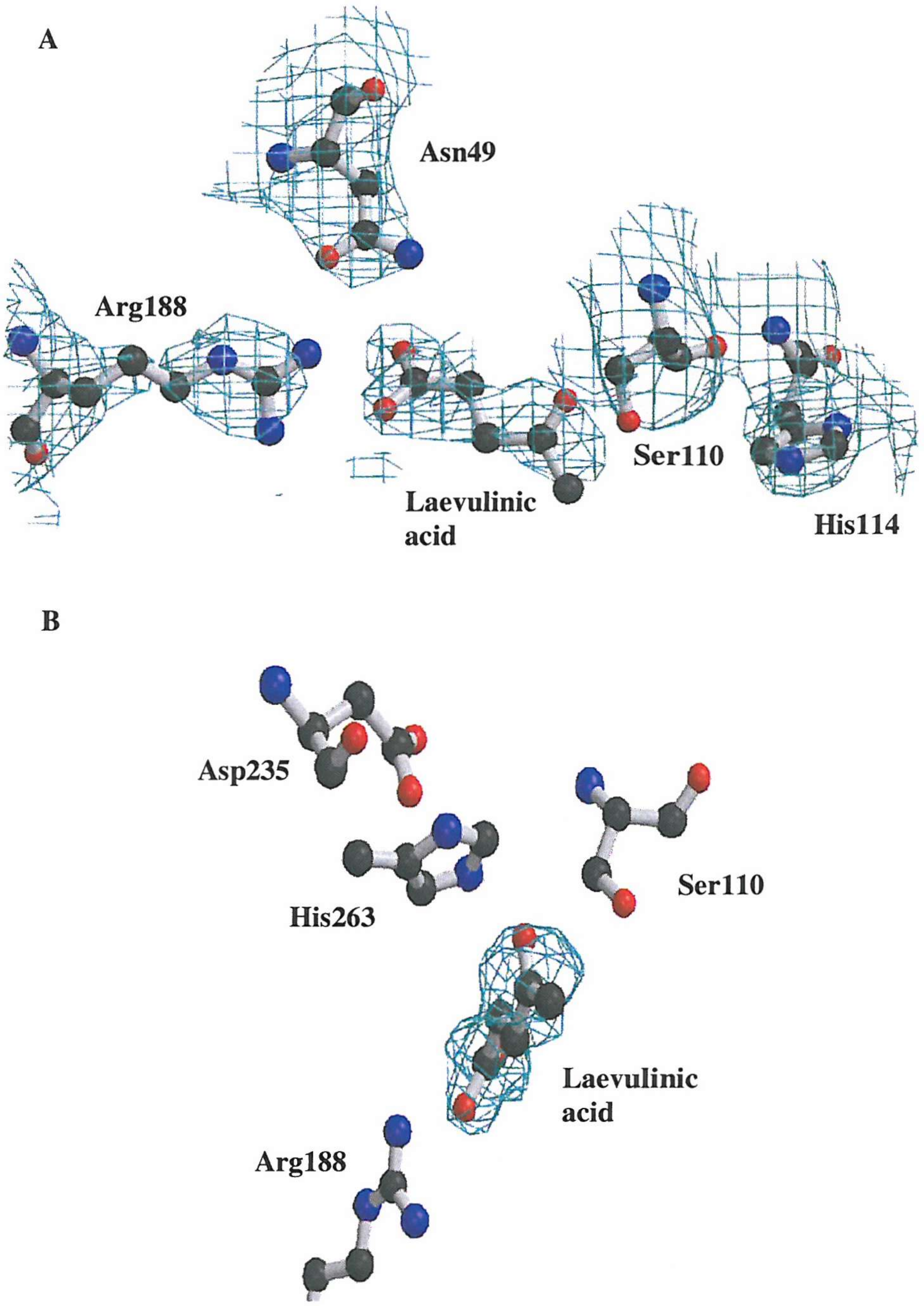


Figure 62: (A) Close-up of ligand binding at active site. The carboxylate group coordinates to Arg188 and Asn49, with the carbonyl group then coming close to the active site serine, Ser110. His114 is also shown. **(B)** Coordination of laevulinic acid shown from another viewpoint, such that it can be seen that the ligand bisects the His263 and Ser110 residues of the catalytic triad. Asp235 of the triad is also shown, as well as Arg188. The *2Fo-Fc* density in both pictures is contoured at 1σ.

5.2. MhpC co-crystallised with 4,6-dioxoheptanoic acid (DOHA)

5.2.1. Data collection and processing of an MhpC/DOHA co-crystal

Data collection from a single crystal was carried out on SRS Daresbury beamline 14.2 (Warrington, UK) at 100K using an ADSC detector. The crystal diffracted to 2.5Å and a total of 164 1°-oscillation frames were measured with a crystal to detector distance of 240mm. The data were processed with MOSFLM and scaled and merged using programs of the CCP4 suite. The resolution of the data was cut back to 2.65Å during processing due to loss of quality of diffraction towards the end of data collection. Data processing statistics are shown in Table 22. The crystal was found to be isomorphous with the crystal that was used to solve the MAD structure of MhpC (Dunn, 2000). That is, it was in monoclinic spacegroup $P2_1$ with unit cell dimensions of a=62.81Å, b=154.11Å, c=80.03Å (β=91.23°).

Total number of reflections	105618
Number of unique reflections	39857
Resolution (Å)	2.65
Completeness (%)	90.5
R_{merge} (%) (outer resolution shell)	10.9 (49.2)
Multiplicity	2.8
Average I/σI (outer resolution shell)	4.0 (1.3)

Table 22: Data processing statistics for MhpC/DOHA co-crystal.

5.2.2. Structural determination and refinement of the MhpC/DOHA co-crystal

5.2.2.1. Determination of structure by molecular replacement

Despite the MhpC/DOHA co-crystal being isomorphous with the MAD structure of MhpC, the technique of molecular replacement was still utilised to solve the structure. The CCP4 program MOLREP was used with one protomer of the native MAD structure as the phasing model. A solvent content of 60.4% (V_m = 3.13 Å^3/Da) was calculated assuming there are four molecules of MhpC in the asymmetric unit of the co-crystal. This was thought to be very likely as the MAD structure also had four protomers per asymmetric unit (Dunn, 2000).

The cross rotation calculations were performed using reflections in the 15.9 to 3.0 Å resolution range with a radius of integration of 25Å. This gave four good peaks, indicating that each MhpC molecule in the asymmetric unit almost certainly had a separate rotation (Table 23).

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	113.17	0.00	73.17	10.05
2	92.59	48.35	270.15	9.75
3	265.04	47.61	274.64	8.36
4	168.89	2.24	191.40	7.94
5	336.46	90.00	130.64	3.73
6	206.78	34.95	281.74	3.56
7	25.34	88.63	309.78	3.56
8	240.79	14.72	238.44	3.49

Table 23: The eight highest peaks following cross rotation calculations performed with MOLREP using data between 15.9 and 3.0 Å.

This was shown to be the case as the translation search, calculated in space group $P2_1$, gave one good peak for each rotation peak (Tables 24, 25, 26 and 27).

Peak Number	X	Y	Z	Dens/σ
1	0.181	0.000	0.371	13.69
2	0.108	0.000	0.006	3.89
3	0.124	0.000	0.088	3.21
4	0.401	0.000	0.052	3.13
5	0.174	0.000	0.478	2.79

Table 24: The five highest translation function peaks, calculated in space group $P2_1$, using the second rotation function solution.

Peak Number	X	Y	Z	Dens/σ
1	0.839	0.760	0.484	28.19
2	0.840	0.790	0.484	10.66
3	0.839	0.373	0.984	10.60
4	0.839	0.157	0.485	10.56
5	0.838	0.730	0.485	10.53

Table 25: The five highest non-crystallographic translation function peaks of the third rotation function calculated after fixing the position of the first MhpC molecule.

Peak Number	X	Y	Z	Dens/σ
1	0.608	0.199	0.210	43.48
2	0.609	0.182	0.710	10.66
3	0.610	0.900	0.710	10.18
4	0.608	0.132	0.711	10.04
5	0.608	0.560	0.209	10.04

Table 26: The five highest translation function peaks calculated for the first rotation function solution after fixing the position of two MhpC molecules.

Peak Number	X	Y	Z	Dens/σ
1	0.380	0.573	0.310	52.59
2	0.383	0.650	0.311	10.21
3	0.381	0.607	0.310	9.33
4	0.379	0.957	0.810	8.90
5	0.876	0.930	0.811	8.70

Table 27: The five highest translation function peaks calculated for the fourth rotation function solution after fixing the position of three MhpC molecules.

When viewed using MOLPACK it can be seen that the four protomers found per asymmetric unit form the tetrameric arrangement typical of the MhpC crystalline structure. There are two tetramers per unit cell, as for the $P2_1$ spacegroup there are two asymmetric units per unit cell, and hence eight MhpC protomers altogether (Figure 63).

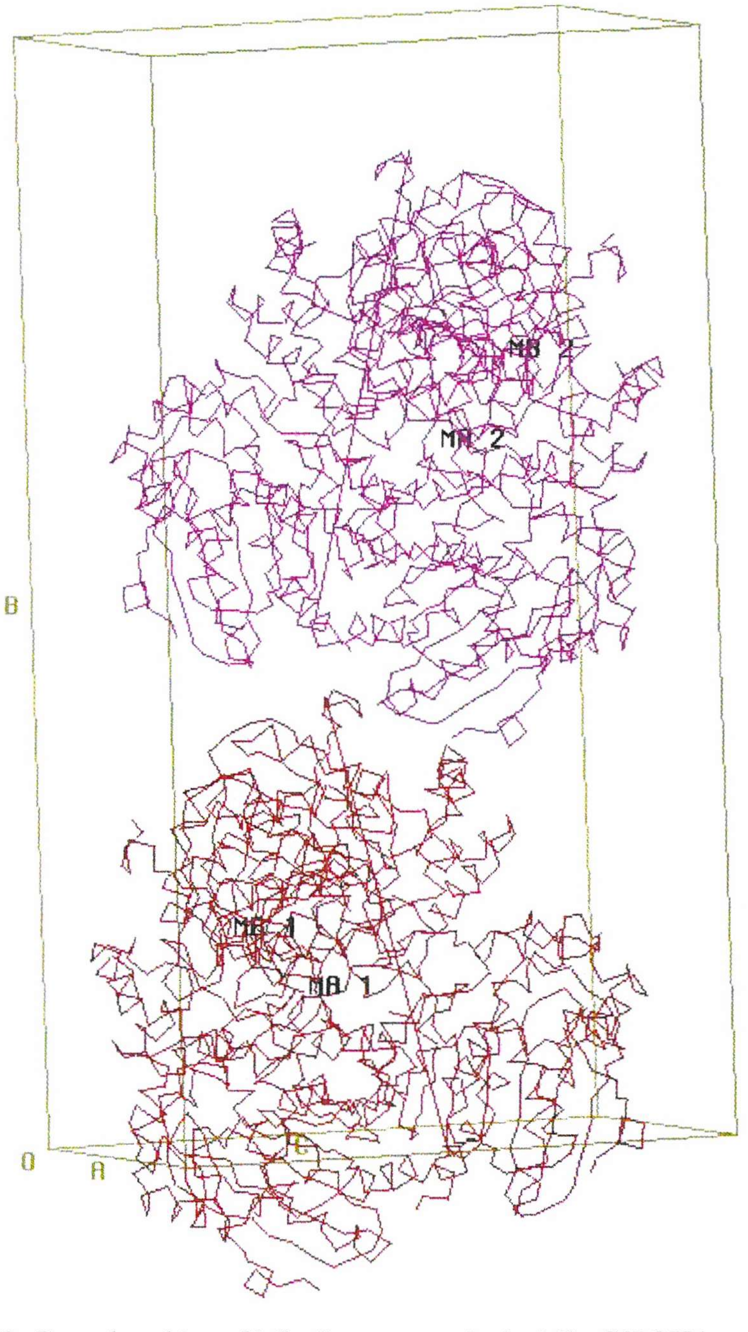


Figure 63: Crystal packing of MhpC protomers in the MhpC/DOHA co-crystal generated using the MOLPACK program of the CCP4 suite. There are four subunits per asymmetric unit and two asymmetric units per unit cell due to the monoclinic spacegroup, $P2_1$, of this crystal form of MhpC.

5.2.2.2. Refinement of the MhpC/DOHA co-crystal structure

The model was refined with programs of the CNS suite using 39675 reflections in the resolution range 20-2.65 Å. A test set of reflections (5% of data) was used for R_{free} calculations. The model was subject to an initial round of refinement involving rigid body refinement, positional refinement and grouped B-factor refinement. NCS restraints were used during this initial round of refinement and the R_{factor} was reduced from 34.6% (R_{free} = 34.7%) to 26.4% (R_{free} = 29.9%). The resulting model was used to calculate sigmaA weighted electron density maps. Map inspection and model building was carried out using QUANTA. Following initial examination of the electron density maps *Fo-Fc* density was observed at the active site which was considered to be that for one molecule of the co-crystallised ligand (Figure 64).

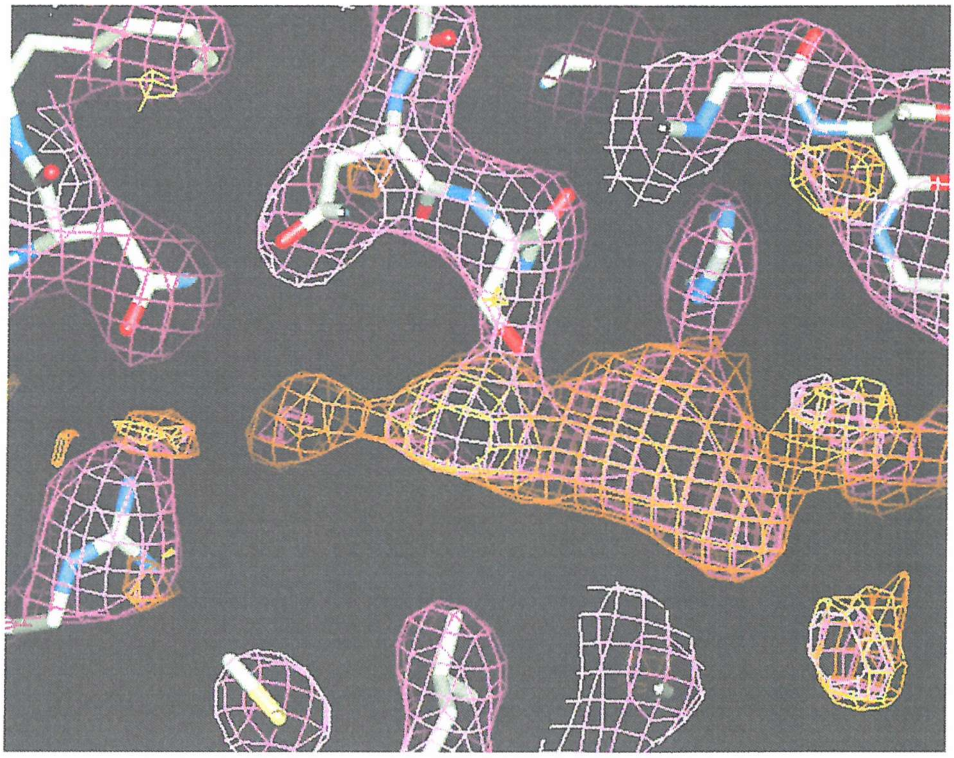


Figure 64: Electron density observed at the active site of the MhpC/DOHA co-crystal which could be that for one molecule of the DOHA ligand. *2Fo-Fc* (purple) density is contoured at 1.0 σ and *Fo-Fc* (orange) density is contoured at 2.0 σ.

5.2.2.3. Re-refinement of the MhpC/DOHA structure with varying NCS restraints

Following difficulties in determining how the DOHA ligand bound at the active site of MhpC, due to the differences in *Fo-Fc* density in the various subunits of the tetramer, then the technique used to solve the MhpC/DKNDA co-crystal structure was attempted (cf. chapter 6.2.2.2.1.).

Initially NCS restraints were used (see section 5.2.2.2. above) in which all four subunits in the asymmetric unit were grouped together during refinement. However, it was thought that as MhpC is a dimer in solution, and that it had been proposed from the MhpC/DKNDA structure that there might be half-site occupancy of the dimer, then one subunit from each dimer could have a different local environment to the other. Therefore, it was decided to redo the refinement with two NCS restraint groups and that one subunit from each dimer would go into each separate group. In this structure it was found that subunits A and D formed one dimer (as described in section 4.3.2.1), with the other dimer of the tetramer being formed by protomers B and C.

The first attempt at refinement using this altered NCS restraint arrangement had subunits A and D in one restraint group and subunits B and C in the other group. Only residues 21 to 284 of each subunit were used in NCS restraints as this had helped previously in refinement and was thought to be due to the greater flexibility of the *N*-terminus affecting the refinement. Refinement programs of the CNS suite were used and consisted of rigid body refinement, simulated annealing, positional refinement and grouped B-factor refinement. The R_{factor} was initially 33.1% (R_{free} = 33.7%) and ended as 24.7% (R_{free} = 29.2%). SigmaA weighted electron density maps were made and visualised, with the model, as before using QUANTA. No dissimilarity in the *Fo-Fc* electron density maps of the four subunits active sites could be seen as compared to the previous refinement using all four subunits as one NCS restraint group.

However, it was realised that the two different NCS groups used actually matched the dimeric arrangement and that this would not in reality show any difference between mismatched subunits if the half-site occupancy theory was correct. In actual fact, one subunit from each dimer should be paired with the same subunit from the other dimer to form a restraint group. Therefore, subunits A and B were paired together and subunits C and D were paired as the other NCS restraint group. This was due to A and B sharing a similar crystal packing environment, as did C and D. For instance, subunits C and D both had symmetry-related molecules packed against helix αF of the lid domain (residues 181-195), which is a normally fairly flexible region of the MhpC structure. Subunits A and B did not have this feature, but both did appear to have a slightly more stable *N*-terminal region, as extra *2Fo-Fc* and *Fo-Fc* electron density was observed, after the previous attempts at refinement, that could be due to some residues of the 'missing' pentapeptide (cf. chapter 4.3.1.).

Refinement was carried out as before using these new NCS restraint groups. Before rigid body refinement the R_{factor} was 33.1% (R_{free} = 33.7%), and following simulated annealing, positional refinement and grouped B-factor refinement the R_{factor} was 23.6% (R_{free} = 28.1%). These values were slightly improved on those obtained using the previous two different NCS restraint procedures. Electron density maps were calculated as before and converted to X-PLOR format. Following examination in QUANTA, along with the model, it could still be seen that there were some discrepancies between subunits of the *Fo-Fc* electron density at the active sites.

It was hoped that some manual model-building and further rounds of refinement may help to improve this density. Therefore, some residues that lacked good *2Fo-Fc* density for their side-chains were changed to alanine residues and regions of the model in poor agreement with the electron density were altered to fit better. The current refinement statistics are shown in Table 28.

Resolution range (Å)	20 – 2.65
R_{factor} (%)	22.8
R_{free} (%)	28.0
Number of reflections in working set	37685
Number of reflections in test set	1990
Number of protein atoms	8688
RMS bond length deviations (Å)	0.008
RMS bond angle deviations (°)	1.39
Overall average B-factor (Å^2)	34.2

Table 28: Current refinement statistics of the MhpC/DOHA co-crystal structure.

5.2.3. Ligand binding within the MhpC/DOHA co-crystal active site

Despite the refinement being more successful when using two NCS restraint groups the electron density at the active site was still difficult to interpret. There was definitely *Fo-Fc* density within the active site cavity in all four subunits, and this did vary between the two groups. However, it was not apparent how the DOHA ligand could fit within this density. Nonetheless, what was revealed by this structure is that Arg188, Asn49 and His114 are all still involved in coordination of the ligand and thus could also possibly be involved in coordination and catalytic turnover of the natural substrate. In the following section, when making comparisons, subunits A and B are from NCS restraint group one and subunits C and D from NCS restraint group two. Nevertheless, it should again be pointed out that subunits A and D form one MhpC dimer and subunits B and C form the other dimer.

5.2.3.1. Comparison of ligand binding between the two NCS restraint groups

In the MhpC/DOHA co-crystal structure it has been observed that a tryptophan residue, Trp264, moves considerably to come into contact with the ligand density in the active sites of subunits C and D. In the other two subunits Trp264 stays in the same position as it does in the native structure, where it packs against another tryptophan residue, Trp267 (Figure 65). This residue in turn interacts with the guanidinium group of an arginine residue (Arg52), an interaction which has also been observed in other structures, as it is well-founded that aromatic rings can act as hydrogen-bond acceptors (Levitt & Perutz, 1988). The relative distance between the aromatic side-chains of the two tryptophan residues is in the region of 4Å, which is not unusual for a van der Waals hydrophobic stacking interaction. However, the distance between the indole group of Trp267 and the guanidinium of Arg52 is approximately 3.4Å. This is rather close for this type of interaction but could be explained by the tryptophan residue having a slightly higher B-factor than residues neighbouring it. Therefore, it may actually be further from the arginine than shown except that this is difficult to determine due to the volume of the *2Fo-Fc* density encompassing it. There is also a hydrogen-bonding network between these three residues. One terminal amine group of the arginine residue is hydrogen-bonded to the amide carbonyl of Trp267 (Arg52 NH1 to Trp267 O = 2.9Å), and the amide nitrogen of this tryptophan is also hydrogen-bonded, in this case to the backbone carbonyl oxygen of Trp264 (Trp267 N to Trp264 O = 2.8Å).

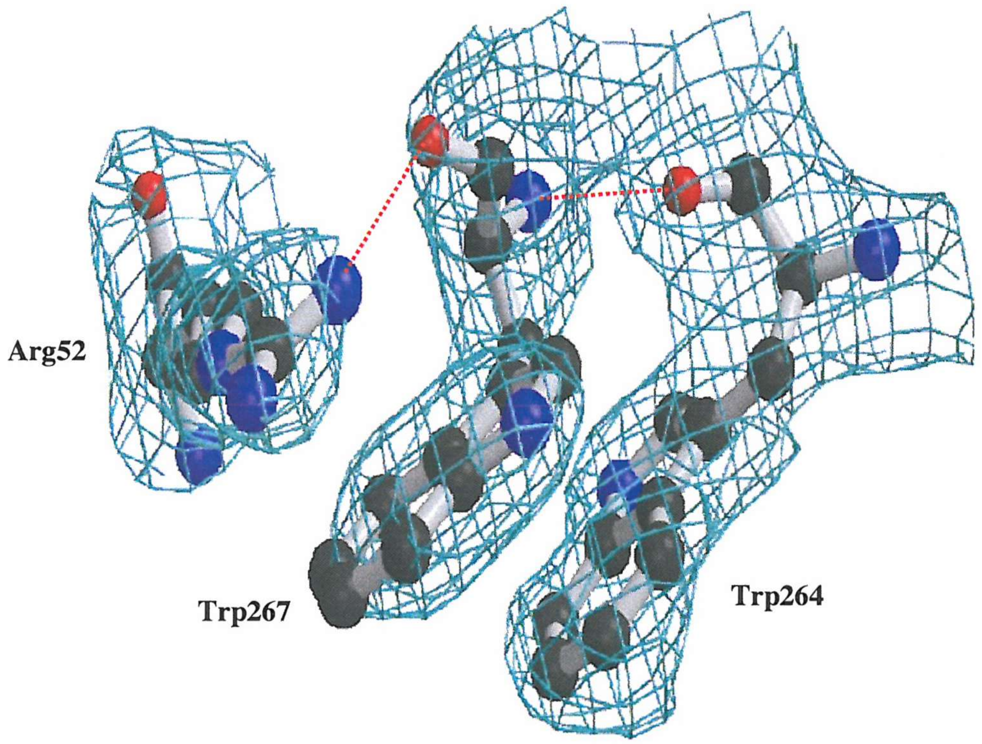


Figure 65: Hydrophobic stacking interaction between Trp264 and Trp267 in subunits A and B of the MhpC/DOHA co-crystal structure. The indole group of Trp267 also interacts with the guanidinium group of Arg52. This three residue stacking is also seen in the native structure of MhpC. The *2Fo-Fc* electron density is contoured at 1σ. Hydrogen-bonding interactions are indicated by dotted red lines.

A histidine residue, His114, occupies two different conformations depending on which NCS restraint group it has been refined in. In subunits A and B it stays pointing towards the active site cavity. Whereas, in subunits C and D it has moved away from the active site cavity. This dual occupancy has been observed before (cf. chapter 6.2.3.2.) and the flexibility of this residue's side-chain could possibly have mechanistic implications. This histidine residue may even be coordinated to a ligand in subunits A and B, although the *Fo-Fc* density in these two subunits active sites is particularly difficult to interpret as it varies between the two. The better density is seen in subunit A (Figure 66). It appears that there is a rather large amount of difference density, possibly too much for a pair of water molecules, towards the His114 end of the active site cavity. There is one spherical section of density close to His114, and another ball of density close to the putative active site serine residue, Ser110. The *Fo-Fc* density in these spheres can be contoured up to 5σ, indicative of a rather electron dense molecule, and hence possibly not a water molecule. However, this density is not unambiguous enough to facilitate model-building of the DOHA ligand into it using the X-BUILD feature of QUANTA. One nitrogen atom of the imidazole ring of His114 points towards the putative ligand density in subunit A, the distance to the middle of the 5σ *Fo-Fc* density from His114 NE2 being about 3.6Å. It appears that the putative active site nucleophile, Ser110, may be covalently bound to this ligand as there is continuous density running from this residues side-chain to the other sphere of density in the active site.

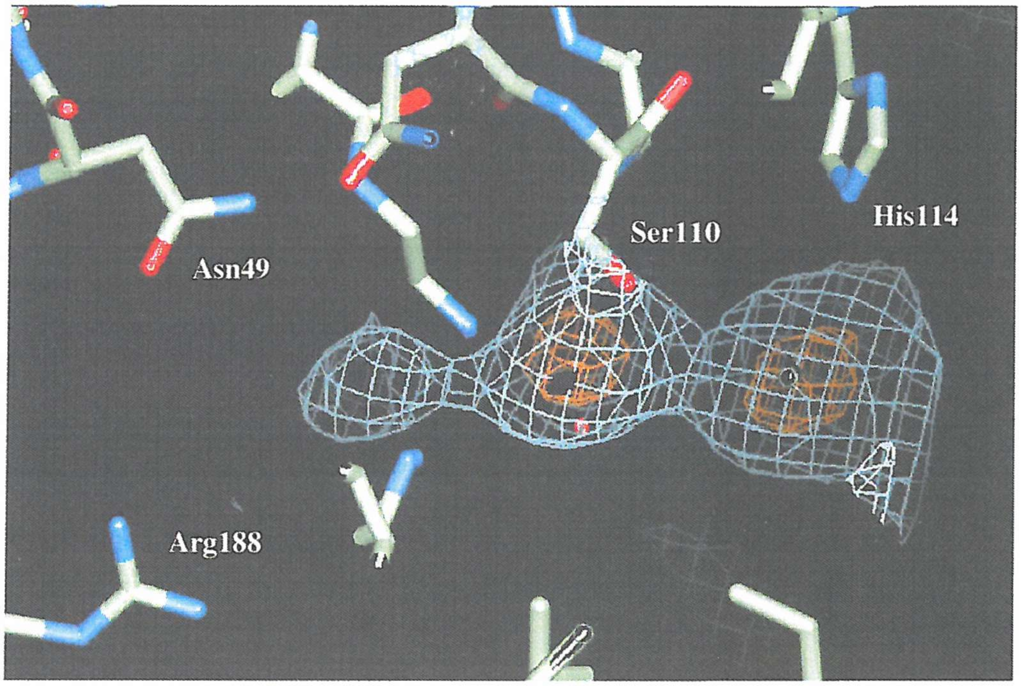


Figure 66: Active site of subunit A in the MhpC/DOHA co-crystal structure. The *Fo-Fc* electron density within the active site may contain a ligand which could coordinate to His114 and Ser110. Thus, the histidine is in the conformation in which it points towards the active site. Blue density is contoured at 2σ, whereas gold density is contoured at 5σ.

Despite there being obvious similarities between the *Fo-Fc* density observed for subunits C and D, as compared to that for subunits A and B, there are also some major differences. In subunits C and D the bulk of the difference density, as well as an ample amount of *2Fo-Fc* density, is mostly at the buried Arg188 end of the active site cavity (Figure 67), in direct contrast to subunits A and B. Again, it appears to be in two bulges, which are interconnected. These two bulges of density have fairly high sigma values, of greater than 5σ, with the interconnecting density having a lower sigma value of between 2 and 3σ. This would seem to indicate that the density is possibly not due to two water molecules being present at this section of the active site as in general water molecules tend to have lower sigma values in difference density maps.

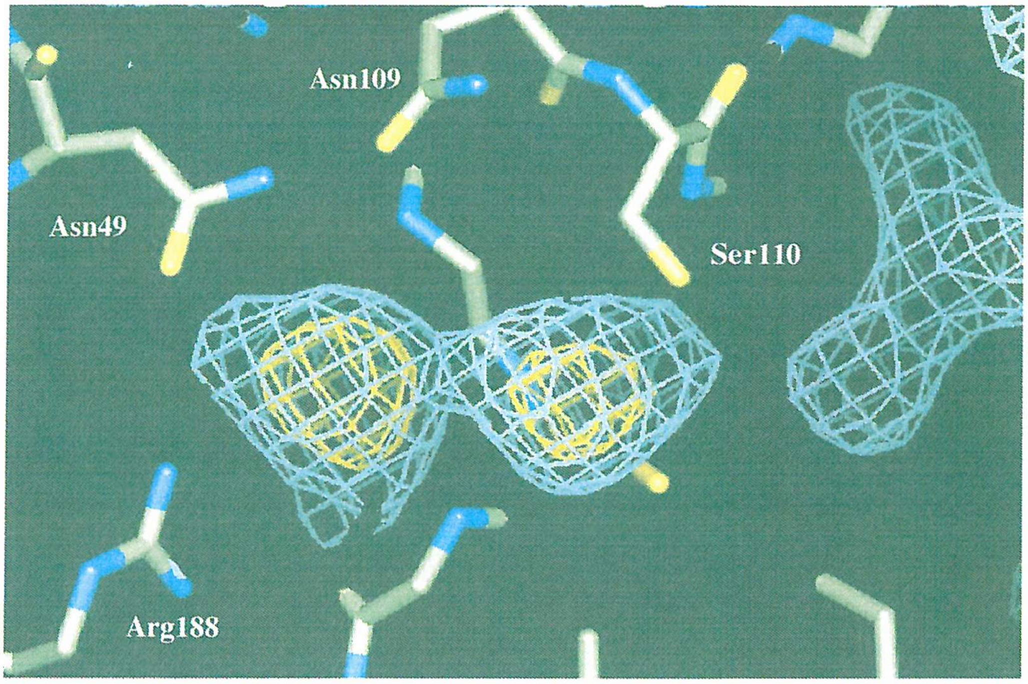


Figure 67: Active site of subunit D in the MhpC/DOHA co-crystal structure. The view is practically identical to that in Figure 66 (above). In this case, however, it can be seen that the bulk of the *Fo-Fc* density is towards the other end of the active site cavity. Thus, the proposed ligand could coordinate with Arg188 and Asn49, and may also interact with the active site serine residue. His114 is in the other conformation away from the active site. Contour levels are as depicted in Figure 66.

It also appears that the end of the *2Fo-Fc* density runs contiguously with that of the Ser110 side-chain (not shown) and so could indicate that the ligand may be covalently bound to the protein at this residue. This is not chemically unreasonable, as one of the carbonyl groups of the DOHA molecule could be attacked by the hydroxyl of the serine side-chain to form a hemi-ketal compound (as in Figure 68). This would seem quite reasonable if it were not for the fact that the density appears to be too short and spherical for this ligand, whereas the DOHA molecule is quite linear and so would be expected to occupy an elongated segment of electron density.

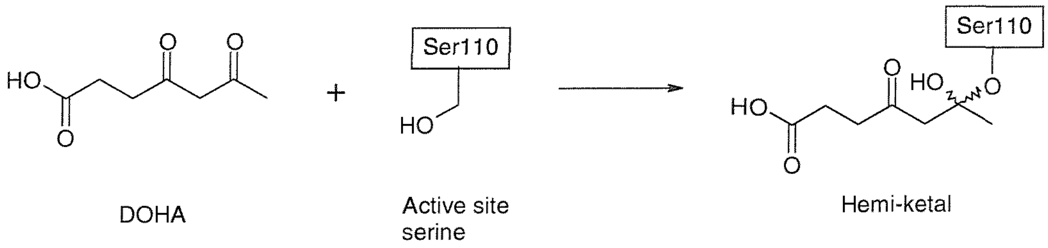


Figure 68: Possible nucleophilic attack of the Ser110 side-chain hydroxyl on the C6 carbonyl of DOHA to form a hemi-ketal adduct. It is also possible that the serine hydroxyl group could attack the other carbonyl group of the ligand.

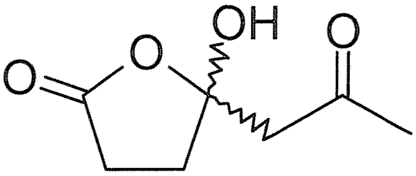


Figure 69: Cyclic *ortho*-ester of 4,6-dioxoheptanoic acid.

On the other hand, if the ligand had become cyclised, as shown in Figure 69, then this may possibly account for the difference density more precisely. This is not totally without precedent, as this compound may actually exist in the cyclic form more readily than in the open chain form under both acidic and, possibly even, basic conditions (Akhtar, 2003). The five-membered ring of this heterocyclic *ortho*-ester could occupy the larger sphere of density that is adjacent to Arg188, and the acetyl functional group could occupy the smaller sphere of density connected to this, and also may form a covalent adduct with the active site serine residue, Ser110. Thus, the tryptophan residue (Trp264) could have moved from its favourable stacking interaction with the other tryptophan, as seen in subunits A and B, so that the side-chain amine group hydrogen-bonds with one of the hydroxyl groups of the cyclised DOHA molecule (Figure 70). The Arg188 side-chain could then coordinate to the polar groups of the ligand, along with Asn49, such as seen for the MhpC/Laevulinic acid co-crystal structure. The Trp264 NE1 group is approximately 4.1Å from the centre of the larger ball of positive electron difference density, and in this new conformation the six-membered ring of the indole group can form a favourable stacking interaction with the side-chain of Phe173 (not shown).

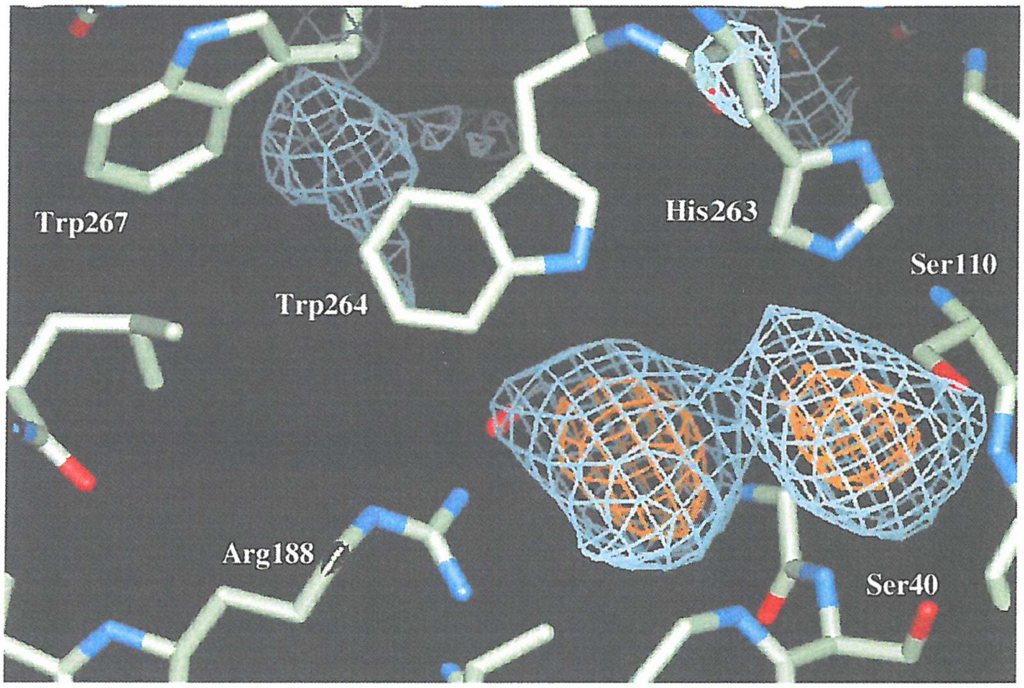


Figure 70: Interaction of Trp264 with a possible ligand in the *Fo-Fc* density at the active site of subunit D. This same interaction occurs for the equivalent tryptophan in subunit C, but is not seen for those in the other two subunits where this residue has a different conformation (see Figure 65). The indole amine group could form a hydrogen bond with a polar group of the suspected cyclic DOHA molecule.

It is interesting to note from the Ramachandran plot of this structure that the active site serine residues in subunits A and B are in a disallowed region, as is the case for the native enzyme and most other α/β-hydrolase fold enzymes, whereas this serine residue is in a generously allowed region of the plot for subunits C and D. This may indicate that the formation of a covalent adduct between the amino acid and the ligand has resulted in a more favourable geometry. In respect to the active site triad residues, then the Asp235 to His263 hydrogen-bonding distances are still respectable in all subunits (ranging from 2.5Å in subunits C and D to 2.7Å in subunits A and B), whereas the His263 to Ser110 distances do not vary between the NCS restraint groups but are all too long for a good hydrogen bond to form (average value of 3.95Å). The distance that appears to vary most between the NCS groups is that from Ser40 OG to Ser110 OG; in subunits A and B it is 4.4Å, whilst in subunits C and D it is 4.7Å. This difference could be attributed to the ligand forming an adduct with Ser110 in subunits C and D, so forcing the other serine away due to steric hindrance.

The arginine residue, Arg188, which appears to be coordinated to the ligand in subunits C and D seems to have multiple conformations in the other two protomers (Figure 71). The density for the side-chain is very well defined for this residue in subunits C and D as supposedly the ligand-to-side-chain interaction 'locks' this residue in place. However, the dual conformation for the side-chain of this residue in the other two subunits may be due to a number of reasons. Firstly, that there is a lack of ligand at this site in these protomers, in itself lending credence to the half-site reactivity hypothesis. Secondly, that there is also a lack of a symmetry-related molecule, which could help to anchor the helix which this residue resides in, and hence causing it to have greater flexibility. This appears to be evinced by the higher B-factors for this residue in subunits A and B (average value = 70Å^2) as compared to the temperature factors for the equivalent atoms in subunits C and D (average value = 60Å^2). So, for NCS equivalence group one the arginine residue not only points towards the active site, as it does for NCS restraint group two, but also away from the active site. In one orientation, towards the active site, the arginine can hydrogen bond with a proline residue (Arg188 NH1 to Pro42 O = 2.65Å), whilst in the other orientation, away from the active site, the same functional group hydrogen-bonds with the adjacent glycine residue on the same loop which contains the proline (Arg188 NH1 to Gly43 O = 3.0Å). This loop is well conserved amongst this class of enzyme, as mentioned previously (cf. chapter 4.3.3.1.).

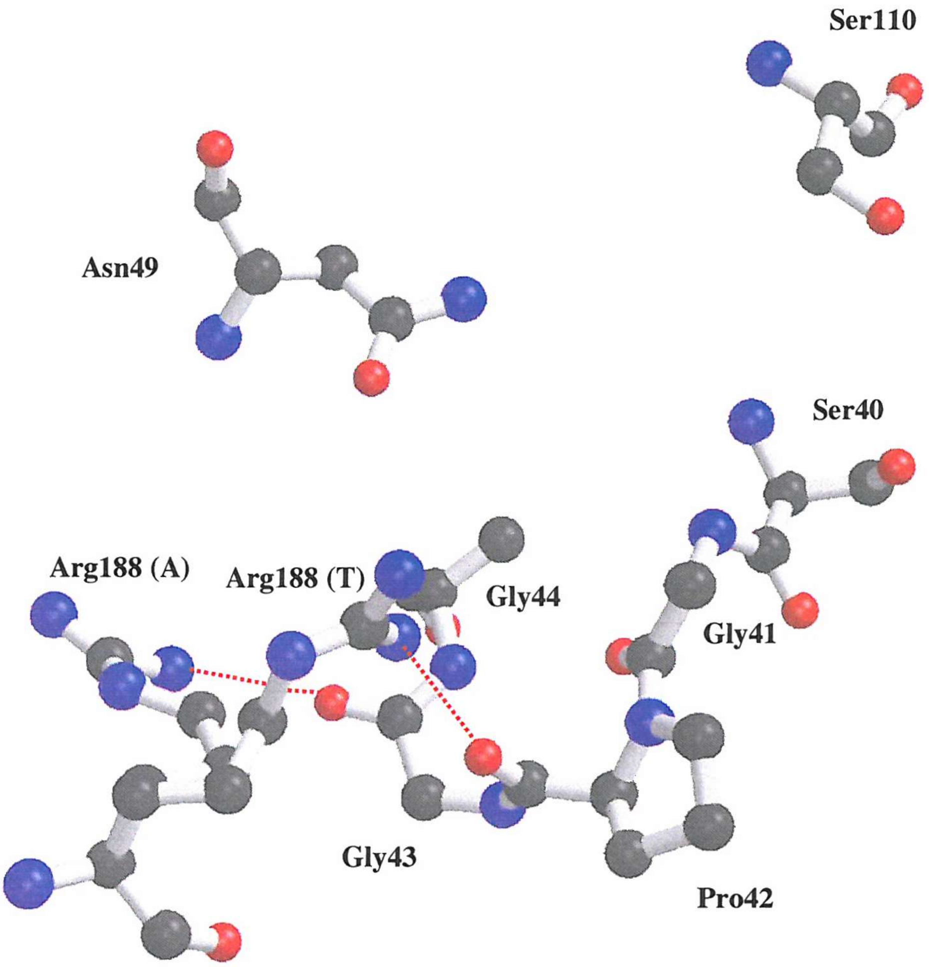


Figure 71: Alternative conformation seen for Arg188 in subunits A and B. The arginine side-chain appears to interact with either Gly43 or Pro42 depending on whether it is pointing towards that active site, Arg188 (T), or away from the active site, Arg188 (A). The positions of Asn49 and Ser110 are also shown to give perspective to the orientations of these residues. Hydrogen bonding of the arginine NH1 groups is shown by red dotted lines.

The aforementioned differences in conformation between Arg188 and Trp264, and their varying subunit dispersion, are not entirely without literature precedent. It has recently been noted that other members of this family of C-C hydrolases also have discrepancies in these residues' positions. The crystal structure of *Janthinobacterium* sp. strain J3 2-hydroxy-6-oxo-6-(2'-aminophenyl)-hexa-2,4-dienoate hydrolase (CarC$_{J3}$), an enzyme involved in carbazole degradation, was determined at 1.86Å resolution (Habe *et. al.*, 2003). This homodimeric enzyme, like BphD and CumD, is a homologue of MhpC and was found to possess the α/β-hydrolase fold. The structures of these three serine hydrolases have been compared and it was found that the side-chain directions of Arg185 (equivalent to MhpC Arg188) and Trp262 (in MhpC: Trp264) in the substrate binding pockets of these enzymes were different, suggesting that these residues may make a conformational change during the catalytic cycle (Habe *et. al.*, 2003). This tallies nicely with what has been reported in this subchapter on the co-crystal structure of MhpC with DOHA. Namely, that not only do these two residues make quite considerable conformational changes, but that there is also a subunit inconsistency between these changes, hence lending more weight to the idea that some form of half-site occupancy could be involved. That is, that MhpC may be affected such that only one subunit of each dimer is catalytically active. This feature will be discussed in greater depth in the section on the structure of MhpC co-crystallised with the non-hydrolysable substrate mimic 2,6-diketo-nona-1,9-dioic acid (DKNDA).

CHAPTER 6

STRUCTURES OF MhpC COMPLEXED WITH SUBSTRATE ANALOGUES

6.1. MhpC co-crystallised with Azelaic acid

6.1.1. Data collection and processing of the MhpC/Azelaic acid co-crystal

Data collection from a single crystal was carried out 'in-house' using an ENRAF Nonius generator, running at 100mA and 50kV, and a MAR image plate detector. The crystal was vitrified with an Oxford Instruments cryostream at 100K. The crystal diffracted to 2.5 Å resolution and a total of 240 1°-oscillation frames were measured with an exposure time of 20 minutes per frame and a crystal to detector distance of 211 mm. The diffraction data were processed with MOSFLM and scaled and merged using programs of the CCP4 suite (data processing statistics are shown in Table 29). The crystal was found to belong to the monoclinic space group C2, with unit cell dimensions of a=99.68 Å, b=113.46 Å, c=66.16 Å (β=131.25°).

Total number of reflections	83485
Number of unique reflections	18221
Resolution (Å)	2.5
Completeness (%)	98.4
R_{merge} (%) (outer resolution shell)	6.0 (18.9)
Multiplicity	4.6
Average I/σI (outer resolution shell)	8.0 (3.5)

Table 29: Data processing statistics for the MhpC/Azelaic acid co-crystal.

6.1.2. Structural determination and refinement of the MhpC/Azelaic acid co-crystal

6.1.2.1. Determination of structure by molecular replacement

As for all previous structures, this structure was also solved by the MR method using MOLREP and a monomer from the MAD model of MhpC as the phasing model. A solvent content of 45.7 % (V_m = 2.28 Å^3/Da) was calculated assuming there are two molecules of MhpC per asymmetric unit.

The cross rotation calculations were performed using reflections in the 19.1 to 3 Å resolution range with a radius of integration of 25 Å. This gave two very good peaks, each peak almost certainly corresponding to a separate rotation for each MhpC molecule in the asymmetric unit of the C2 unit cell (Table 30).

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	329.00	52.76	201.83	16.30
2	30.90	52.70	342.94	16.22
3	128.09	85.22	250.31	4.56
4	270.19	52.11	335.31	4.54
5	156.76	87.27	90.05	4.51

Table 30: The five highest peaks following cross rotation calculations performed with MOLREP using reflections between 15.1 and 3.0 Å.

The translation search, also carried out using MOLREP, proved this assumption as two good translation peaks were found, one for each of the separate rotation solutions, when performed in monoclinic spacegroup C2 (Tables 31 and 32).

Peak Number	X	Y	Z	Dens/σ
1	0.319	0.000	0.030	20.21
2	0.347	0.000	0.034	4.64
3	0.281	0.000	0.197	3.31
4	0.269	0.000	0.268	3.19
5	0.269	0.000	0.418	2.92

Table 31: The five highest translation function peaks, calculated in spacegroup C2, using the second rotation function solution.

Peak Number	X	Y	Z	Dens/σ
1	0.317	0.284	0.609	83.61
2	0.318	0.246	0.608	30.38
3	0.317	0.366	0.608	29.50
4	0.818	0.251	0.608	28.70
5	0.817	0.193	0.608	28.61

Table 32: The five highest non-crystallographic translation function peaks of the first rotation function solution, calculated after fixing the position of the initial MhpC molecule in the asymmetric unit.

The packing of the two MhpC protomers in the asymmetric unit of the C2 unit cell, for this co-crystal, can be seen when viewed using MOLPACK (Figure 72). This shows that no clashes occur between subunits in the unit cell and that as the C2 spacegroup has four asymmetric units per unit cell then overall there are eight protomers of MhpC.

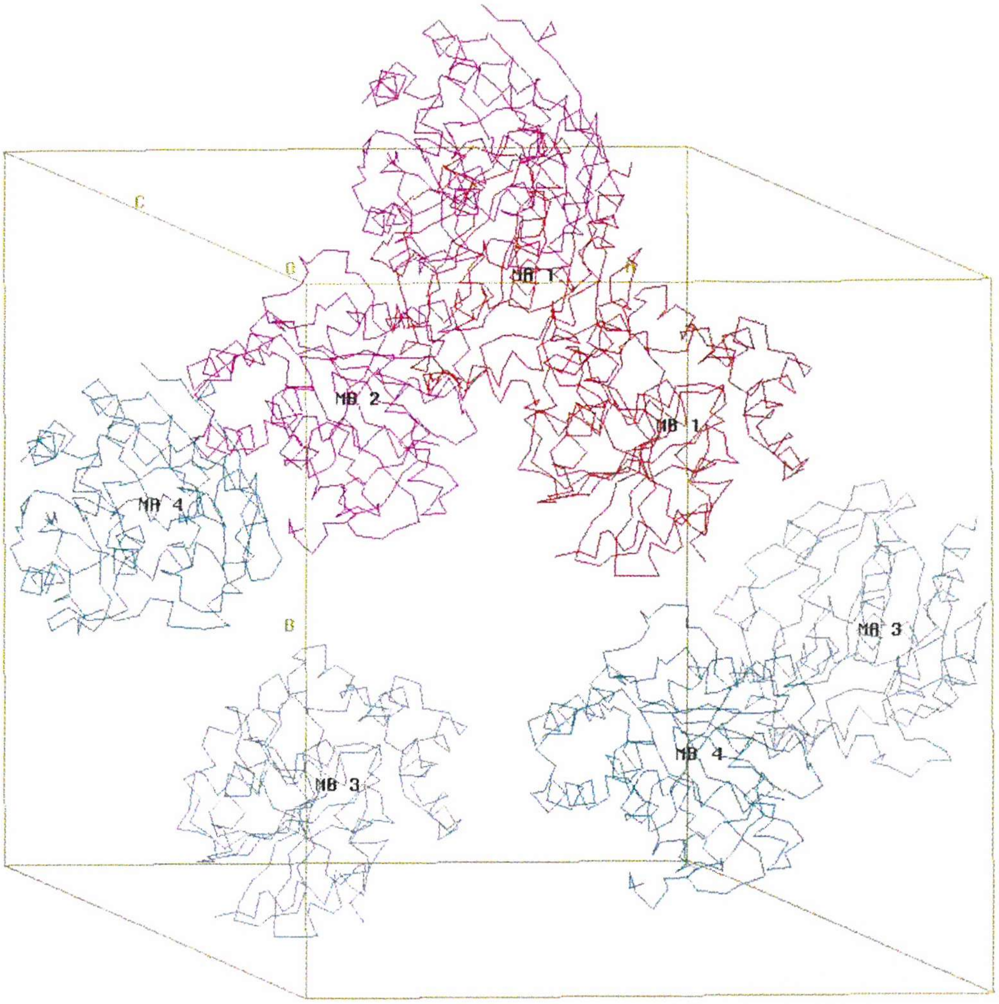


Figure 72: Crystal packing of MhpC molecules in the MhpC/Azelaic acid co-crystal generated using the MOLPACK program of the CCP4 suite. There are two subunits per asymmetric unit and four asymmetric units per unit cell owing to the monoclinic spacegroup C2 of this crystal form of MhpC.

6.1.2.2. Refinement of the MhpC/Azelaic acid co-crystal structure

The model was refined with programs of the CNS suite using 18215 reflections in the resolution range 20 to 2.5 Å. A test set of reflections (5% of data) were kept aside and used for R_{free} calculations. The model was subject initially to rigid body and positional refinement with NCS restraints. This lowered the R_{factor}

from 36.1% (R_{free} = 37.1%) to 31.0 % (R_{free} = 34.9 %). The resulting model was used to calculate sigmaA weighted electron density maps, which were then inspected and modelled using the graphical display package QUANTA. Following examination of the averaged maps *Fo-Fc* density was observed which seemed to indicate the binding of one molecule of azelaic acid per subunit of MhpC. Hence, one molecule of azelaic acid was modelled into the difference density in each active site region of the two protomers of MhpC in the asymmetric unit and further refinement was carried out. At the outset this also involved rigid body and positional refinement, but as refinement progressed also incorporated grouped and individual B-factor refinement with loosening and eventually total loss of NCS restraints.

During these further rounds of refinement as the R_{factor} became lower and the model more accurate then solvent molecules were also incorporated and any changes to side-chains that were not initially in electron density were made. Residues that lacked electron density for their side-chains were also changed to alanine residues. The refinement statistics after six rounds of refinement, thus far, are shown in Table 33.

Resolution range (Å)	19.6 – 2.5
R_{factor} (%)	24.7
R_{free} (%)	29.4
Number of reflections in working set	17330
Number of reflections in test set	885
Number of protein atoms	4360
Number of solvent molecules	31
Number of ligand atoms	26
RMS bond length deviation (Å)	0.007
RMS bond angle deviation (°)	1.25

Table 33: Refinement statistics for the MhpC/Azelaic acid co-crystal.

6.1.3. Co-ordination of Azelaic acid within the MhpC active site cavity

As mentioned in Chapter 5.1.3, a ligand which more closely mimics the natural substrate of MhpC, the RFP, would give a better indication of both the residues involved in substrate binding and those important for catalysis. In the case of azelaic acid, or nona-1,9-dioic acid, then this ligand contains the main skeleton of the RFP but unfortunately has none of the important functional groups. That is, it contains carboxylate groups in the C1 and C9 positions as does the RFP, but lacks the dienol moiety (2,4-diene and 2-hydroxyl group) and the C6 carbonyl which are believed to be the crucial functionalities for catalysis to occur. This compound is also symmetrical and quite flexible, due to being a fully saturated alkyl chain, whereas the RFP substrate of MhpC is not in both cases.

From the structural evidence obtained from the MhpC/Azelaic acid co-crystal it can be seen that there is *2Fo-Fc* electron density for the entire ligand within the active site of an MhpC subunit (Figure 73).

Figure 73: Subunit of MhpC showing that azelaic acid binds within the active site cavity, between the core domain and the lid domain. The nucleophile elbow is depicted in red, and the loops bearing Asp235 and His263 are depicted in orange and purple respectively. The *2Fo-Fc* density for the ligand is contoured at the 1σ level.

Upon closer inspection of the docking of azelaic acid within the active site cavity it can be said that one carboxylate binding site for this ligand is at the opposite end of the active site cavity to the carboxylate binding site seen for laevulinic acid (cf. Chapter 5.1.3.; Figure 62). This carboxylate group binds to a water molecule, itself coordinated to three backbone amide groups, and His114 (Aze O to His114 ND1 = 2.86 Å; Figure 74). The water molecule is coordinated to the amide nitrogen atoms of

Gly136, Thr137 and Gly138. These residues occur on the loop following strand β6, which connects the core domain to the lid domain, and appear to form a kind of binding cleft. This could possibly be due to the residues of this loop all being fairly small, the motif in this region being GGGTGG, and hence having no larger side-chains which could cause steric hindrance within this region of the active site. The water molecule is almost equally coordinated to both oxygen atoms of the carboxylate group, being 3.13 Å from one and 3.15 Å from the other. The distances from the water molecule to Gly138 N, Thr137 N and Gly136 N are 2.96 Å, 2.85 Å and 3.24 Å, respectively, and therefore it could hydrogen-bond to two of these three residues easily. The histidine residue, His114, occupies a different position in this structure than it does in the laevulinic acid co-crystal structure (cf. Chapter 5.1.3.; Figure 62A). The side-chain of the histidine has flipped to a new conformation away from this carboxylate group of the ligand, and as such this movement also puts the His114 residue further away from the active site cavity. The distance from the ND1 atom of the His114 side-chain to the Ser110 OG in this structure is 8.6 Å, whereas in the MhpC/Laevulinic acid co-crystal structure the same distance is 4.9 Å. The carboxylate group of azelaic acid actually displaces the His114 side-chain and occupies the same position in the active site that the side-chain occupies in the MhpC/Laevulinic acid structure. The movement of the His114 side-chain is therefore helped by having the GGGTGG motif close to it, as there will be no larger side-chains in the vicinity which could limit the movements of this residue. Also, in this altered conformation one of the nitrogen groups of the histidine imidazolium can hydrogen-bond to a serine hydroxyl group (His114 NE2 to Ser85 OG = 3.5Å).

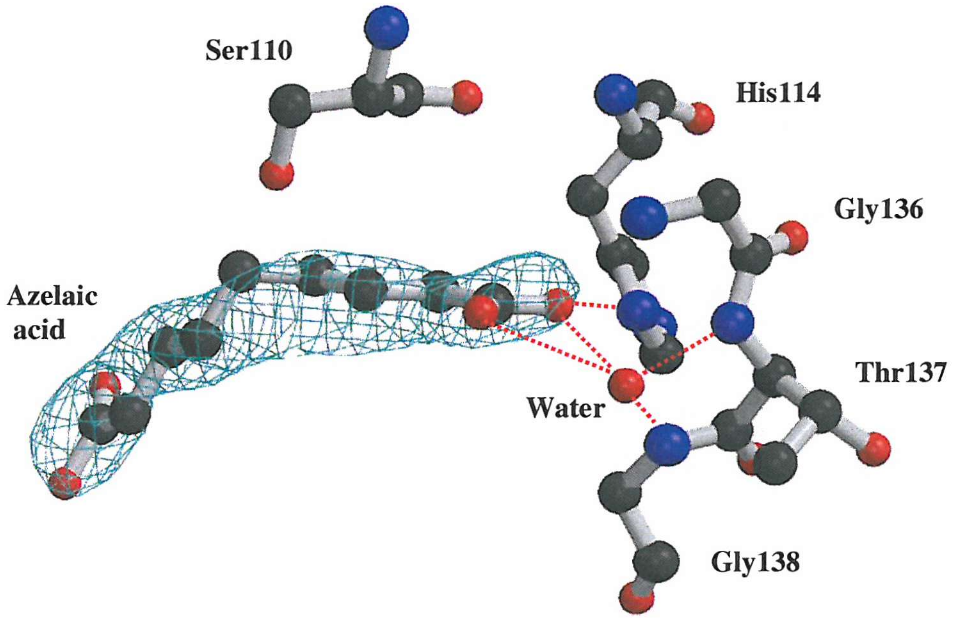


Figure 74: Interaction between one carboxylate group of azelaic acid and residues in the active site of MhpC. The carboxylate coordinates to a water molecule and His114, which has a different conformation in this structure than those obtained previously. The *2Fo-Fc* density for the ligand is contoured at 1.0σ. Ser110 of the catalytic triad is also shown so that the close proximity of the ligand to the active site residues can be seen. Putative hydrogen bonds are shown by dotted red lines.

The other carboxylate group of the ligand does not appear to coordinate to anything specifically within the active site cavity as it is in a region that is towards the lid domain of the subunit, away from the catalytic residues and expected substrate binding pocket. The area that it occupies is quite hydrophobic in nature, consisting of Ile151, Leu154, Phe202 and Leu206, but the carboxylate group does hydrogen bond to Asn155, the distance from Aze O to Asn155 ND2 being 3.17 Å (Figure 75). These residues occur on helices αD (residues Thr148 to Arg159) and αG (residues Arg196 to Ala208) of the lid domain, which are stabilised by a hydrophobic stacking interaction between the side-chains of Tyr158 and Phe202.

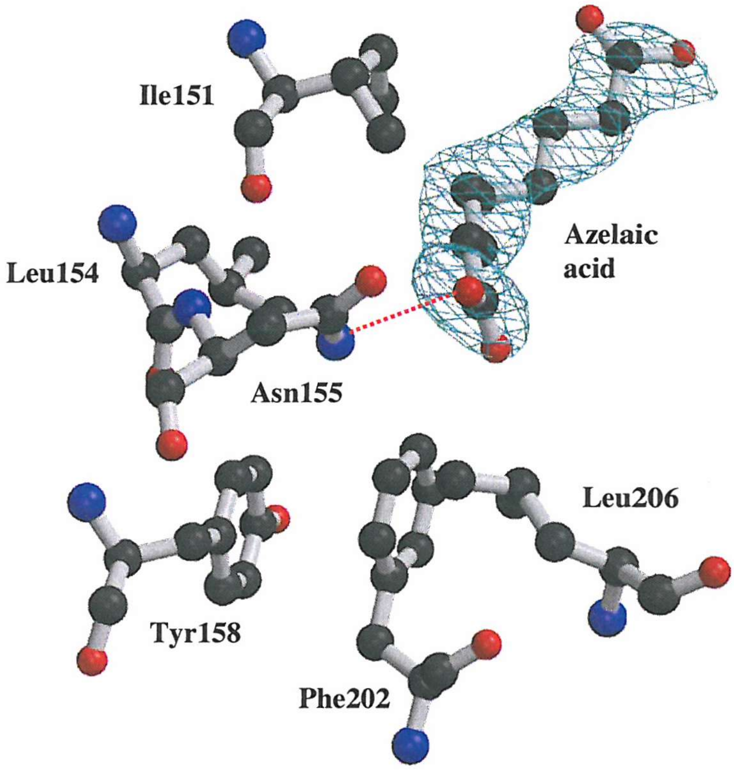


Figure 75: Coordination of an azelaic acid carboxylate group away from the active site region and towards the lid domain. The binding region is quite hydrophobic in nature, but there is a possible hydrogen-bonding interaction between Asn155 and the ligand (depicted by a dotted red line). The *2Fo-Fc* electron density for the ligand is contoured at the 1.0σ level.

It appears that the azelaic acid ligand has a kink in its backbone. Thus, as one of the carboxylate groups binds near His114, then it would be expected that the aliphatic chain would go past the His263 and Ser110 residues of the catalytic triad and the other carboxylate would bind near Arg188 and Asn49. This being the case observed for the carboxylate of laevulinic acid. However, it is forced into adopting a conformation where this part of the molecule is pointing away from the Arg188 end of the active site binding domain. This could be due to the inherent flexibility of the molecule, as it would be expected to occupy a number of different conformations due to its straight chain nine-carbon backbone. It could also be due to the unexplained density that runs contiguously with that of the Ser110 side-chain which is observed in this structure (Figure 76), and has been observed in a number of other structures (cf.

Chapter 4.3.3.4.). If this is a covalent modification then the azelaic acid ligand would not be able to occupy the active site, as the normal substrate would, and so be forced to adopt a different orientation. For instance, the modified serine side-chain could be occupying the location that the ligand would normally reside in and so force the ligand to occupy a different position. The modification of this serine residue's side-chain could also explain the increased distance from the histidine residue of the catalytic triad, His263, to the serine. The distance from His263 NE2 to Ser110 OG being about 3.82 Å, slightly too long for a good hydrogen-bonding interaction. The Asp235 to His263 bonding distance is still acceptable, being 2.64 Å from Asp235 OD2 to His263 ND1. The modelling of a water molecule into the extraneous density led to no decrease in difference density, as there was still *2Fo-Fc* and *Fo-Fc* density present extending from the Ser110 side-chain and into the active site. It is interesting to note that in the laevulinic acid co-crystal structure there is no extra density for the Ser110 side-chain.

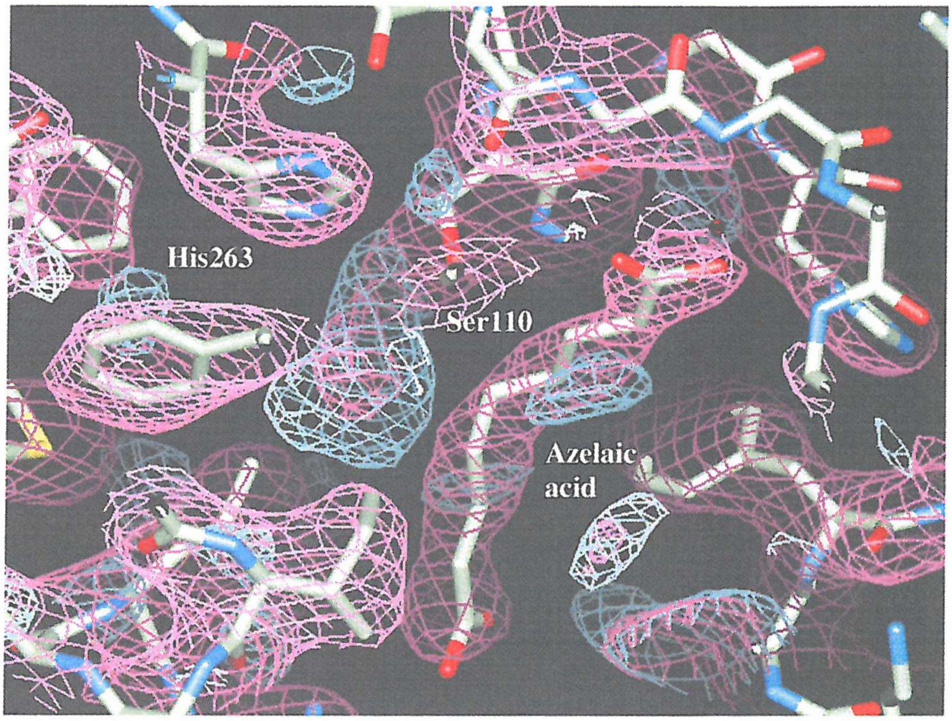


Figure 76: Extraneous density observed for the side-chain of Ser110, possibly due to a covalent modification. This may explain why azelaic acid adopts a conformation in which the backbone is kinked and so one end of the molecule is forced away from the active site region towards the lid domain of the protein. *2Fo-Fc* density (purple) is contoured at 1.0σ and *Fo-Fc* density (blue) is contoured at 2.0σ.

The carboxylate binding position seen for laevulinic acid, not observed in this structure, is instead occupied by a water molecule, which coordinates to Arg188 and Asn49 (Figure 77). It was hoped that this structure would reveal both carboxylate binding sites of the natural MhpC substrate. However, by taking information from both this co-crystal structure and that of laevulinic acid, it can be speculated that one carboxylate of the RFP coordinates at the Arg188 end of the active site and that the other end of the substrate binds at the His114 end of the active site. This would lead to the substrate passing closely to both Ser110 and His263 of the catalytic triad, in particular the C-6 carbonyl of the substrate would be close to the catalytic triad residues, and so could account for either of the proposed mechanisms. A ligand which mimics more closely the functional groups of the natural substrate might therefore provide better evidence for which of the proposed mechanisms is more likely.

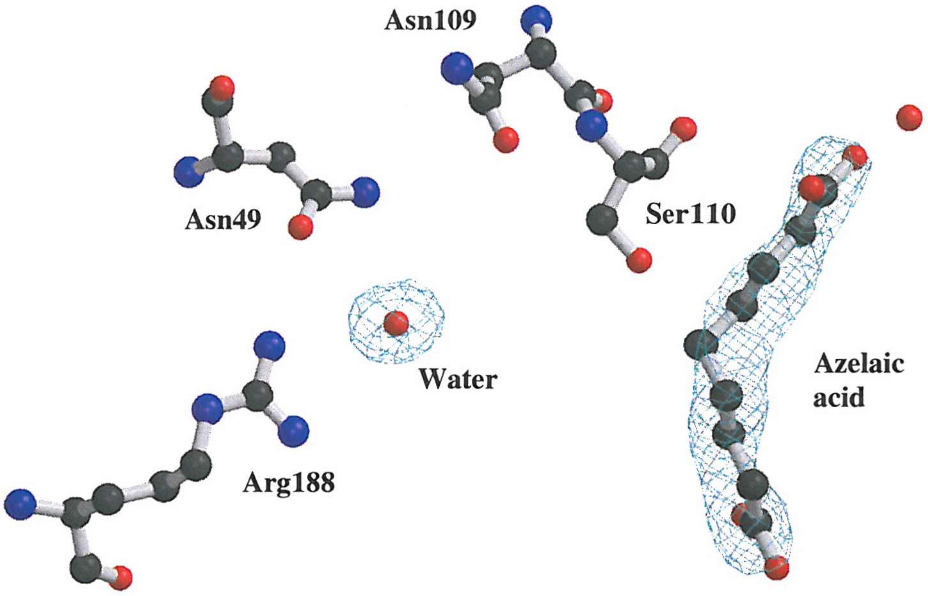


Figure 77: Coordination of a water molecule to Arg188 and Asn49 in the active site cavity of MhpC. This water molecule is in a similar position to the carboxylate group of laevulinic acid in that co-crystal structure. The ligand, azelaic acid, is also shown, as well as the active site serine, Ser110, and adjacent residue, Asn109. The *2Fo-Fc* density is contoured at 1.0σ.

6.2. MhpC co-crystallised with 2,6-diketonona-1,9-dioic acid (DKNDA)

6.2.1. Data collection and processing of the MhpC/DKNDA co-crystal

Data from an MhpC crystal co-crystallised with DKNDA was collected at 100K on ESRF beamline ID29 (Grenoble, France) using an ADSC scanner. Two datasets were collected; one at a crystal to detector (X-D) distance of 250mm and another at a X-D distance of 300mm to collect the overloaded reflections. A total of 185 1°-oscillation frames were measured for the former and 90 2°-oscillation frames collected for the latter. The datasets were processed with MOSFLM to a resolution of 2.8Å (X-D = 250mm) and 3.2Å (X-D = 300mm).

No	PEN	LATT	a	b	c	alpha	beta	gamma	Possible spacegroups
16	193	oC	199.81	199.91	79.69	89.9	90.1	78.0	C22, C2221
15	193	mC	199.91	199.81	79.69	90.1	90.1	102.0	C2
14	193	tP	125.84	155.27	79.69	90.0	90.01	90.0	P4, P41, P42, P43, P422, P4212, P4122, P41212, P4222, P42212, P4322, P43212
13	193	mC	199.81	199.91	79.69	89.9	90.1	78.0	C2
12	148	oC	79.69	320.60	125.84	89.9	90.1	104.4	C222, C2221
11	148	mC	79.69	320.62	125.84	90.0	90.1	75.6	C2
10	148	mC	320.60	79.69	125.84	90.1	90.1	75.6	C2
9	147	mC	79.69	263.80	155.27	90.0	90.0	72.6	C2
8	147	oC	79.69	263.80	155.27	90.0	90.0	107.4	C222, C2221
7	146	mC	263.80	79.69	155.27	90.0	90.0	72.6	C2
6	1	oP	79.69	125.84	155.27	90.0	90.0	90.1	P222, P2221, P21212, P212121
5	1	mP	79.69	125.84	155.27	90.0	90.0	90.1	P2, P21
4	1	mP	125.84	79.69	155.27	90.0	90.0	90.1	P2, P21
3	0	mP	79.69	155.27	125.84	90.0	90.1	90.0	P2, P21
2	0	aP	79.69	125.84	155.27	90.0	90.0	89.9	P1
1	0	aP	79.69	125.84	155.27	90.0	90.0	90.01	P1

Table 34: Bravais lattices generated by the auto-indexing function of MOSFLM for the MhpC/DKNDA co-crystal diffraction data. Only solutions with a penalty of less than 200 are listed. There was initially some ambiguity between solutions 5 and 6.

Initially, there was a degree of spacegroup ambiguity as the auto-indexing of this data, in MOSFLM, gave identical Bravais lattice penalties for both the monoclinic point group 2 and the orthorhombic point group 222 (Table 34). The cell dimensions for both were a = 79.69 b = 125.84 c = 155.27, and the fact that all the angles for the monoclinic form were almost 90° led to further uncertainty in the choice of spacegroup. The spacegroup P2 was chosen, as this gave good predictions for spots on the diffraction patterns and a structure had previously been solved with the spacegroup $P2_1$ and fairly similar cell dimensions (cf. chapter 5.2.1.). As an MhpC co-crystal structure had already been determined in spacegroup $P2_1$, so the diffraction patterns were checked for screw axes and consequently one was found (Figure 78). This lead to a reindexing of the data so that the screw axis was along *k* and so that, after post-refinement, the final cell dimensions were a = 80.19 b = 155.67 c = 125.98, with the β angle thus being the one that was just greater than 90° (β = 90.42°). Upon retrospective analysis of the symmetry of the diffraction pattern in HKLVIEW it became clear that point group 2 was the correct choice, as only two layers had mm symmetry (*hk0* and *0kl*) (Figure 94 and Figure 95), whereas if the spacegroup had been orthorhombic then all three zeroth layers would have had mm symmetry. The *h0l* layer showed no mirror symmetry, only inversion symmetry (Figure 96), as did the *h1l* layer (Figure 99). At the next level up, if the correct spacegroup were monoclinic then you would expect to see only one mirror plane for *1kl* and *hk1*, perpendicular to the *k* axis, whereas an orthorhombic point group would still give two mirror planes. In this case, the former situation was observed (Figure 97 and Figure 98), lending further support to $P2_1$ being the correctly chosen spacegroup for this co-crystal. All of the symmetry elements used to determine whether the point group was 2 or 222 can be seen in the diffraction patterns in Appendix A.

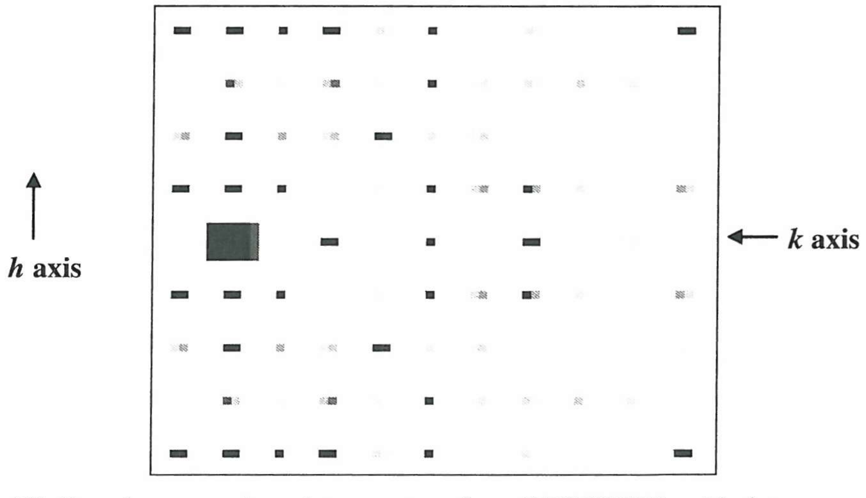


Figure 78: Pseudo-precession picture output from HKLVIEW, with data processed in spacegroup P2, showing the systematic absences along the reciprocal space lattice unique *k* axis, indicative of a 2_1 screw axis.

The two datasets were then sorted together in SORTMTZ, a CCP4 program, and scaled and merged using SCALA, another program of the CCP4 suite. The final data processing statistics for these amalgamated datasets are shown in Table 35. The combined resolution when scaling the data was to 2.8Å, but the lower resolution data was cut back to 3.5Å from the 3.2Å at which it was initially processed. Despite this the merging R_{factor} was still high, as had been the case for almost all MhpC crystal structures solved, and was possibly due to an inherent difficulty in obtaining good quality crystals for this protein. All crystals tended to diffract poorly, hence the highest resolution obtained, thus far, for MhpC being only 2.1Å (cf. chapter 4.1.).

Total number of reflections	379981
Number of unique reflections	75626
Resolution (Å)	2.8
Completeness (%)	99.5
R_{merge} (%) (outer resolution shell)	17.6 (41.2)
Multiplicity	5.0
Average I/σI (outer resolution shell)	2.8 (1.7)

Table 35: Data collection statistics for the MhpC/DKNDA co-crystal.

6.2.2. Structural determination and refinement of the MhpC/DKNDA co-crystal

6.2.2.1. Determination of structure by molecular replacement

Once more, the phases for the final model were determined by the technique of molecular replacement, using MOLREP and the MAD MhpC coordinates as the phasing model. A solvent content of 60.9% (V_m = 3.17 $Å^3$/Da) was calculated assuming there are eight molecules of MhpC in the asymmetric unit of the co-crystal. This number of protomers is double that of the number in the MhpC/DOHA structure, despite them both having more or less the same solvent contents. However, this is to be expected as the unit cell is almost precisely twice the size of the MAD MhpC and MhpC/DOHA co-crystal structures, which both have four protomers per asymmetric unit. That is to say, the c dimension of the unit cell is twice the length in this case, but the a and b cell dimensions are almost identical. This might have led to the assumption that the crystal was twinned, but this is not shown in any of the tests used to detect twinning, such as examining the graph of cumulative intensity distributions from TRUNCATE. If the curve for acentric data is hyperbolic, and not sigmoidal, then this indicates that the crystal is untwinned. A sigmoidal curve far below that of the theoretical curve for acentric data would indicate twinning, as shown in Figure 79A. This has not been observed in these plots for the MhpC/DKNDA co-crystal data output from TRUNCATE, as the curves for the theoretical and actual acentric data are almost co-incidental, hence suggesting that this data is not twinned (Figure 79B).

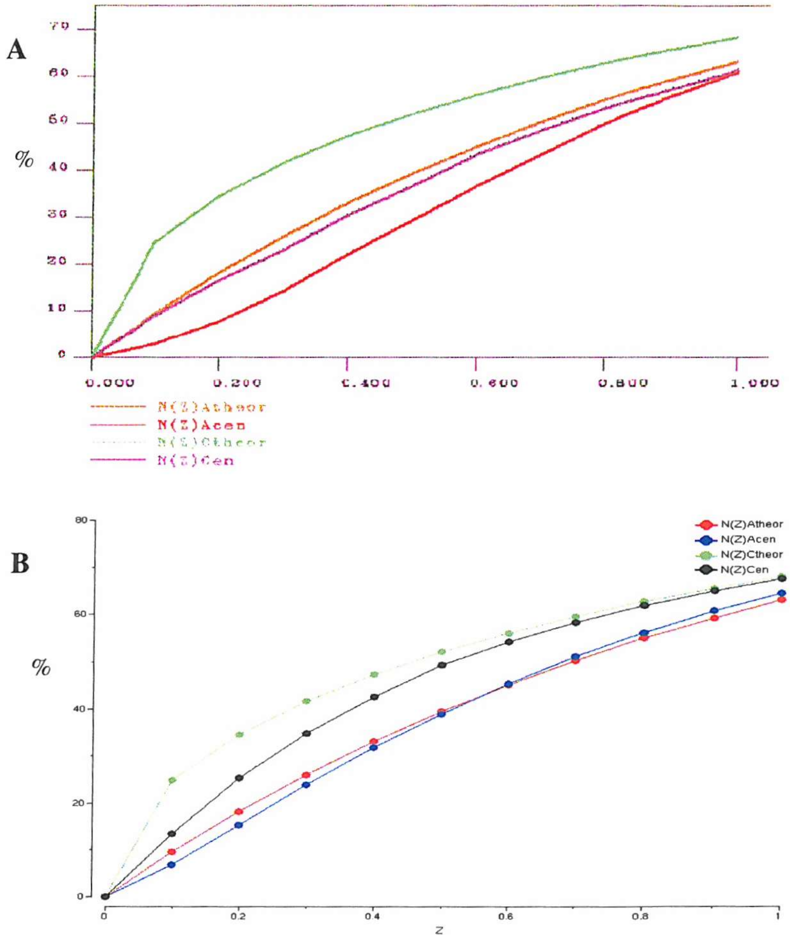


Figure 79: **(A)** Example of the cumulative intensity distributions for a twinned dataset from the output of TRUNCATE. This shows that the acentric data for a twinned crystal gives a sigmoidal curve lower than the theoretical acentric data curve (*Source*: www.ccp4.ac.uk); **(B)** Cumulative intensity distributions from the output of TRUNCATE for the MhpC/DKNDA co-crystal data illustrating no apparent crystal twinning.

Based on the above conclusions it would be expected that the asymmetric unit would contain two MhpC tetramers (as described in chapter 4.3.2.4.), and so a tetramer was used as the search model in molecular replacement. The cross rotation calculations were performed using reflections in the 126 to 3.0 Å resolution range with a radius of integration of 30Å. This gave four good peaks, as shown in Table 36, which is consistent with more than one peak for each tetramer. This occurs because each tetramer can have more than one orientation due to non-crystallographic two-fold rotational symmetry (222 symmetry).

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	338.19	90.00	179.26	18.93
2	158.31	90.00	179.19	18.76
3	206.07	90.00	179.27	18.31
4	26.20	90.00	179.23	18.22
5	52.26	2.29	102.27	5.02
6	76.45	2.29	126.43	4.89
7	249.84	2.71	84.84	4.77
8	263.32	1.80	120.82	4.50

Table 36: The eight highest peaks following cross rotation calculations performed with MOLREP using data between 126 to 3.0 Å.

Following this, the translation search, carried out in spacegroup $P2_1$, gave good peaks for all four rotation function solutions, but only two were chosen by MOLREP, for rotation function solutions 1 and 3, as the other two were related to these by a shift of 180° around α due to the rotational symmetry mentioned previously. The two translation function solutions are shown in Tables 37 and 38.

Peak Number	X	Y	Z	Dens/σ
1	0.070	0.000	0.303	18.24
2	0.583	0.000	0.043	3.37
3	0.088	0.000	0.059	3.18
4	0.480	0.000	0.467	2.99
5	0.138	0.000	0.420	2.95

Table 37: The five highest translation function peaks, calculated in spacegroup $P2_1$, using the third rotation function solution.

Peak Number	X	Y	Z	Dens/σ
1	0.952	0.515	0.200	50.44
2	0.953	0.485	0.200	29.98
3	0.929	0.483	0.196	27.05
4	0.901	0.487	0.196	20.51
5	0.930	0.549	0.196	19.24

Table 38: The five highest translation function peaks using the first rotation function solution, after fixing the position of the first MhpC tetramer.

Thus, two tetramers were found in the asymmetric unit, consequently showing that eight protomers per asymmetric unit was a correct assumption. Also, searching for another translation function peak, for one of the other rotation function solutions, gave no increase in density/σ levels above noise (Table 39).

Peak Number	X	Y	Z	Dens/σ
1	0.428	0.222	0.196	16.51
2	0.427	0.332	0.197	16.49
3	0.428	0.485	0.696	16.45
4	0.428	0.517	0.697	16.32
5	0.428	0.285	0.196	15.72

Table 39: The translation function search for rotation function two, after two tetramers had already been found, showing that no more tetramers were present in the asymmetric unit.

Furthermore, by checking the packing of the solution obtained, using the CCP4 program MOLPACK, it could be seen that the two tetramers were nicely ordered within the centre of the unit cell, with further subunits arranged above and below them, which could apparently form further tetramers across the unit cell boundaries, as each unit cell in a P2 monoclinic crystal is known to contain two asymmetric units (Figure 80).

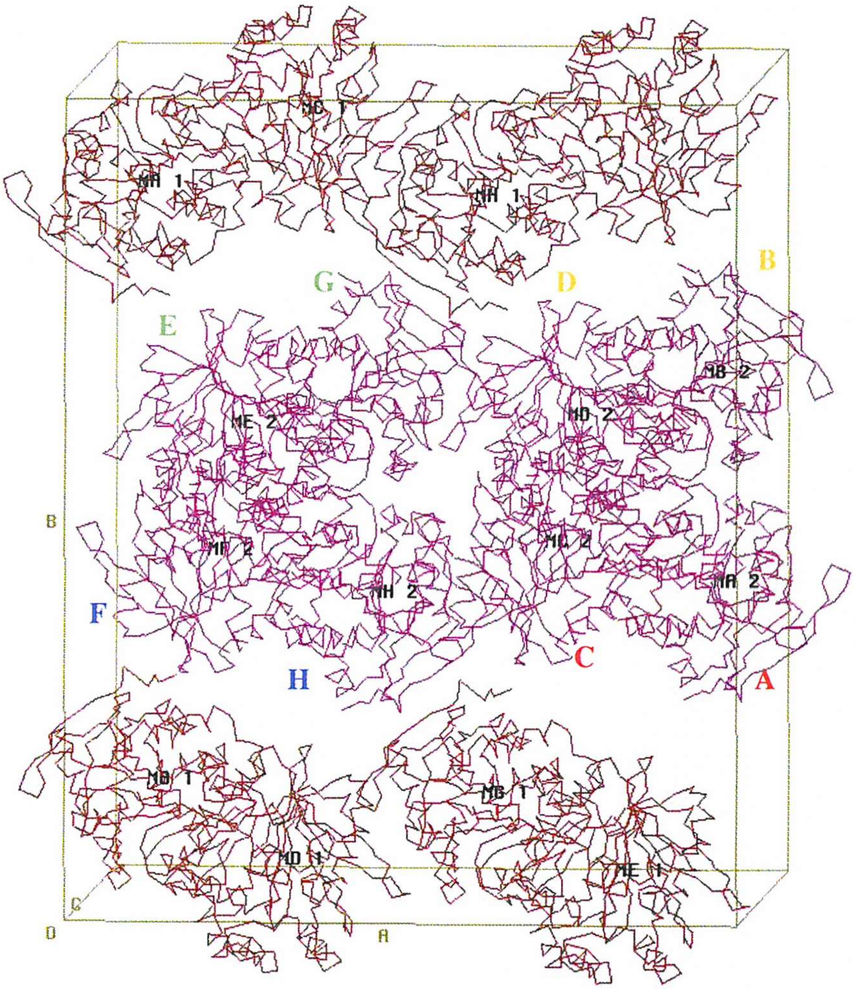


Figure 80: Crystal packing of MhpC protomers in the MhpC/DKNDA co-crystal unit cell, as visualised using MOLPACK. There are 16 MhpC subunits in the unit cell because the monoclinic spacegroup P2 has two asymmetric units per unit cell, and there are two MhpC tetramers per asymmetric unit in this case. Dimers are shown as same coloured letters.

6.2.2.2. The epic refinement of the MhpC/DKNDA co-crystal structure

6.2.2.2.1. Investigation of subunit disparities within the MhpC dimer

Previous work on a number of structures from MhpC/DKNDA co-crystals, between 2.4 and 3.5Å resolution, had led to the belief that there was some ambiguity between the two subunits of each dimer. Specifically, the electron density at the active site for the proposed ligand varied considerably between subunits in the asymmetric unit. However, after careful examination of both *2Fo-Fc* and *Fo-Fc* electron density at the active sites in all of these structures it seemed that the density for this potential ligand was more extensive in some subunits than others, and that when the subunit dispersion was investigated it was revealed that this discrepancy may be based on the quaternary arrangement of the protein.

Therefore, it was decided to refine the model using NCS restraints, and to have these restraints in two groups. As a result, one subunit was taken from each dimer at random and put into a separate NCS restraint group. The asymmetric unit of this co-crystal contains two tetramers, and consequently four dimers, with the subunit distribution of these dimers being: A with C, B with D, E with G and F with H, as shown in Figure 80. Thus, for the initial refinement of this model NCS restraints were applied in the following manner: subunits A, B, E and F in the first group and subunits C, D, G, and H in the second group. In addition, the whole polypeptide chain for each subunit was not used in restraints but only that portion from residues 21 to 284. This was because it had previously been observed that the *N*-terminal region varied between subunits as well, and that some electron density for the five unobserved residues at this terminus was present in some protomers, but not in others. This phenomenon being almost certainly due to the flexibility of this region, as is the case for *N*-terminal regions of proteins in general. In some other structures of MhpC symmetry related molecules help to stabilise this region and hence some of the *N*-terminal residues are visible in the electron density maps.

The model was refined with programs of the CNS suite using 75457 reflections in the resolution range 129 to 2.8 Å. A test set of reflections (5% of data) was set aside and used for R_{free} calculations. The model was initially subject to a round of refinement that included rigid body refinement, positional refinement, with

NCS restraints as previously described, and grouped B-factor refinement. This reduced the R_{factor} from 43.2% (R_{free} = 42.6%) to 25.7% (R_{free} = 28.3%). The resultant model was used to calculate sigmaA weighted electron density maps.

After the aforementioned refinement had been completed, and analysis of the electron density maps had been inconclusive regarding the putative ligand density within the active site, it was realised that the NCS restraints groups used had actually been chosen by a misguided approach. That is, by randomly choosing one subunit from each dimer for each restraint group no account had been made to distinguish the variability between each subunit of an MhpC dimer. As a result, the groups chosen contained subunits that had mixed environments, such as in their crystal contacts, rather than identical environments.

Accordingly, prior to altering these groups a control refinement was conducted by including all subunits in one NCS restraint group and carrying out a round of refinement. As before, only residues 21 to 284 were included in the NCS restraints. Refinement, using programs of the CNS suite, consisted of rigid body refinement, positional refinement and grouped B-factor refinement and decreased the R_{factor} from 43.2% (R_{free} = 42.7%) to 25.7% (R_{free} = 28.2%), not too dissimilar to that obtained previously. SigmaA weighted electron density maps were made and the model and maps examined in QUANTA. These were used to determine what the best choice of subunits for the restraint groups should be, based on their relative local environments and crystal packing amongst other criteria.

It became immediately obvious that one helix in particular varied significantly amongst subunits of the dimers. This was helix αF of the lid domain (residues 181 to 195), which for one subunit had a symmetry-related molecule packed against it, while in the other subunit of the dimer it was solvent exposed. Thus, the *2Fo-Fc* electron density for the side-chains of these helical residues was much better in the former case, whereas in the latter case there was more of the model out of the density and greater amounts of *Fo-Fc* density, indicating possible movements of these residues side-chains due to increased flexibility in this region. Therefore, it was decided that those subunits of each dimer that had this helix with a symmetry-related molecule packed against it would constitute one NCS restraint group, and the other subunit of each dimer, without this feature, would make up the other group. This led to the subunit distribution being: subunits A, B, G and H in one restraint group, and subunits C, D, E and F comprising the other group, quite unlike the random grouping chosen for the initial refinement of the MhpC/DKNDA co-crystal.

As in the previous two cases, programs of the CNS suite were used in refinement bringing the R_{factor} from 43.2% (R_{free} = 42.7%), initially, to a final value of 25.2% (R_{free} = 28.1%). These values were only marginally better than those produced by the other two restraint methods, and so it was decided to redo the refinement including simulated annealing, as the differences between the subunits from each group may be great enough to cause quite large shifts in structure, especially in the *N*-terminal and helical lid regions. Thus, the refinement was repeated, with simulated annealing (slowcool from 5000K in 50K steps) bridging the rigid body and positional refinement steps, prior to making electron density maps. The CNS program NCS_AVERAGE_MAP was used to make a sigmaA weighted annealed omit map, in which map making is followed by a single cycle of electron density averaging. NCS restraints were also incorporated when running this program. The application of simulated annealing, and the making of averaged electron density maps, was shown to make a difference in the refinement process, as the R_{factor} attained was notably lower afterwards, in comparison to that obtained for previous rounds of refinement. So that the initial R_{factor} of 43.2% (R_{free} = 42.7%) eventually fell to 24.2% (R_{free} = 27.8%), a full 1.5% lower (excluding the R_{free} value) than that obtained by the two previous methods of either randomly choosing subunits for restraint groups or putting all subunits as one group.

6.2.2.2.2. Further refinement of the MhpC/DKNDA co-crystal structure

Once a seemingly correct procedure for refinement had been determined, that of getting the correct subunit distributions for the NCS restraint groups, then further refinement was undertaken, with model-building, to make the electron density for the ligand at the active site better and to try and build a model of DKNDA into this density.

After the initial refinement, examination of the averaged electron density maps revealed extensive *2Fo-Fc* and *Fo-Fc* density within one active site of each dimer, and less density within the other subunit. For instance, in subunit A there was considerable density that was contiguous with the Ser110 side-chain and extended towards both ends of the active site, towards Arg188 in one direction and towards

His114 in the other direction, as shown in Figure 81A. In contrast, the opposing subunit of this dimer, subunit C, had less density within the active site, although there was still some present. Most of this was concentrated towards the Arg188 end of the active site cavity (Figure 81B). It should be pointed out that the histidine residue His114 occurs in a different conformation dependent on which subunit of the dimer it is in. In those subunits with the most extensive difference density in the active site then His114 has shifted away from the cavity, whereas the other subunits have His114 pointing towards the active site cavity. This feature can be seen by examining Figures 81A and 81B. The placement of this residue's side-chain occurred during simulated annealing, that is to say the movement was carried out automatically by the program. This lends further support to the idea that there is some subunit disparity within the MhpC dimeric arrangement, as the conformational changes for this residue are consistent within every dimer of the asymmetric unit. It had previously been observed for other MhpC structures that this residue's side-chain could exist in alternate conformations (cf. chapters 5.1.3. and 6.1.3.; Mohammed, 2001).

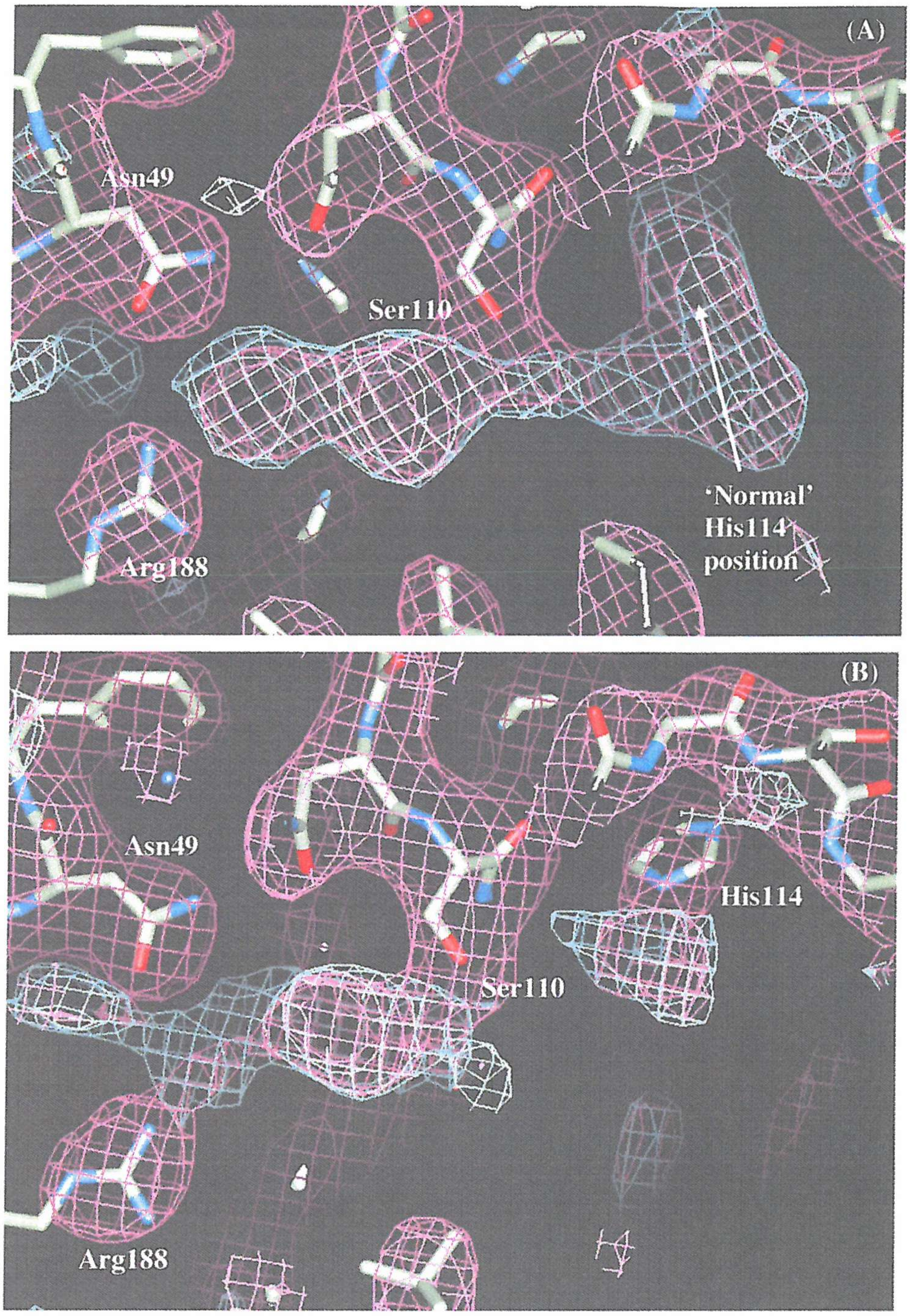
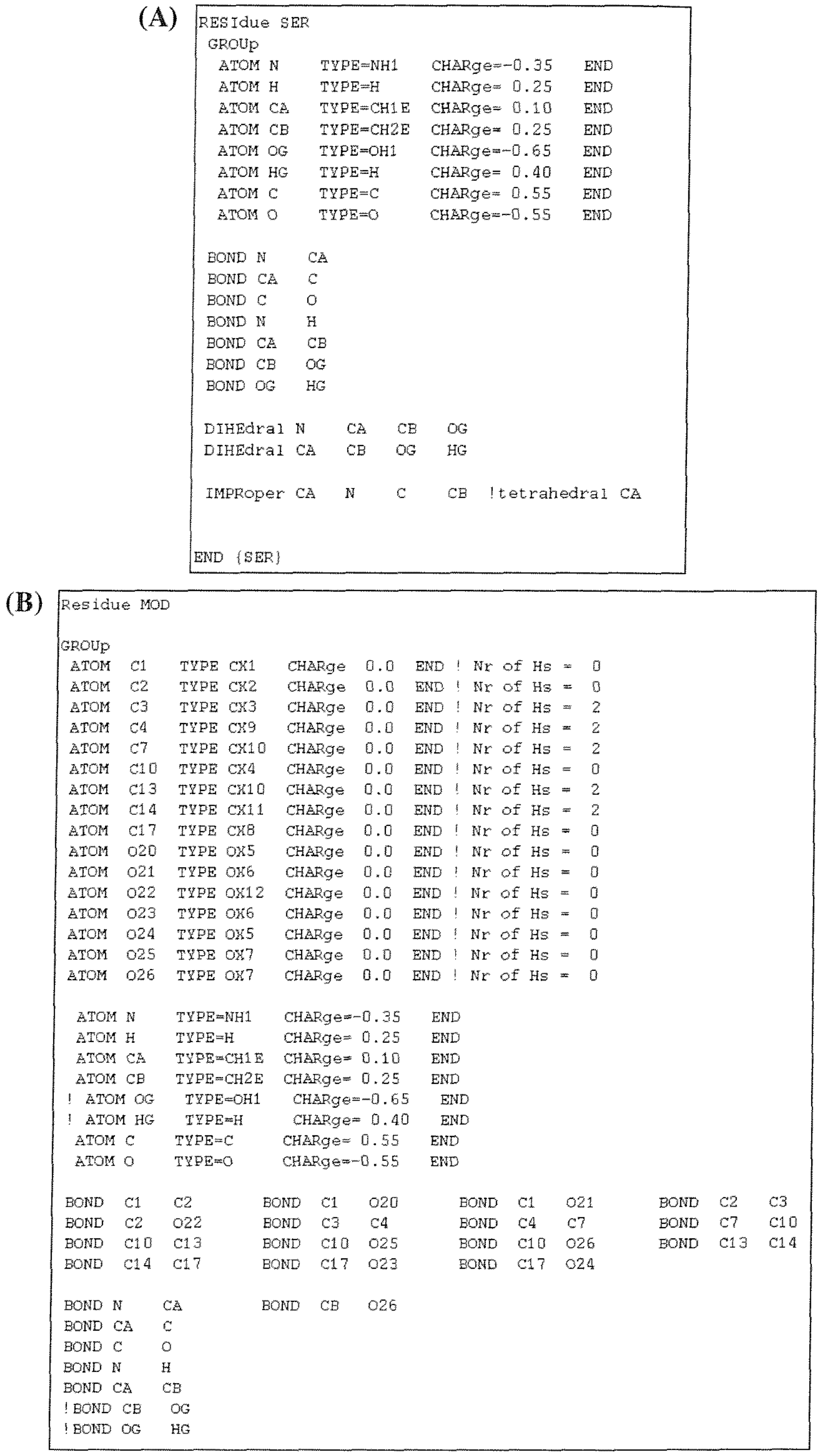


Figure 81: **(A)** Active site cavity of subunit A showing extensive *2Fo-Fc* (purple) and *Fo-Fc* (blue) electron density for the DKNDA ligand, which differs considerably from the density observed in subunit C shown in **(B)**. Both of these subunits form one MhpC dimer, hence indicating a difference in subunit diversity within the MhpC quaternary structure. *2Fo-Fc* density contoured at 1σ, *Fo-Fc* density contoured at 2σ.

As the ligand density appeared so convincing then general model-building of the MhpC structure was conducted, followed by a further round of refinement, before deciding the best way to proceed. Certain changes were made, such as had been carried out for other structures. For example, changing amino acid side-chains to alanine if no electron density was present for them or by moving sections of the model that were out of density into electron density if possible. Refinement consisted of those same procedures carried out in the initial refinement and lowered the R_{factor} a fraction to 24.21% (R_{free} = 27.61%).

6.2.2.2.3. Covalent modification of the active site serine residue

The extra *2Fo-Fc* density within the active site cavities of subunits A, B, G and H ran contiguously with that of the Ser110 side-chain. This led to the conclusion that the ligand may have become covalently attached to this residue. This is not chemically unfeasible as the carbonyl group in the C6 position of DKNDA could be nucleophilically attacked by the hydroxyl of the active site serine to form a hemi-ketal. Thus, one strategy utilised to model-build the ligand into the active site was to create a modified residue which incorporated both the atoms of the ligand and those of the serine residue. The ChemNote feature of QUANTA was used to make a model of the ligand as the *gem*-diol form, such that it had two hydroxyl groups at the C6 position. Once this had been done, the ligand was moved so that it occupied as closely as possible the difference density within the active site of subunit A, with the C9 carboxylate group at the Arg188 end of the active site and the C1 carboxylate at the His114 end, thus positioning the C6 area close to the Ser110 OG. Both the protein model and the ligand model were exported from QUANTA as PDB files and LSQKAB (CCP4) was used to create the ligands for subunits B, G and H by applying the correct rotations and translations from that in subunit A. The program XPLO2D was used to create a dictionary for the ligand; that is to say, topology and parameter files were created that contained information about the bond distances and angles for this molecule. As a modified amino acid residue was required then the normal protein parameter and topology files utilised by the CNS suite of programs were altered so as to contain these new details for a residue with the acronym MOD (Figure 82).



```
RESIdue SER
 GROUp
  ATOM N     TYPE=NH1    CHARge=-0.35   END
  ATOM H     TYPE=H      CHARge= 0.25   END
  ATOM CA    TYPE=CH1E   CHARge= 0.10   END
  ATOM CB    TYPE=CH2E   CHARge= 0.25   END
  ATOM OG    TYPE=OH1    CHARge=-0.65   END
  ATOM HG    TYPE=H      CHARge= 0.40   END
  ATOM C     TYPE=C      CHARge= 0.55   END
  ATOM O     TYPE=O      CHARge=-0.55   END

 BOND N    CA
 BOND CA   C
 BOND C    O
 BOND N    H
 BOND CA   CB
 BOND CB   OG
 BOND OG   HG

 DIHEdral N    CA   CB   OG
 DIHEdral CA   CB   OG   HG

 IMPRoper CA   N    C    CB  !tetrahedral CA


END {SER}
```

```
Residue MOD

GROUp
 ATOM   C1    TYPE CX1    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   C2    TYPE CX2    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   C3    TYPE CX3    CHARge  0.0  END ! Nr of Hs =  2
 ATOM   C4    TYPE CX9    CHARge  0.0  END ! Nr of Hs =  2
 ATOM   C7    TYPE CX10   CHARge  0.0  END ! Nr of Hs =  2
 ATOM   C10   TYPE CX4    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   C13   TYPE CX10   CHARge  0.0  END ! Nr of Hs =  2
 ATOM   C14   TYPE CX11   CHARge  0.0  END ! Nr of Hs =  2
 ATOM   C17   TYPE CX8    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O20   TYPE OX5    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O21   TYPE OX6    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O22   TYPE OX12   CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O23   TYPE OX6    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O24   TYPE OX5    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O25   TYPE OX7    CHARge  0.0  END ! Nr of Hs =  0
 ATOM   O26   TYPE OX7    CHARge  0.0  END ! Nr of Hs =  0

  ATOM N     TYPE=NH1    CHARge=-0.35   END
  ATOM H     TYPE=H      CHARge= 0.25   END
  ATOM CA    TYPE=CH1E   CHARge= 0.10   END
  ATOM CB    TYPE=CH2E   CHARge= 0.25   END
!  ATOM OG    TYPE=OH1    CHARge=-0.65   END
!  ATOM HG    TYPE=H      CHARge= 0.40   END
  ATOM C     TYPE=C      CHARge= 0.55   END
  ATOM O     TYPE=O      CHARge=-0.55   END

BOND   C1    C2         BOND   C1    O20         BOND   C1    O21         BOND   C2    C3
BOND   C2    O22        BOND   C3    C4          BOND   C4    C7          BOND   C7    C10
BOND   C10   C13        BOND   C10   O25         BOND   C10   O26         BOND   C13   C14
BOND   C14   C17        BOND   C17   O23         BOND   C17   O24

BOND N    CA            BOND   CB    O26
BOND CA   C
BOND C    O
BOND N    H
BOND CA   CB
!BOND CB   OG
!BOND OG   HG
```

Figure 82: (A) Example of an entry in a protein topology file, used by the CNS suite of programs, which contains the information about bond distances and angles for a serine residue; and **(B)** part of a similar entry for the newly created residue MOD, which is that of a serine residue covalently modified by DKNDA.

The atomic coordinates for the ligand were incorporated into the PDB file for the MhpC model, so that they were in the place of Ser110 for subunits A, B, G and H, an example of which is shown in Figure 83A. The subunits C, D, E and F still had the coordinates for the normal unmodified Ser110 amino acid (Figure 83B). The coordinates for the OG of Ser110 in subunits A, B, G and H were removed so that a bond could be formed between the CB atom and that of O26 of the ligand, such that the hemi-ketal would be created. Thus, a command line in the topology file was included which would generate this new bond between the side-chain and the ligand to form the covalently modified residue. The CNS program GENERATE was run with the newly created PDB file and the new protein topology and parameter files. This program creates a molecular topology file (MTF) which can then be used by other programs of the CNS suite during refinement, and contains all the relevant bonding information for the protein model and any other molecules, such as ligands, DNA/RNA, water molecules or carbohydrate that may need to be included. Following on from GENERATE, the refinement proceeded as per normal. Consequently, rigid body refinement and positional refinement were performed and sigmaA weighted averaged electron density maps were created. On examination of the model in QUANTA it was observed that although the bond between the CB of Ser110 and the O26 of DKNDA had been created, the peptide bond between Ser110 and Met111 had become severed. Despite numerous attempts to overcome this problem, such as the repositioning of the ligand in respect to the serine residue so that there would be less constraint on the bond angles formed, the peptide bond was always broken following positional refinement. Thus, it was deemed that this was due to a problem within the CNS suite of programs themselves and as a result an alternative approach to forming the modified serine residue was undertaken.

(A)

```
ATOM    804  N   MOD A 110      22.366  -7.468  26.868  1.00 52.10           A
ATOM    805  CA  MOD A 110      22.235  -6.048  26.500  1.00 52.10           A
ATOM      1  C1  MOD A 110      19.061   1.413  26.031  1.00  0.00           A
ATOM      2  C2  MOD A 110      19.901   0.152  26.043  1.00  0.00           A
ATOM      3  C3  MOD A 110      21.303   0.229  26.649  1.00  0.00           A
ATOM      4  C4  MOD A 110      21.982  -1.138  26.565  1.00  0.00           A
ATOM      5  C7  MOD A 110      23.392  -1.135  27.157  1.00  0.00           A
ATOM      6  C10 MOD A 110      24.071  -2.503  27.071  1.00  0.00           A
ATOM      7  C13 MOD A 110      25.481  -2.498  27.663  1.00  0.00           A
ATOM      8  C14 MOD A 110      26.159  -3.866  27.578  1.00  0.00           A
ATOM      9  C17 MOD A 110      27.562  -3.789  28.185  1.00  0.00           A
ATOM     10  O20 MOD A 110      19.665   2.495  26.574  1.00  0.00           A
ATOM     11  O21 MOD A 110      17.893   1.349  25.802  1.00  0.00           A
ATOM     12  O22 MOD A 110      19.524  -0.819  25.463  1.00  0.00           A
ATOM     13  O23 MOD A 110      28.312  -4.715  28.250  1.00  0.00           A
ATOM     14  O24 MOD A 110      27.715  -2.645  28.889  1.00  0.00           A
ATOM     15  O25 MOD A 110      24.095  -2.959  25.743  1.00  0.00           A
ATOM     16  O26 MOD A 110      23.280  -3.484  27.693  1.00  0.00           A
ATOM    808  C   MOD A 110      22.051  -5.873  24.972  1.00 52.10           A
ATOM    809  O   MOD A 110      21.012  -6.228  24.411  1.00 52.10           A
```

(B)

```
ATOM   5143  N   SER C2110      -0.792 -17.789  48.975  1.00 46.13           C
ATOM   5144  CA  SER C2110      -1.817 -16.786  49.313  1.00 46.13           C
ATOM   5145  CB  SER C2110      -3.202 -17.303  48.892  1.00 66.16           C
ATOM   5146  OG  SER C2110      -4.237 -16.386  49.206  1.00 66.16           C
ATOM   5147  C   SER C2110      -1.791 -16.475  50.817  1.00 46.13           C
ATOM   5148  O   SER C2110      -0.818 -15.910  51.329  1.00 46.13           C
```

Figure 83: Example of the coordinates for the modified serine residue (MOD) in subunit A (**A**) and the unmodified Ser110 in subunit C (**B**), both subunits forming an MhpC dimer. The file was in the PDB format.

Instead of creating a modified serine residue it was decided to leave the protein and ligands separate and use a special command, known as a 'patch' command, to form a bond between the CB of Ser110 and O26 of DKNDA. Thus, in GENERATE two separate coordinate files were needed, one for the protein model and one for the ligand, along with the related topology and parameter files. The coordinates for the protein had to be adjusted so that the Ser110 residues in subunits A, B, G and H were replaced by an alanine residue, hence they would have no OG atom present, whilst leaving Ser110 in subunits C, D, E and F intact. The ligand coordinate file contained information for all four DKNDA molecules, in the correct positions to occupy the active sites of subunits A, B, G and H. The 'patch' command was included in GENERATE so that the atoms could remain close enough without causing any error due to 'bumping' of the ligand with the protein and possible disturbance of the protein structure, such as happened before when the Ser110 to

Met111 peptide bond was disrupted. Following rigid body, positional and grouped B-factor refinement the R_{factor} was 25.08% (R_{free} = 28.03%). On checking the B-factors for the ligand molecules in the output coordinate file it was seen that they were extremely high (>100) and on analysis of the electron density for the ligand, in the newly created averaged maps, a large swath of negative difference density was seen to cover the entirety of the ligand, as shown in Figure 84.

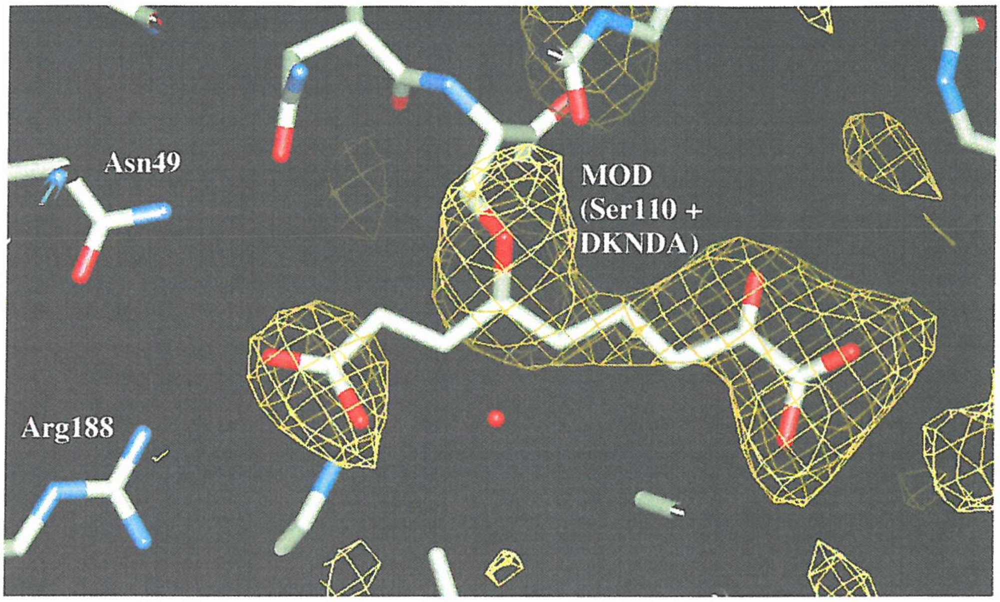


Figure 84: An illustration of the large amount of negative difference density (yellow) covering the entirety of the DKNDA ligand in the active site of subunit A. The *Fo-Fc* density is contoured at -2.0σ.

This led to the suggestion of carrying out occupancy refinement on the ligand, using the QGROUP program of the CNS suite, which takes each ligand as a separate group and refines the overall occupancy for that molecule. Firstly, the B-factors were reset to a value of 20 and then occupancy refinement was carried out by two different methods, either starting from an initial occupancy of zero or one for the atoms of the ligand molecules. Refinement of occupancies from a value of one gave final values for DKNDA of 90, 81, 76 and 83% for subunits A, B, G and H, respectively; whereas refinement from an initial value of zero gave final values of 89, 79, 73 and 79% for the equivalent subunits. It can be seen that these values differ by only 2 to 4% between either method, and so the values chosen for subsequent use were those that were refined from an initial occupancy of zero. Therefore, on re-refinement of the B-factors with these new occupancies the values obtained were much lower, being between 60 and 65 for all DKNDA molecules, not too dissimilar to those values obtained for the B-factors of other regions of the MhpC model. Following generation of sigmaA weighted averaged electron density maps the new R_{factor} was 24.1% and the R_{free} was 27.1%. Inspection of these maps revealed much less negative difference density for the ligand (Figure 85) and helped to justify the use of occupancy refinement. This also leads to the interesting conclusion that not only may MhpC have some form of half-site reactivity with respect to the ligand, but also that each subunit of the dimer that contains the ligand is still only approximately three-quarters occupied.

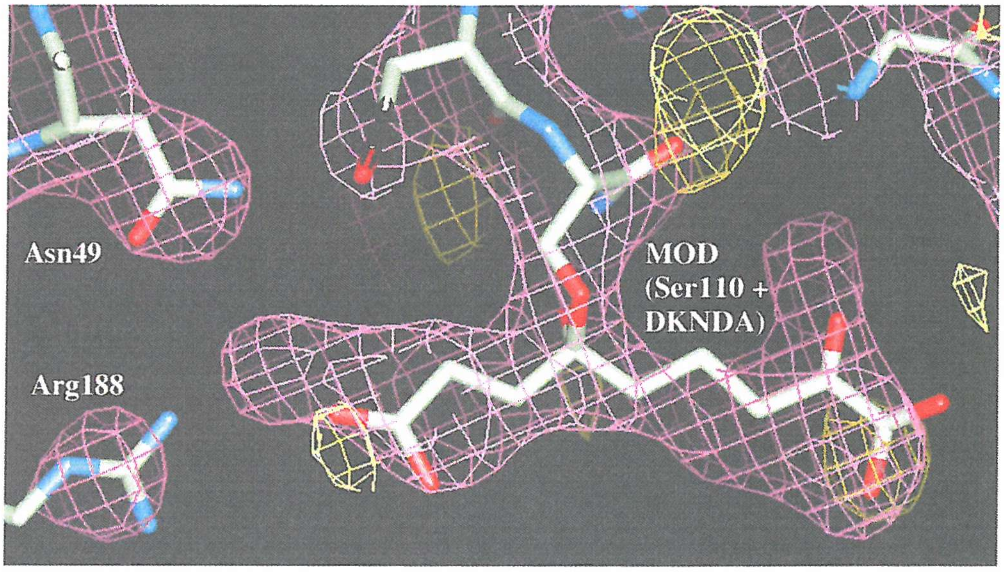


Figure 85: The active site cavity of subunit A subsequent to occupancy refinement and grouped B-factor refinement. This time there is very little negative difference density (yellow) for the ligand, whilst the positive density (purple) is reasonably good. *2Fo-Fc* density was contoured at 1.0σ and negative *Fo-Fc* density at -2.0σ.

Model-building and refinement of the MhpC/DKNDA co-crystal structure is still ongoing, especially regarding the more accurate placement of the ligand within the electron density at the active site. It appears that a water molecule may be present between the C9 carboxylate of the DKNDA molecule and Arg188/Asn49, and as such this has now been incorporated into the model. The following round of refinement saw good *2Fo-Fc* density for this water molecule and so all that remains is to ensure the fit of the ligand to the density is as precise as possible taking into account possible discrepancies between its position in the active sites of subunits A, B, G and H. Current refinement statistics for this model are shown in Table 40.

Resolution range (Å)	129 – 2.8
R_{factor} (%)	24.1
R_{free} (%)	27.1
Number of reflections in working set	71624
Number of reflections in test set	3833
Number of protein atoms	17280
Number of ligand atoms	64
RMS bond length deviation (Å)	0.013
RMS bond angle deviation (°)	1.66
Overall average B-factor (Å^2)	42.4

Table 40: Refinement statistics for the MhpC/DKNDA co-crystal.

6.2.3. Analysis of the MhpC/DKNDA co-crystal structure

The position of the DKNDA molecule within a subunit of MhpC is as predicted from those co-crystal structures obtained with both laevulinic acid and azelaic acid (cf. chapters 5.1. and 6.1.). Namely, one carboxylate group of the ligand occupies a position towards the Arg188 end of the active site and the other carboxylate is located at the other end of the active site cavity near His114. Thus, the ligand runs most of the way through the region enclosed by the core and lid domains of the protein. The DKNDA molecule is covalently bound, through its C6 carbonyl group, to the active site serine residue, Ser110, to form a hemi-ketal intermediate. All of these binding features are shown in Figure 86. However, the ligand is only present in one MhpC subunit per dimer. For instance, subunit A contains DKNDA bound to Ser110, whereas it is not observed in subunit C.

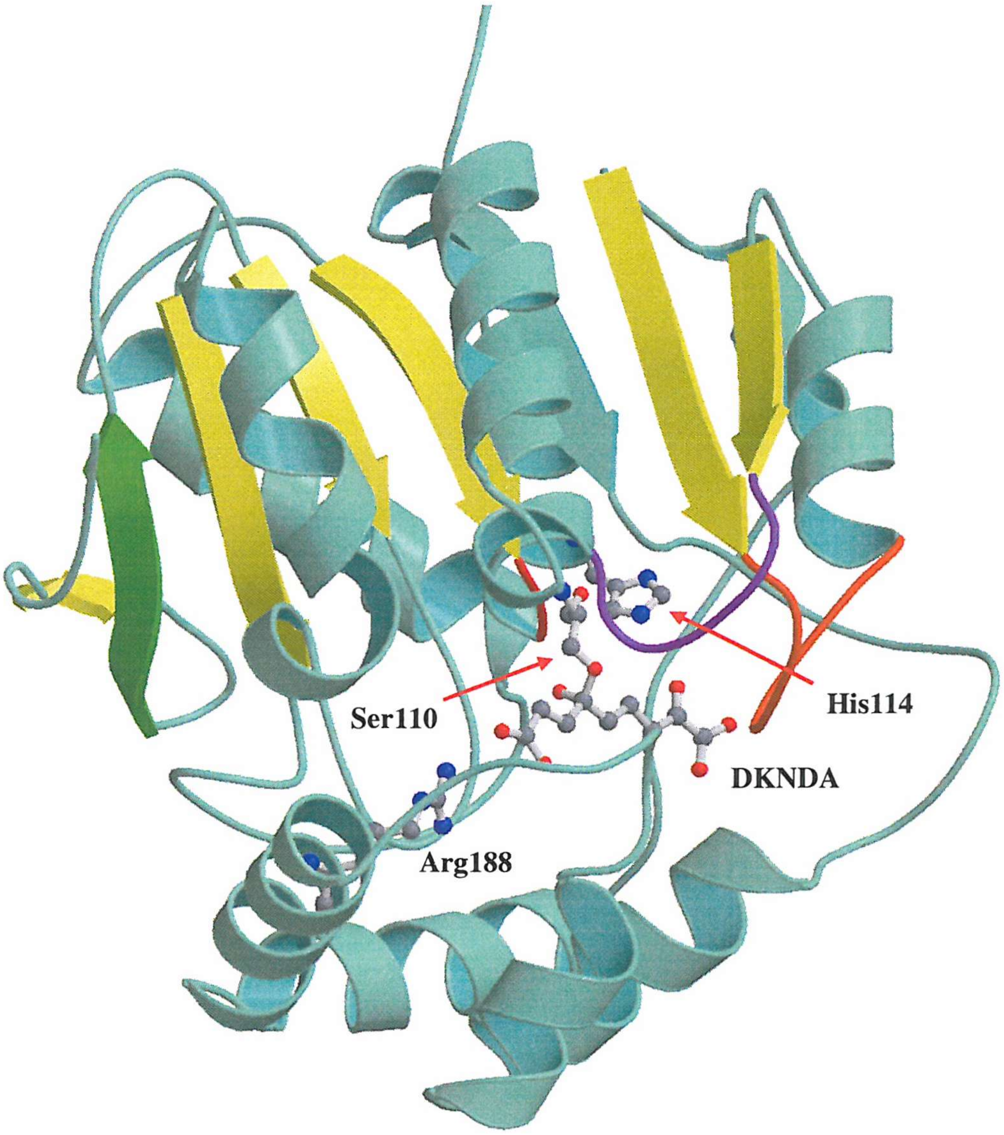


Figure 86: A typical MhpC protomer containing the ligand DKNDA, which is covalently attached to the protein through a serine residue. The ligand occupies the full extent of the active site cavity, located between the core and lid domains, and coordinates to a number of amino acid residues. The ligand was found to occupy only one subunit of each MhpC dimer. Colour scheme as in Figure 35.

6.2.3.1. Binding mode of DKNDA at the MhpC active site

The DKNDA molecule binds almost as predicted from the coordination of the ligands in the previously discussed co-crystal structures. It was predicted that one carboxylate, most probably the C9 carboxylate, would coordinate to the conserved Arg188 and Asn49 residues, and that the C1 carboxylate would bind somewhere near His114. This being the case, then the C6 carbonyl of the ligand would be expected to pass close to the active site histidine and serine residues.

Unexpectedly, once the structure of MhpC co-crystallised with DKNDA had been obtained it could instantly be seen that there were some differences to what was expected. Most significantly, it appeared that the ligand was covalently bound to the active site serine residue. Also, it became obvious that the C9 carboxylate did not coordinate directly to Arg188 and Asn49, but instead this interaction was mediated through a water molecule. This actually mirrors the observation in the azelaic acid co-crystal structure in which a water molecule is co-ordinated to these two residues (Figure 77). Nevertheless, it can be seen that the C1 carboxylate does indeed come close to His114. However, even here interpretation of ligand binding is hampered by the dual occupancy of this residues side-chain (discussed in greater depth in the following sub-chapter).

Comparison of hydrogen-bonding lengths between the catalytic triad residues of those subunits which contain the ligand and those which do not is quite forthcoming. The distance between the aspartate and histidine residues is quite similar for both subunits of the dimer (subunit A – Asp235 OD2 to His263 ND1 = 2.85Å; subunit C – Asp235 OD2 to His263 ND1 = 2.77Å). The major difference comes when comparing the distances between the serine and histidine residues of the catalytic triad. It appears that the bound ligand in some way brings about a shortening of the distance. So that, in subunit A the distance is 3.33Å, whereas in subunit C it is 3.93Å, a quite substantial increase when it comes to the formation of a hydrogen bond.

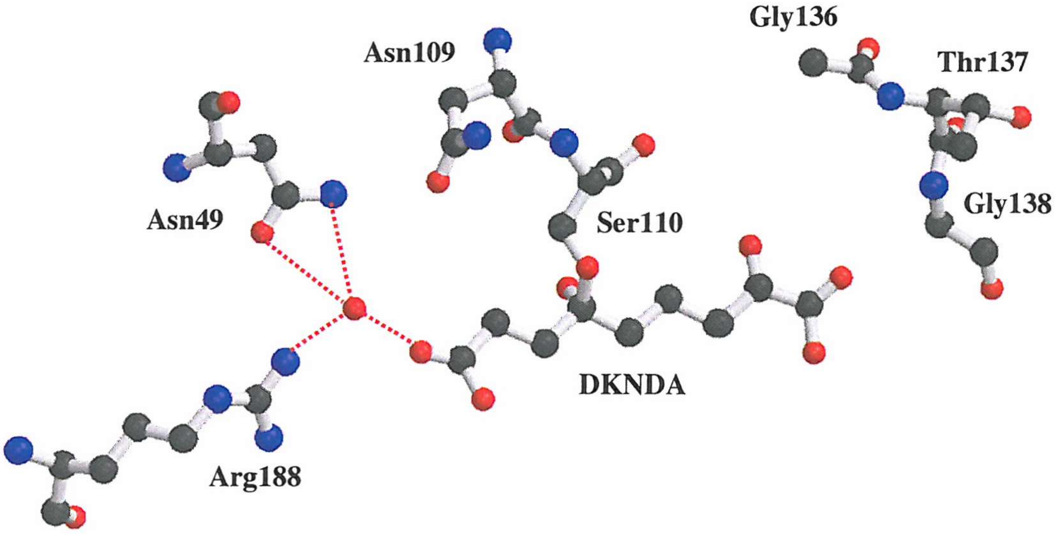


Figure 87: Binding of DKNDA within the active site of subunit A. The C9 carboxylate is coordinated to a water molecule, which itself is coordinated to Arg188 and Asn49. The C6 carbonyl of the ligand is covalently attached to the protein *via* the Ser110 hydroxyl, forming a hemi-ketal intermediate. Hydrogen-bonding is indicated by dotted red lines.

The ligand is bound in a non-planar kinked conformation, unlike the fully extended form expected if it were non-covalently attached to the enzyme. This leads to the C9 carboxylate not extending all the way to the Arg188 and Asn49 residues as expected, and instead facilitates the insertion of a water molecule between these groups (Figure 87). This water molecule can make four hydrogen bonds, two with Asn49 (HOH to Asn49 ND2 = 3.3Å and HOH to Asn49 OD1 = 3.4Å), one with Arg188 (HOH to Arg188 NH2 = 2.5Å) and one with the DKNDA molecule itself (HOH to DKNDA O23 = 2.5Å).

In addition, the electron density surrounding the C6 carbon of the ligand does not appear to be of the correct shape to account for a tetrahedral arrangement around this atom as would be expected. However, it is presumed that the C6 carbon does exist in a tetrahedral arrangement, as there is chemical precedent that a carbonyl can be nucleophilically attacked by an alcohol to form a hemi-ketal compound. The C6 position would then also have a hydroxyl functional group which could position itself in the oxyanion hole of the enzyme. The oxyanion hole appears to be formed by the backbone amide groups of Met111 (Met111 N to DKNDA O25 = 3Å) and Ser40 (Ser40 N to DKNDA O25 = 3.3Å), and the side-chain hydroxyl group of Ser40 (Ser40 OG to DKNDA O25 = 2.8Å), as all these groups are within hydrogen-bonding distance of the hydroxyl (Figure 88). This type of tripartite oxyanion hole has been observed in other α/β-hydrolase fold structures, such as the crystal structure of the acetyl esterase, HerE, from *Rhoddococcus* sp. strain H1 (Zhu *et. al.*, 2003).

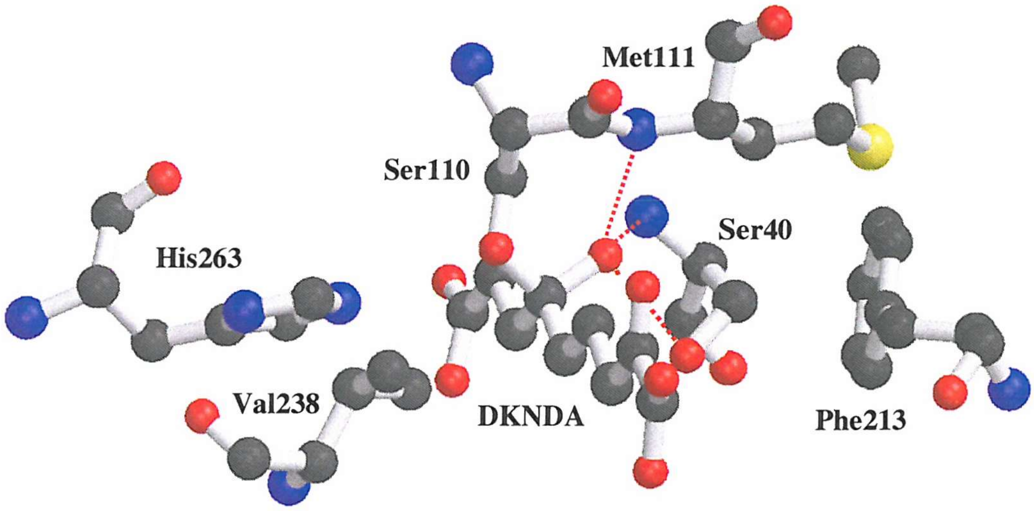


Figure 88: Putative tripartite oxyanion hole formed by the backbone amides of Ser40 and Met111, and the side-chain hydroxyl of Ser40. The C6 hydroxyl of the DKNDA hemi-ketal binds within this cleft. The interaction between part of the ligand, equivalent to the dienol moiety of the RFP, and the side-chains of Phe213 and Thr238 can also be seen. Hydrogen-bonding between the oxyanion hole forming residues and the C6 hydroxyl of the hemi-ketal intermediate is shown by dotted red lines.

Unlike the C1 carboxylate group of azelaic acid which coordinates to the His114 side-chain and a water molecule, itself coordinated by three backbone amide groups, the equivalent functional group of DKNDA does not appear to be coordinated to anything in particular within the active site. However, if the imidazole group of His114 is in a position close to the active site then it would bring the C2 carbonyl group of the ligand very near to this residue, which could have mechanistic implications as discussed in a later section. Of course, the ligand is trapped in this position due to the covalent attachment and so it does not particularly matter if this end of the molecule is less rigidly fixed than the other.

As the ligand only occupies one subunit of each dimer, then it is easy to overlook the other subunit and miss any interesting features that could be revealed by it. As it happens, there is some unusual electron density that extends from Ser110 of the unmodified subunits, as shown in Figure 81B, and it has been attempted to position a water molecule into this density. This water molecule could then interact with the hydroxyl groups of both Ser110 and Ser40, as well as the NE2 of His263 (Figure 89). The position adopted by the water molecule is not within the oxyanion hole however, but is in a very good position for the putative catalytic water that is required to fulfil the *gem*-diol hypothesis. This favours base-catalysed attack of water on the RFP^k, instead of the serine residue itself acting as a nucleophile. The water molecule hydrogen bonds to both the backbone amide and side-chain hydroxyl group of Ser40 (HOH to Ser40 N = 3.2Å; HOH to Ser40 OG = 2.9Å), and so at first appeared to occupy the oxyanion hole. However, the distance from the methionine backbone amide to the water is much greater in this case (HOH to Met111 N = 5.1Å). Moreover, this water molecule could hydrogen bond to the active site serine (HOH to Ser110 OG = 2.7Å) and also possibly coordinate with the triad histidine, although the distance is slightly too long (HOH to His263 NE2 = 3.7Å). It is interesting to note the very close proximity of the two active site serine residues in subunit C (Ser40 OG to Ser110 OG = 3.3Å), which may serve a mechanistic purpose, as discussed later.

The His114 side-chain is free in this subunit to occupy both conformations equally, and it appears that in one position it could be bound to another water molecule within the active site quite close to Ser110. There is no density near the Arg188 and Asn49 residues for either a ligand or a water molecule in subunit C, but neither does it appear that Arg188 could exist in two conformations as was observed in the MhpC/DOHA co-crystal structure. Another residue that exists in two conformations in the DOHA structure, Trp264, also does not appear to vary its position in either subunit of the dimer in this structure, indicating that the manners of binding for these two ligands are quite different in some respects.

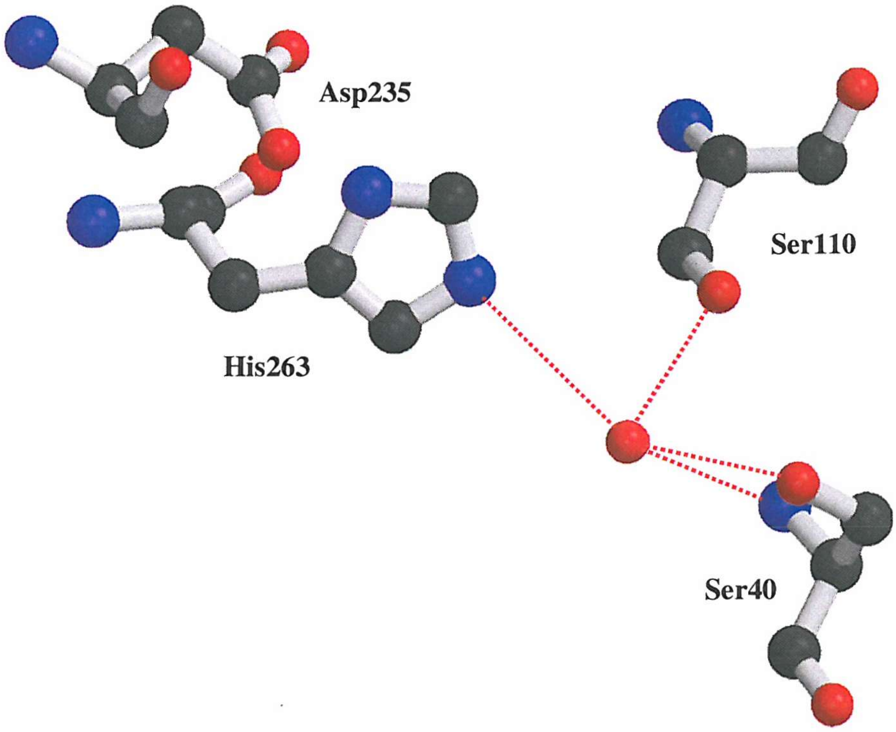


Figure 89: Active site of subunit C illustrating a water molecule in close proximity to the active site serine residue. This water molecule is within hydrogen-bonding distance of Ser110, Ser40, and possibly His263, as shown by the dotted red lines, and could have mechanistic implications for the MhpC catalysed reaction.

6.2.3.2. The implication of His114 in the keto-enol tautomerisation of the RFP

In this structure His114 occurs in a dual conformation in each subunit of the MhpC dimer, including the subunit which contains the DKNDA molecule, unlike what was predicted during initial refinement of the structure (cf. chapter 6.2.2.2.2.). However, it appears that the histidine side-chain occupies one conformation for about 50% of the time and the other conformation for also about 50% of the time in the subunits that do not contain ligand, whereas in those subunits with DKNDA present then the His114 side-chain prefers the orientation away from the active site. The exact proportions spent in either conformation in these subunits are difficult to identify precisely.

The feature of dual occupancy was first observed as a large amount of positive *Fo-Fc* density was present next to this residue in several structures of MhpC, as depicted in Figure 90, and it was at one point thought that a water molecule could occupy this density (Dunn, 2000). However, with the very high σ levels for this density, and the position of the electron density being so close to His114, then it was realised that if the side-chain χ angles were adjusted slightly then this density would be well occupied by the imidazole group. Thus, this change was made during model building and following refinement in this new position then good *2Fo-Fc* density was observed for this residue, but also a large amount of positive difference density for the previous conformation.

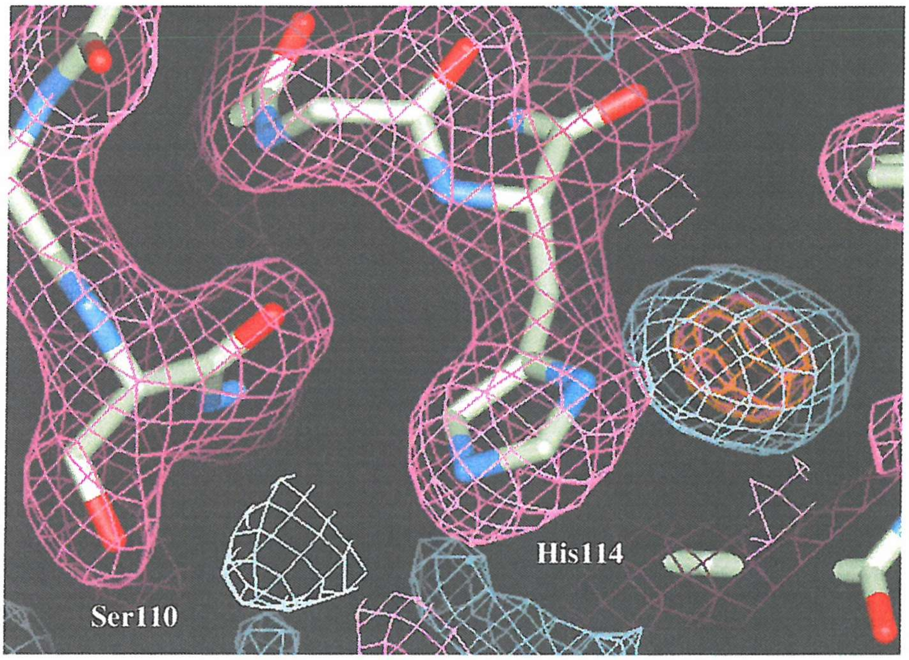


Figure 90: His114 in the orientation towards the active site, but with large amounts of positive difference density next to it, indicating a dual conformation and not a water molecule as previously suspected. The *2Fo-Fc* density (purple) is contoured at 1.0σ, whilst the *Fo-Fc* density is contoured at 2σ (blue) and 5σ (gold).

Therefore, it was realised that this residue could adopt either conformation and, in fact, does appear to do so approximately equally in both positions. In the orientation away from the active site, as also observed in the azelaic acid co-crystal structure, the His114 side-chain can hydrogen bond to a serine residue (Ser85 OG to His114 ND2 = 2.5Å) and also form hydrogen-bonds to either a threonine or proline backbone carbonyl (Pro214 O to His114 NE2 = 3.05Å; Thr137 O to His114 NE2 = 3.21Å). Whereas, in the orientation towards the active site the histidine is in a position where it does not appear to hydrogen-bond to any residue but forms a stacking interaction with Phe213. However, when a ligand is bound within the active site, then this phenylalanine packs against it on one side, with Val238 packed against the other side, as illustrated previously in Figure 88. This occurs on a region of the DKNDA ligand that would be equivalent to the diene moiety of the substrate for MhpC, and so could represent a good hydrophobic interaction in that instance. It also appears that another phenylalanine residue shifts dependent on which conformation His114 adopts. If it is pointing towards the active site then Phe216 is closer to Ser85, whilst when His114 is hydrogen-bonded to Ser85 then the Phe216 side-chain is shifted almost 2Å further away.

It should be noted that the presence of azelaic acid, in that co-crystal structure, almost certainly causes the histidine to remain in the one conformation, as was thought to be the case for the MhpC subunit which contains the DKNDA ligand in this structure. The His114 side-chain cannot move back towards the active site without clashing with the carboxylate group of azelaic acid, whereas the equivalent carboxylate of DKNDA does not occupy an identical position and so there is greater freedom for the histidine to move. It is this unusual feature of the histidine residue, that of its ability to exist equally in two quite contrasting conformations, one of which would bring it close to the C2 hydroxyl of the substrate (equivalent to the C2 carbonyl of DKNDA), which brings about the notion that it may play some part in the catalytic process of the enzyme.

As a result, this has led to the speculation that this residue may be involved in the initial tautomerisation of the RFP, that leads to the formation of the keto-intermediate in the first step of either of the proposed mechanisms. This histidine could abstract the proton from the C-2 hydroxyl group of the substrate, which would be located within H-bonding distance. This could lead to the formation of the keto-intermediate, on the condition that C5 of the substrate could be protonated by a nearby enzyme residue acting as an acid, such as His263, the catalytic triad histidine. This would also leave His263 in the deprotonated state, suitable for its involvement in a further important step of the mechanism, either the abstraction of a proton from a nearby water molecule or directly from the hydroxyl group of the active site serine residue.

Doubt has been cast over whether His114 could function in the keto-intermediate formation described above as it does not appear to be a very well conserved residue amongst related enzymes. In the sequence alignments (Figure 50) there are no other enzymes, such as the closely related BphD enzymes, that have a histidine residue in this position, instead most have an alanine or glycine residue. These enzymes would be expected to utilise similar residues to MhpC in their mechanisms and they are also postulated to undergo this initial keto-enol tautomerisation. However, a recent search of the SWISS-PROT/TrEMBL database (Bairoch & Apweiler, 2000) has highlighted a number of enzymes related to MhpC that do have a histidine residue in this position, namely 2-hydroxy-6-oxo-6-phenyl-hexa-2,4-dienoate hydrolase from *Rhodococcus* sp. (Lau *et. al.*, 1996) and 2-hydroxy-6-(2'-hydroxyphenyl)-6-oxo-hexa-2,4-dienoic acid hydrolase from *Terrabacter* sp. strain DBF63 (Kasuga *et. al.*, 1997). Therefore, there is the possibility that one group of these enzymes has evolved to use His114 to initiate the keto-intermediate formation, whilst the other group utilises another residue within the active site to catalyse the same reaction.

6.2.3.3. The MhpC/DKNDA structure suggests a case for half-site reactivity

As previously mentioned, only one DKNDA molecule was observed per MhpC dimer and as such this may indicate that binding of this ligand could be as a result of half-site reactivity. There are several explanations for the phenomenon of half-site reactivity in enzymes, such as the induced asymmetry and pre-existing asymmetry models, as well as that of allostery, in which binding sites are near symmetry axes of the enzyme dimers so that steric or electrostatic constraints come into play. An example of pre-existing asymmetry accounting for the half-site reactivity of active site cysteine residues of glyceraldehyde-3-phosphate dehydrogenase has been made (Bernhard & MacQuarrie, 1973). The explanation given is that the subunit dimerisation creates a conformational change in one of the two sites that then becomes unfunctional. In the case of induced asymmetry, binding of a ligand to one site induces a conformational change of the second binding site that subsequently becomes inaccessible. A relevant example of such behaviour is the case of 6-diazo-5-oxonorleucine binding to cystidine triphosphate synthetase (Levitzki, *et. al.*, 1971).

The three-dimensional structures of MhpC appear to rule out the hypothesis that the active sites of each subunit in a dimer may be too close together. If the dimeric arrangement is as proposed in Chapter 4.3.2.1. then the active sites of each subunit are located far from each other. However, if either of the other two possible dimers, as based on the crystalline tetrameric arrangement, were indeed correct then at least one of these could possibly support this premise. In the dimeric arrangement described in Chapter 4.3.2.3. the active sites are much closer together, and if the lid domains have to make conformational changes to accommodate the substrate, or ligand, then there is the possibility that they would obstruct one another, and hence only one active site could be exposed at a particular time.

Although the above possibility seems unlikely, nevertheless it is not apparent from the structure described herein how either the pre-existing asymmetry or induced asymmetry models could be applied. No major structural changes seem to be apparent between the subunits of each dimer in the MhpC/DKNDA co-crystal. However, it has been observed that there are a number of subunit discrepancies within the co-crystal structure of MhpC and DOHA (cf. Chapter 5.2.3.1.). In this structure, Trp264 and Arg188 are shown to have different conformations dependent on which protomer of the dimer they are in, and this has been further reported in the literature relating to homologues of MhpC, such as in BphD from *Rhodococcus* sp. strain RHA1 (Nandhagopal *et. al.*, 2001) and the CarC enzyme from *Janthinobacterium* sp. strain J3 (Habe *et. al.*, 2003). As these residues are well conserved amongst members of the C-C hydrolase family then there could also be half-site reactivity in these enzymes that has not yet been reported, owing possibly to the fact that no other related enzyme has had its structure determined with a substrate analogue.

The structural evidence described for half-site reactivity in MhpC can also be backed up by some biochemical evidence. The 'leakiness' of the MhpC active site has previously been reported (Henderson & Bugg, 1997) and was not accounted for with any degree of certainty. In retrospect, taking into account this new structural information, then a case can be made for the kinetic data with respect to the mechanism of the MhpC catalysed reaction. The MhpC catalysed reaction was analysed by stopped flow kinetics which provided evidence for the formation of a discrete keto-intermediate. However, the accurate modelling of the concentration data could only be achieved using a branched kinetic mechanism in which the intermediate was released at a rate comparable to its catalytic turnover. Release of the intermediate from the MhpC active site was also predicted from ^2H exchange experiments (Lam & Bugg, 1997). Thus, approximately 50% of the keto-intermediate tautomer is released from the active site of MhpC and so may be consistent with a half-site reactivity model for this enzyme. That is, if only one subunit of each MhpC dimer can cause cleavage of the substrate, but both subunits can bring about the keto-intermediate formation, then this would satisfactorily explain not only the kinetic data, but may also help to explain the fact that only one DKNDA molecule binds per dimer. However, the crucial point is, what mechanism could possibly bring about this diversity within the dimer? The required catalytic apparatus must be present in both subunits of a dimer to cause keto-enol tautomerisation, but there must be a difference between these subunits so that one active site releases the intermediate, whereas the other active site contains the relevant amino acid residues to bring about C-C cleavage of this intermediate. Thus, as MhpC is a homodimer then some structural change must bring about this disparity, but at present there appears to be very little notable variation between the two subunits of a dimer in the MhpC/DKNDA co-crystal, apart from the varying conformations of the His114 imidazole group.

6.2.3.4. The mechanistic implications of the MhpC/DKNDA co-crystal structure

It is postulated that the binding of DKNDA at the active site would in some way mirror the binding of the natural substrate in its ketonised form (RFP^k), as all it lacks is the diene moiety of the RFP which thus makes it non-cleavable. However, as mentioned previously, the DKNDA molecule is covalently attached to the protein through a serine residue and as such would seem to indicate that this residue is therefore reactive. This residue is Ser110, the putative nucleophile of the active site catalytic triad, which in other serine hydrolases is deemed to be reactive as indicated by its strained conformation on the 'nucleophile elbow'. This strained conformation is also indicated in the Ramachandran plot obtained for this structure where Ser110 occurs in a disallowed region, even when bound to the DKNDA molecule.

Nonetheless, the serine and ligand appear to react to form a hemi-ketal compound, which is not quite equivalent to the acyl enzyme intermediate formed by the serine proteases, and as such does not necessarily implicate this residue as the catalytic nucleophile for the MhpC-catalysed reaction.

However, the position of the ligand in this structure is quite similar when compared to some of the other crystal structures of serine hydrolases that have ligands bound. Normally, the catalytic triad residues occur to the side of the substrate binding pocket, such that the catalytic serine residue points towards the incipient carbonyl that is attacked. In the case of MhpC it appears that the DKNDA ligand is more or less straddled by the Ser110-His263 pair in a unusual conformation, a situation also observed in the covalent modification of Ser195 of bovine γ-chymotrypsin, a typical serine protease, by the pentapeptide TPGVY (PDB accession code 1AB9). In this structure the distance between the acylated serine hydroxyl and the histidine residue of the catalytic triad is 3Å, with the distance between the histidine and the aspartate of the triad being only 2.7Å. These values are not inconsistent with those for the hydrogen-bonding distances of the equivalent residues in the hemi-ketal adduct of MhpC by DKNDA. So, this structural evidence tends to favour the formation of an acyl-enzyme intermediate in the MhpC-catalysed reaction by analogy with the serine proteases.

Yet, confusion still remains over the mechanism adopted by MhpC, as it is unclear from even this structure if the nucleophilic attack at the C-6 carbonyl occurs *via* a water molecule or the hydroxyl of the active site serine. From this co-crystal structure the formation of an acyl-enzyme intermediate is favoured as explained above; but, on the other hand, this is in direct contrast with all previous biochemical work carried out on MhpC which has failed to provide evidence for the existence of an acyl-intermediate (Fleming *et. al.*, 2000).

As can be seen from Figure 85, there is continuous electron density running from the side-chain of Ser110 around the ligand, DKNDA. This would indicate that the ligand is covalently bound to the enzyme, because if it were not then the ligand would make many unfavourable van der Waals contacts with several of the residues within the active site. In particular, if the ligand was instead present as the *gem*-diol form, as would be the case for base-catalysed attack of water at the C6 carbonyl, then the two hydroxyl groups of the *gem*-diol would be approximately 2.3 Å from the oxygen of the serine side-chain which is highly unfeasible. Therefore, the crystallographic evidence does seem to point to the existence of an acyl-enzyme intermediate, and hence that the catalytic triad of MhpC does indeed mimic that of the classical serine proteases.

Furthermore, there is the possibility that an oxyanion hole, extremely important in the reaction mechanism of serine proteases, does exist in MhpC. As described in section 6.2.3.1., within hydrogen-bonding distance of the C6 hydroxyl group of the hemi-ketal adduct are the backbone amides of Ser40 and Met111, the residue adjacent to the catalytic serine residue. Also, the side-chain hydroxyl of Ser40 is within hydrogen-bonding distance of the DKNDA C6 hydroxyl group (Figure 88). This is extremely reminiscent of the action of cutinase, another α/β-hydrolase fold enzyme, whose Ser42 side-chain stabilises the oxyanion transition state (Nicolas *et. al.*, 1996). In cutinase, the oxyanion hole is formed by two main-chain NH groups, that of Ser42 and the residue next to the catalytic serine, and the Ser42 OG. The S42A mutant of cutinase shows a 450-fold reduction in activity, illustrating the importance of this residue. A similar interaction had previously been observed in the crystal structure of *R. miehei* lipase with n-hexylphosphonate ethyl ester bound (Brzozowski *et. al.*, 1991). A direct covalent bond was formed between the nucleophilic Ser144 and the phosphate group of the inhibitor, with the ester hydroxyl group stabilised by not only the backbone amide of Ser82 but also the hydroxyl group of Ser82 as well, an oxyanion interaction similar to that observed in cutinase. This has also been observed in acetylxylan esterase from *Trichoderma reesei*. From the 1.9Å resolution structure of this α/β-hydrolase fold enzyme it was seen that an oxyanion hole could be formed by the main-chain amide group and the side-chain hydroxyl group of a threonine residue (Hakulinen *et. al.*, 2000). Only a small number of enzymes, such as *S. scabies* esterase, subtilisin and papain, also use side-chains to stabilise the oxyanion, and so MhpC could be one of the first C-C hydrolases that this has also been observed in.

6.2.3.5. The proposal of a catalytic tetrad in MhpC suggests a link between the C-C hydrolase family and clan SC of the serine proteases

The foregoing mechanistic conclusions, based on the co-crystal structure of MhpC with the non-cleavable substrate analogue DKNDA, have led to the proposal that MhpC may actually contain a catalytic tetrad, featuring not only the standard triad residues of Asp235, His263 and Ser110 but also the other active site serine residue, Ser40. It can be seen from Figure 50 that this residue is highly conserved amongst a number of C-C hydrolases from single phenol compound degradation pathways, but not BphD, which has either a glycine or an alanine in the equivalent position.

This assumption is similar to the case of the three unrelated clans of serine proteases (chymotrypsin-like; subtilisin-like; and α/β-hydrolase fold) in which there is the development and maintenance of a catalytic tetrad, where the fourth member is a serine or cysteine residue whose side-chain stabilises other residues of the standard triad (Krem & Di Cera, 2001). In clan SC one residue has been identified (Ser57) that fits the conservation and dichotomy criterion for discrete evolutionary markers. Catalytic serine residues utilise TCN codons to a greater degree than non-conserved serine residues, primarily because TCN codons are the primordial serine codons, and as can be seen from the genetic code for MhpC both Ser110 and Ser40 are encoded for by TCN codons (cf. Chapter 4.1.1.). By analysing the sequence motifs surrounding evolutionary markers and active site residues for the SC clan of serine proteases and C-C hydrolases it can be clearly seen that these two distinct sets of enzymes may well be evolutionarily related (Table 41). This is not surprising considering that both of these groups of hydrolase enzymes have adopted the same structural framework, that of the α/β-hydrolase fold, and the same basic catalytic machinery.

SC family	Ser57	Ser146	Asp338 & Val/Cys341	His397
S9	GGPG_**S**X	GW**S**YGG	GXX**D**XN**V/C**	GAG**H**
S10	GGPGCS**S**	GE**S**YAG	GDX**D**XX**V/C**	DEX**H**
C-C hydrolase family	HG**S**GPG	GN**S**M/FGG	GRX**D**XX**V/I**	CG**H**W

Table 41: Sequence motifs surrounding the evolutionary markers (catalytic tetrad) for Clan SC of the serine proteases and a comparison to the proposed tetrad observed in C-C bond hydrolases of the α/β-hydrolase fold family. Note: the numbering system for the serine proteases is not the same as that for C-C hydrolases.

Clan SC of the serine proteases are those enzymes which possess the α/β-hydrolase fold (cf. Chapter 1.9.1.3.), and so it would be expected that they would share features in common with other α/β-hydrolase fold enzymes. However, not all C-C hydrolases are α/β-hydrolase fold enzymes, for example murine fumarylacetoacetate hydrolase (cf. Chapter 1.10.1.; Timm *et. al.*, 1999), and so there are also some distinct differences. The surprising feature, however, is the level of similarity over these short motifs for many members of these disparate families indicating some form of evolutionary relatedness.

So, for instance, the proposed fourth member of the tetrad in clan SC is a serine residue and occurs in a motif which is almost a reverse of that seen in the C-C hydrolases. That is, the GPG motif coming before the serine instead of after the serine. This type of motif is thought to be crucial for oxyanion hole formation and seems to be designed so that the small glycine residues and proline would put this serine in a sharp turn with little steric hindrance, thus bringing it close to the catalytic triad residues at the active sites of these enzymes. This is almost the same as the conserved HGXP motif observed in other α/β-hydrolases (Arand *et. al.*, 1994), which is normally located in a sharp *cis*-proline turn, and is stabilised by the hydrogen bond between the histidine ND1 and the backbone carbonyl oxygen of the following glycine residue. The similarity of the 'nucleophile elbow' amongst many hydrolytic enzymes is very well known, where the triad serine is located on a very sharp turn, and this also has been conserved in that the glycine residues allow the turn to be so sharp with no steric hindrance between residues preceding and following the catalytic serine.

It is also interesting to note that there almost always seems to be a hydrophobic residue following the active site serine residue; either a tyrosine residue for the serine proteases or a methionine or phenylalanine for the C-C hydrolases. This hydrophobicity may or may not play a role in catalysis, but it is known that this residue's backbone amide plays a key role in binding of the oxyanion intermediate in both serine proteases (ie. trypsin) and α/β-hydrolases (ie. cutinase). Evidence that this residue may also play a role in catalysis has been found by mutagenesis studies on the serine hydrolase XylF (Diaz & Timmis, 1995). The substitution of Phe by Met in position 108 of XylF (following the triad Ser107) created an enzyme with increased thermostability and altered substrate specificity.

The motif around the catalytic aspartate residue is very well conserved amongst both families of enzymes, with the glycine, aspartate and either valine or isoleucine (cysteine for clan SC) residues being totally conserved. The residue following the aspartate tends to be either an arginine or a glutamine for the C-C hydrolases, but is more variable in the serine proteases.

The final member of the tetrad, the catalytic histidine residue, seems to have a less well conserved motif between the two classes of enzymes. Even amongst the two families of clan SC serine proteases the motif varies quite dramatically, with only the histidine itself being the common link. For the C-C hydrolases, on the other hand, the tetrapeptide motif containing the catalytic histidine is completely conserved, being in all but one case CGHW (Figure 50). This one abstainee from the norm is the BphD from *Pseudomonas* sp. strain KKS102 (Kimbara *et. al.*, 1989), which appears not to have a triad histidine at all and so sets this enzyme apart from the others in any case.

CHAPTER 7

GENERAL SUMMARY

7.1. Introduction

The main remit for this research was to determine which of the two proposed mechanisms for MhpC was more likely, based on the newly gained structural evidence. Initially, the MAD structure of MhpC had been determined (Dunn, 2000) just prior to co-crystallisation trials of MhpC with substrate and product analogues, thus enabling the use of molecular replacement to solve the co-crystal structures.

The initial assumption was that MhpC would behave mechanistically much like the serine hydrolases, due to them having similar structural motifs such as the 'nucleophile elbow' (Ollis *et. al.*, 1992), and also based on biochemical studies of related C-C hydrolases that belong to the α/β-hydrolase fold family (Diaz & Timmis, 1995; Ahmad *et. al.*, 1995). In these proteins, the carbonyl of the substrate is attacked nucleophilically by the serine residue of the catalytic triad to form an acyl-enzyme intermediate. However, biochemical studies performed on MhpC had led to the conclusion that no acyl-enzyme intermediate was formed and indicated instead that a *gem*-diol intermediate was formed (Fleming *et. al.*, 2000). This was indicative of a mechanism involving base-catalysed attack of water on the substrate, as seen in the aspartic protease family (Veerapandian *et. al.*, 1992).

Thus, it was hoped that by obtaining the structure of native MhpC and of MhpC complexed with substrate and product analogues then some details of the residues involved in substrate binding and catalysis would be revealed.

7.2. Difficulties in crystal growth and data collection

Crystallisation conditions had initially been discovered for MhpC (Dunn, 1999; Mohammed, 2000) and so numerous ligands were soaked into preformed MhpC crystals. Upon subsequent analysis it was found that these ligands did not bind and so co-crystallisation was attempted instead, in which the ligand and protein were incubated together prior to crystallisation. Again difficulties were faced, in that the usual conditions no longer produced crystals, possibly due to the influence of the added chemicals. Therefore, several screening strategies were carried out for each protein/ligand combination which gave fairly good quality crystals for several of the co-crystal complexes.

Afterwards, when diffraction data were collected on several of these co-crystals it was revealed that most of them were poor diffractors of X-rays, even at synchrotron sources. A typical diffraction pattern is shown in Figure 91, and illustrates the poor quality of diffraction shown by a large number of the co-crystals. All the co-crystal structures of MhpC have only been resolved to between 2.5 and 3.4Å, and even the native structure has only been determined at 2.1Å. The spots of the diffraction patterns tended to be more 'comet-shaped' than spherical, resulting in rather a smudgy image. This made initial indexing of diffraction data difficult, as well as further data processing, as shown by the rather large merging R_{factor}s obtained for many of the co-crystals.

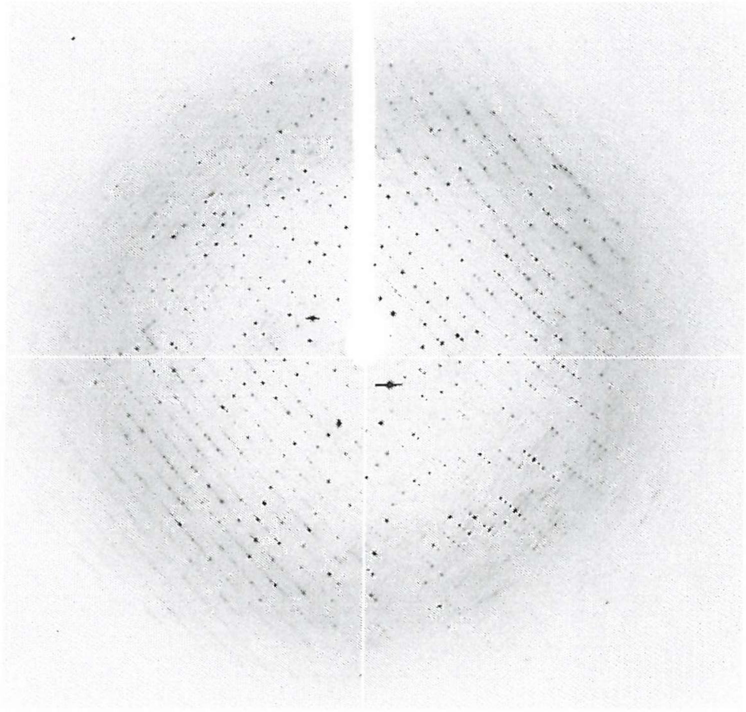


Figure 91: Example of a typical diffraction image from any of the MhpC/ligand co-crystal data collections showing 'smudginess' of the pattern which made data processing awkward.

7.3. Analysis of the native MhpC structure indicated some unusual features

It has been revealed that MhpC does indeed possess the α/β-hydrolase fold, comprising eight mostly parallel β-strands interspersed with α-helices in the repeating pattern βαβ. The protein also consists of two domains, a core domain comprising residues 1-133 and 217-284 and a lid domain of residues 134 to 216. This α-helical lid domain is a feature of several analogous lipases and is thought to act in allowing substrate entry to the active site of the enzyme, which is sandwiched between the two domains. Space-filling models of an MhpC protomer have revealed that this may well be the case, as there is little room for access of the rather large substrate of MhpC into the active site cavity without some conformational change occurring to assist in its subsequent docking at the catalytic binding site.

The catalytic triad residues are located in the positions expected of an α/β-hydrolase, with the serine (Ser110) occurring on the 'nucleophile elbow' between strand β5 and helix αC, and the aspartate (Asp235) and histidine (His263) being located on loops following strands β7 and β8, respectively. These residues are located together in an elongated funnel between the core and lid domains of the protein. A loop following strand β3, comprising residues 37 to 45, extends down into this funnel from the core domain and comes close to the region occupied by the catalytic triad residues. From sequence alignments with related enzymes it was seen that many residues in this loop are either conserved or invariant.

MhpC exists as a homodimer in solution (Lam & Bugg, 1997) and a very tight association has been observed in the crystal structure that could account for this oligomerisation. This interaction is the anti-parallel association of the β8 strands of two subunits to form an extended 16-strand β-sheet, a feature observed in several related enzymes, such as CumD from *P. fluorescens* IP01 (Fushinobu *et. al.*, 2002). In MhpC this dimer interface buries 2236Å^2 and forms numerous van der Waals contacts, hydrogen bonds and salt bridges and is likely to be relatively stable in solution. The other two postulated MhpC dimers, based on the tetrameric assembly observed in the crystal structure, are less likely to be stable in solution and so can presumably be discounted as the physiological association.

In terms of the MhpC mechanism then the native structure is not especially revealing. Due to only a few lower resolution (2.5 to 3.4Å) structures being available, initially, it was observed that the hydrogen-bonding distances between the catalytic triad residues were quite different to those seen for the typical serine hydrolases. The aspartate to histidine distance was of a suitable magnitude, being around 2.7Å, but the histidine to serine distance was in the order of 4.0Å, far too great for a hydrogen-bond to form. This led to the proposal that unless a conformational shift occurred on substrate binding, to bring these residues closer, then the catalytic triad may actually be just a dyad of Asp-His and that the serine residue may have a non-nucleophilic role. This would support the mechanism which proposes that a water molecule could act as the nucleophile in the MhpC-catalysed reaction. This appeared to be substantiated by numerous other reports in the literature of unusual changes to the typical triads of these types of enzymes. An example being that of the human acyl thioesterase I in which the position normally occupied by the serine hydroxyl is actually occupied by a water molecule (Devedjiev *et. al.*, 2000). It was also proposed in Chapter 4.3.3.3 that as there is another serine residue, Ser40, close to Ser110 in the active site then both of these residues could be involved in the stabilisation of a *gem*-diol intermediate formed by the base-catalysed attack of water on the substrate. However, the more recent structure obtained for native MhpC, to 2.1Å resolution, cast some seeds of doubt on these suggestions. This structure revealed that His263 and Ser110 could indeed form a hydrogen bond, as the distance was found to be around 3.2Å. It also revealed that there were two water molecules within the active site of the enzyme. One water appeared to occupy a region thought to be analogous to the oxyanion hole of other serine hydrolases, formed by the backbone amides of Met111 and Ser40, as well as the side-chain hydroxyl of Ser40, which is more unusual. The other water was within hydrogen-bonding distance of both His263 and Ser110 and so could be catalytically important. That is, it could be the water which hydrolyses either the *gem*-diol or acyl-enzyme intermediate, and which ultimately leads to cleavage of the MhpC substrate.

Nevertheless, an unexpected observation was revealed by some structures of MhpC. This being that there was a possible covalent modification of the active site serine residue, as evinced by a continuation of the *2Fo-Fc* density from this residues side-chain. From earlier studies it appeared to be due to neither a water molecule nor a metal ion, which led to the assumption that it was some form of covalent modification (Mohammed, 2001). Numerous efforts were made to determine exactly what was causing this extraneous electron density, as described in Chapter 4.4. There was no point mutation at this position as determined by sequencing of the *mhpC* gene; whilst tryptic digest followed by LC-MS analysis to discover a possible modification of this residue also led to little success in its identification. A number of amino acid residues are naturally modified by enzymes, such as when kinases bring about phosphorylation of a side-chain. However, although there appears a chance that MhpC may actually be phosphorylated, as determined by ^{31}P NMR spectroscopy, there is no clear evidence of this being the case at Ser110. Therefore, for the time being, this phenomenum still remains a mystery, even in the cases of other enzymes which appear to share this unusual feature, such as the C-C hydrolase BphD (Nandhagopal *et. al.*, 2001) and the bromoperoxidase from *S. aureofaciens* (Hecht *et. al.*,1994), themselves both enzymes of the α/β-hydrolase fold family. It does seem likely that modification of this reactive serine is common. An arsenic adduct of an acetyl esterase has recently been observed in which the As(III) arsenical is bound to the active site serine residue (Zhu *et. al.*, 2003). Thus, as cacodylate has been present in some crystallisation conditions then this may account for the unexpected density in some MhpC structures, but cannot account for all of them. It is also interesting to note from the co-crystal structures solved thus far that when a ligand is bound at the active site then the unusual density on the Ser110 side-chain disappears in some cases but is still present in others, the former being true for the MhpC/Laevulinic acid structure and the latter for the MhpC/Azelaic acid structure.

Another interesting observation from the MhpC structure was a possible explanation for the inhibition of serine hydrolases by some thiol-directed reagents. It had been reported that MhpC was inhibited by bulky sulfhydryl directed reagents but not by smaller reagents (Lam & Bugg, 1997). A feature also revealed in the related C-C hydrolase XylF (Diaz & Timmis, 1995) and the diketo acid hydrolase from beef liver (Hsiang *et. al.*, 1972). By analysis of the sequence alignments in Figure 50 it was revealed that Cys261 of MhpC was a fairly well conserved residue amongst this class of enzyme and that the close proximity of this cysteine to the catalytic histidine could account for the inhibitory effects observed. A bulky reagent would be expected to either perturb this region of the protein or block the catalytic histidine and hence disrupt catalysis by the enzyme, whereas this appears less likely for a smaller reagent. Thus, it seems logical that by analogy this could be extended to other enzymes of this family in which the unusual feature of serine hydrolase inhibition by thiol-directed compounds have been observed.

7.4. Analysis of the MhpC co-crystal structures

The co-crystal structures of MhpC with two substrate analogues and two product analogues have been determined. The substrate analogues, azelaic acid and DKNDA, vary in that they contain differing functional groups but are both basically nona-1,9-dioic acids. In fact, azelaic acid does only constitute this framework, whereas DKNDA has keto groups in the C2 and C6 positions and so would be expected to be similar to the keto-intermediate of the natural substrate. The difference between DKNDA and RFP^k is the lack of the diene functionality which makes the former compound non-cleavable and so very useful for investigating the active site binding of MhpC. The product analogues, laevulinic acid and DOHA, are quite dissimilar in that each one mimics one of the different products of the MhpC-catalysed reaction. It was thought that laevulinic acid would mimic part of the succinate product of the reaction, as it possesses one of the carboxylate groups, but instead of the other carboxylate it has a carbonyl. This could actually be a bonus as this carbonyl group would then be equivalent to the C6 carbonyl of the RFP. The other product of the MhpC reaction is 2-hydroxypenta-2,4-dienoate and the ligand used to imitate this was DOHA. This compound contains the carboxylate group in the correct position but lacks the dienol moiety. Instead, owing to the extended length of this compound, it can actually be said to mimic part of the RFP in both orientations, as it possesses a carbonyl group in both the C4 and C6 positions. From the C1 carboxylate of the RFP there is a keto group in the C6 position and from the C9 carboxylate end there is a keto group in the C4 position. Thus, the DOHA ligand serves a dual purpose and might even possibly occupy two varying positions within the MhpC binding site.

The four co-crystal structures have all contributed some knowledge towards how the MhpC substrate could bind within the active site. In some instances they have revealed amino acid residues that could possibly co-ordinate to the functional groups of the substrate and in other respects they have revealed residues that may be involved catalytically.

7.4.1. The co-crystal structures have delineated the MhpC active site pocket

The structure of MhpC with laevulinic acid has implicated two residues in the active site that could co-ordinate to a carboxylate group, those residues being Asn49 and Arg188. Both of these residues are well conserved amongst enzymes in this family. The carbonyl group of this ligand comes into close proximity with Ser110 of the catalytic triad, which is important as this functional group is equivalent to that of the C6 carbonyl of the RFP. An interesting observation from this structure was that the ligand forced itself between the histidine and serine residues of the triad, so increasing the distance between them to about 4.4Å.

Another carboxylate binding site has been inferred from the structure of MhpC with azelaic acid, this time at the other end of the active site cavity towards a histidine residue, His114. The position of this histidine's side-chain was different to that in the laevulinic acid structure, as it had swung away from the active site region due to the influence of the azelaic acid carboxylate. A GGGTGG motif of part of the MhpC polypeptide chain near this histidine allows for the freedom of movement observed. The position of the other carboxylate binding site was occupied by a water molecule in this structure, perhaps owing to the covalent modification of the active site serine blocking part of the binding cavity from the remainder of the azelaic acid molecule.

Thus, it was assumed that both carboxylate binding sites for the RFP had possibly been determined. The co-crystal structure with DKNDA confirmed these sites, but also contained some interesting differences. For one thing the ligand was covalently bound to the active site serine and for another the water molecule at the Arg188 end of the active site was still present, which was not anticipated. Also, the structure with DKNDA seemed to have exposed the presence of a tripartite oxyanion hole, as predicted from the native structure. In the DKNDA structure the C6 hydroxyl group of the hemi-ketal adduct appeared to occupy the oxyanion hole, which is not unexpected for this class of enzyme.

The DOHA structure could also have a covalent adduct to the active site serine as electron density was also observed from the side-chain of Ser110 that ran continuously with that for the putative ligand. The distances between Ser110 and His263 in both the azelaic acid structure and the DOHA structure were in the region of 3.8 to 4.0Å, still too long for a good hydrogen bond to form.

The DOHA co-crystal structure, despite not yet having a model built into the electron density at the active site, also appeared to confirm the involvement of the Asn49, Arg188 and His114 residues in substrate binding. There was *2Fo-Fc* density at both ends of the active site, but in different subunits of each dimer, possibly indicating that a ligand is co-ordinated to the Arg188 site in one protomer and at the His114 site in the other protomer of the dimer. Another difference revealed between the subunits of each dimer was the orientation of the His114 side-chain, which in one subunit occupied a position equivalent to that in the laevulinic acid structure and in the other subunit of the dimer occupied a position equivalent to that seen in the azelaic acid structure. Furthermore, the DOHA structure revealed conformational shifts for both Arg188 and Trp264 dependent upon which subunit of the dimer was examined and so these residues may have mechanistic possibilities, a fact which is backed up in the literature. For instance, differences in the positions of equivalent residues have been observed in CumD (Fushinobu *et. al.*, 2002), BphD (Nandhagopal *et. al.*, 2001) and CarC (Habe *et. al.*, 2003), all C-C hydrolases related to MhpC.

7.4.2. The co-crystal structures have revealed some mechanistic possibilities

The binding mode of the ligands, as described above, have led to some interesting insights into possible mechanistic details for the MhpC-catalysed reaction. The covalent modification of MhpC by DKNDA being a leading example. The DKNDA co-crystal structure has led to a whole array of amino acids being identified that could be involved in co-ordination and catalysis of the RFP, as illustrated in Figure 92, as it is presumed that this ligand can mirror the binding of the natural substrate for MhpC.

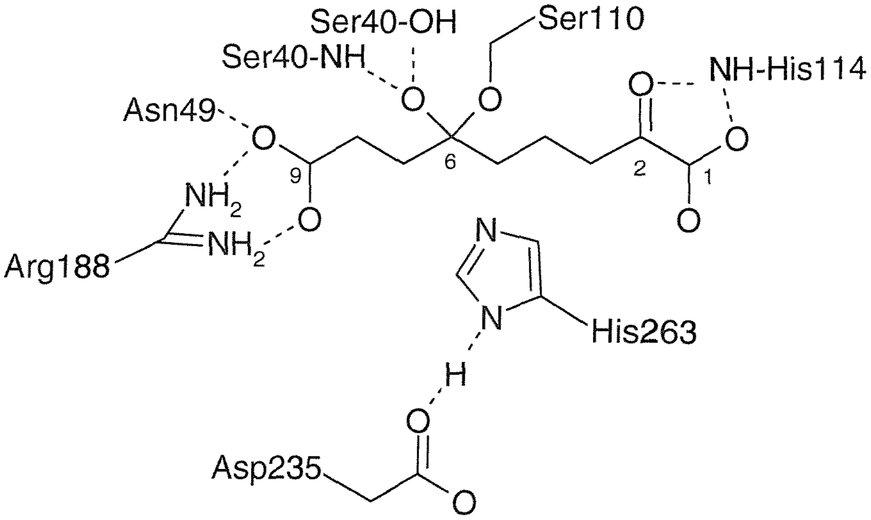


Figure 92: The residues that could be crucial in substrate binding and catalysis by MhpC, as revealed by co-ordination of the non-cleavable analogue DKNDA within the active site of the enzyme. This may resemble the tetrahedral intermediate formed by the natural substrate, in which case there would be a double bond between the C3 and C4 atoms of the ligand.

Most structures have indicated that His114 can have two conformations, the exception being the azelaic acid structure in which the histidine imidazole is fixed in position by one of the ligands carboxylate groups. This dual occupancy appears to show some diversity within different subunits of the MhpC dimer in both the DOHA and DKNDA structures and as such it is thought that this residue could play a catalytic role. This histidine could act as the base which deprotonates the C2 hydroxyl of the RFP to form the keto-intermediate, but only if another residue within the active site could concomitantly supply a proton to C5 of the substrate. One residue that might possibly be in the correct position to carry out this proton transfer is the active site histidine itself, His263. In which case, the histidine NE2 would then be in a deprotonated state, and as such it would be realistically feasible that this residue could then abstract a proton from either a water molecule or the Ser110 side-chain hydroxyl, which is a necessary step in the MhpC-catalysed reaction.

Another mechanistic discovery made from the DKNDA structure was that MhpC might have half-site reactivity as it was observed that only one subunit per dimer had the ligand bound. This cannot be accounted for by analysis of the structures of each subunit in the dimer as the only noticeable difference is the conformational change of the His114 side-chain. Nevertheless, there is some biochemical evidence to back up the assumption that half-site reactivity might be occurring. It had previously been found that 50% of the keto-intermediate was released by the MhpC active site. There is a 'leakiness' in this enzyme which has not been sufficiently accounted for (Henderson & Bugg, 1997). Therefore, if both subunits of each dimer could catalyse the initial enol-keto tautomerisation of the substrate, but then only one subunit had the necessary catalytic apparatus to bring about cleavage, then this could explain both the kinetic and structural evidence. A mechanistic proposal based on the evidence presented in the sections above is illustrated in Figure 93.

As only one subunit of each dimer contained the DKNDA ligand it was found that the other subunit contained a water molecule within the active site of the enzyme, but not occupying the putative oxyanion hole. In the protomer containing the DKNDA adduct the C6 hydroxyl of the hemi-ketal occupies a position where it can hydrogen bond to both backbone amides of Met111 and Ser40, as well as the side-chain OG of Ser40. This type of tripartite oxyanion binding cleft has been reported for other enzymes of the α/β-hydrolase fold. However, the water molecule in the other protomer does not occupy an equivalent position to the hemi-ketal hydroxyl, even though it interacts with both the NH and OG atoms of Ser40, for it is much further from the NH of Met111. Instead, the water is located almost equidistant from the OG atoms of Ser110 and Ser40, and quite near NE2 of His263, in a position not dissimilar to that observed for one of the water molecules in the 2.1Å native structure of MhpC. Another difference noticed between the catalytic residues of the subunits of each dimer was the distances between Ser110 and His263 of the catalytic triad. In those subunits which contained a water molecule then the distance was too great for a hydrogen bond to form (~3.9Å). Conversely, in those subunits where Ser110 had been modified by DKNDA, the separation of Ser110 and His263 was within hydrogen-bonding range (~3.2Å). This raises the possibility that on binding of substrate the protein might undergo a conformational change which could bring these residues closer together, which could be of great mechanistic significance.

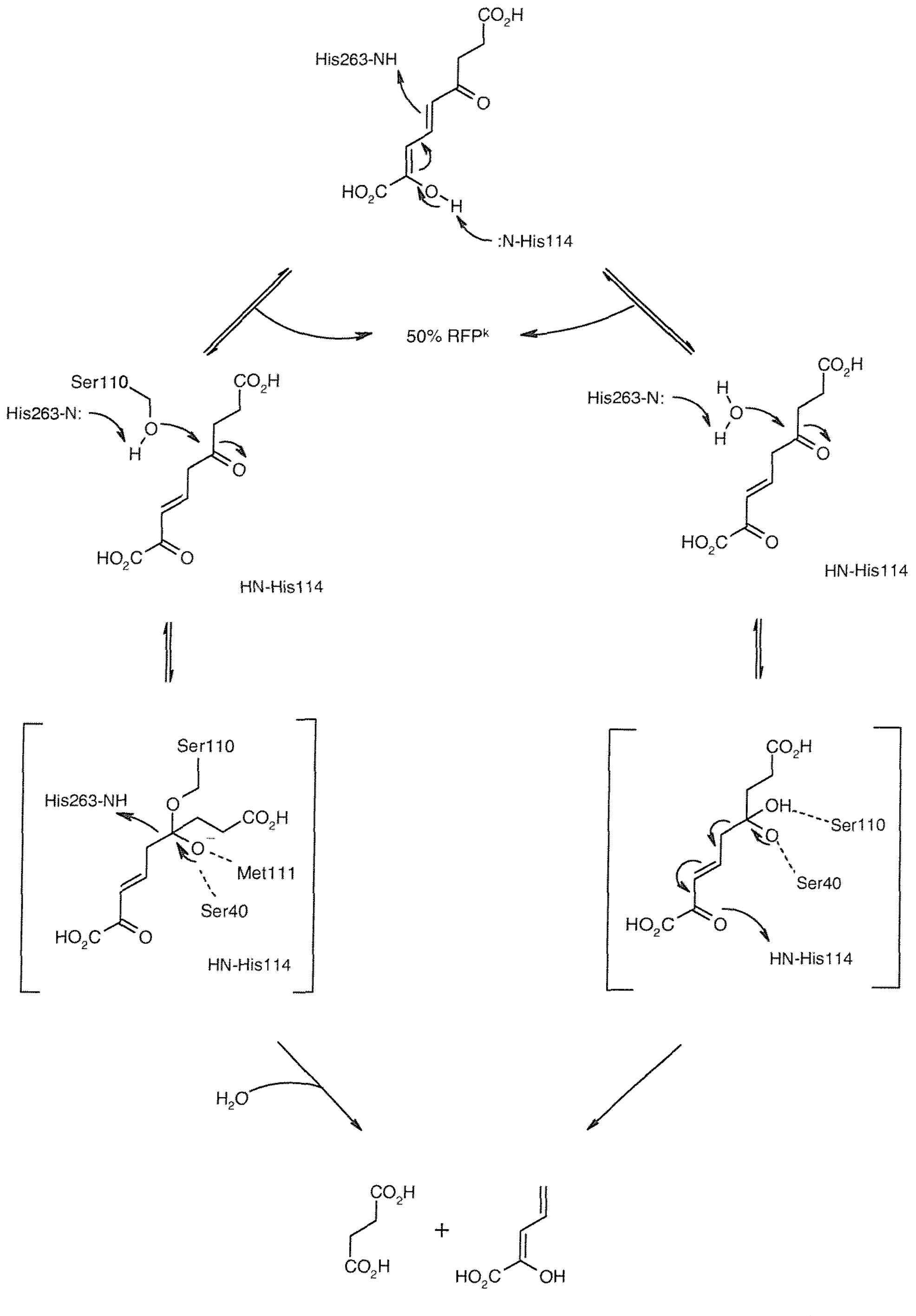


Figure 93: Possible mechanism for MhpC involving initial ketonisation of substrate by the tandem action of His114 and His263, followed by release of 50% of the keto-intermediate by the enzyme owing to half-site reactivity. The cleavage of the remaining bound substrate is brought about by either base-catalysed (His263) attack of water or nucleophilic attack of the active site serine. In both cases Ser40 could be crucial in stabilising the intermediate formed.

Thus, although a great deal has been revealed about the mechanism of MhpC by examination of the co-crystal structures the key mechanistic question, that of whether a water molecule or the active site serine acts as the nucleophile, has not been resolved. There is much crystallographic evidence that supports the biochemical evidence, such as that obtained for the half-site reactivity of MhpC, but also there are some major contradictions, such as the formation of a hemi-ketal intermediate which suggests MhpC might possibly be capable of forming an acyl-enzyme intermediate. However, the rate of hemi-ketal formation would be expected to be much slower than the catalytic rate and so the binding of DKNDA may not implicitly reflect that of the natural substrate. It had also previously been discovered that KNDA, an analogue of DKNDA, was a slow-binding irreversible inhibitor of MhpC (Fleming *et. al.*, 2000).

What does appear clear from these crystallographic studies on MhpC is that there is some evidence to suggest the maintenance of a catalytic tetrad within the C-C hydrolase family, which once again links these enzymes evolutionarily with clan SC of the serine proteases, along with their shared protein fold.

7.5. The crystal structures of MhpC could lead to further mechanistic studies

Obviously the main mechanistic question still needs to be unravelled and so further investigations into the MhpC-catalysed reaction, based on information obtained from the crystallographic studies, can be commenced to resolve these issues. The DKNDA co-crystal structure has identified several key active site residues which are positioned to participate in substrate binding and catalysis. The roles of these residues could be explored by site-directed mutagenesis, with those mutants of particular interest being of the catalytic triad Ser110, Asp235 and His263. Mutating these residues to alanines would presumably make them catalytically inactive for C-C cleavage, as was observed for XylF (Diaz & Timmis, 1995). However, they may retain some activity for the keto-enol tautomerisation step, which could be investigated. The putative acid/base residue His114 could also be mutated as this amino acid could play a key role in not only the initial keto-intermediate formation but also in the feature of half-site reactivity. Residues involved in substrate binding, such as Arg188 and Asn49, could also be modified to alter substrate specificity, which might have far-reaching consequences for bioremediation and biotransformation endeavours. The other residue of the catalytic tetrad, Ser40, which is thought to be involved in oxyanion stabilisation, could be replaced, whereby if a notable decrease in rate is obtained then it may be possible to co-crystallise the natural substrate with the enzyme.

CHAPTER 8

REFERENCES

Abeles, R. H., Frey, P. A. & Jencks, W. A. (1992) *Biochemistry*, Jones & Bartlett.

Abramowicz, D. A. (1990) *Crit. Rev. Biotechnol.* **10**, 241-249.

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

Akhtar, M. (2003) Personal Communication.

Albertson, N. H. & Koomey, M. (1993) *Mol. Microbiol.* **9**, 1203-1211.

Andreoni, V. & Bestetti, G. (1986) *Appl. Environ. Microbiol.* **52**, 930-934.

Arai, H., Yamamoto, T., Ohishi, T., Shimizu, T., Nakata, T. & Kudo, T. (1999) *Microbiology (UK)* **145**, 2813-2820.

Arand, M., Grant, D. F., Beetham, J. K., Friedberg, T., Oesch, F. & Hammock, B. D. (1994) *FEBS Lett.* **338**, 251-256.

Arpigny, J. L., Feller, G. & Gerday, C. (1993) *Biochim. Biophys. Acta* **1171**, 331-333.

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

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1995) In *Short Protocols in Molecular Biology* 3rd Edition, Wiley & Sons, Inc.

Bachovchin, W. W. & Roberts, J. D. (1978) *J. Am. Chem. Soc.* **100**, 8041-8047.

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

Bairoch, A., Gasteiger, E. & Gattiker, A. (2003) *Applied Bioinform.* **1**, 107-108.

Barnes, M. R., Duetz, W. A. & Williams, P. A. (1997) *J. Bacteriol.* **179**, 6145-6153.

Bateman, R. L., Bhanumoorthy, P., Witte, J. F., McClard, R. W., Grompe, M. & Timm, D. E. (2001) *J. Biol. Chem.* **276**, 15284-15291.

Bedard, D. L., Unterman, R., Bopp, L. H., Brennan, M. J. & Haberl, M. L. (1986) *Appl. Environ. Microbiol.* **51**, 761-768.

Bernhard, S. A. & MacQuarrie, R.-A. (1973) *J. Mol. Biol.* **74**, 73-78.

Bild, G. S. & Morris, J. C. (1984) *Arch. Biochem. Biophys.* **235**, 41-47.

Blakley, E. R. & Simpson, F. J. (1964) *Can. J. Microbiol.* **10**, 175-185.

Blow, D. (2002) In *Outline of Crystallography for Biologists*, Oxford University Press.

Blow, D. M., Birktoft, J. J. & Hartley, B. S. (1969) *Nature* **221**, 337-340.

Blundell, T. L. & Johnson, L. N. (1976) In *Protein Crystallography*, Academic Press, N. Y.

Bolin, B. (1973) The carbon cycle. In *Chemistry in the Environment: Readings from Scientific American* (Hamilton, C. L., ed.) W. H. Freeman, San Francisco.

Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A. & Uson, I. (2001) *Structure* **9**, 483-491.

Botos, I., Meyer, E., Nguyen, M., Swanson, S. M., Koomen, J. M., Russell, D. H. & Meyer, E. F. (2000) *J. Mol. Biol.* **298**, 895-901.

Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248-254.

Brady, L., Brzozowski, A., Derewenda, Z. S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J. P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thin, L. & Menge, U. (1990) *Nature* **343**, 767-770.

Branden, C.-I. (1980) *Q. Rev. Biophys.* **13**, 317-338.

Brenner, C., Bevan, J. & Fuller, R. J. (1994) *Methods Enzymol.* **244**, 152-167.

Brenner, S. (1988) *Nature* **334**, 528-530.

Brock, D. J. H. & Williamson, D. H. (1968) *Biochem. J.* **110**, 677-682.

Brzozowski, A. M., Derewenda, U., Derewenda, Z. S., Dodson, G. G., Lawson, D. M., Turkenburg, J. P., Bjorkling, F., Huge-Jensen, B., Patkar, S. A. & Thim, L. (1991) *Nature* **351**, 491-494.

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

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., GrosseKunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998) *Acta Cryst.* **D54,** 905-921.

Bugg, T. D. H. (1993) *Biochim. Biophys. Acta* **1202**, 258-264.

Bugg, T. D. H. & Winfield, C. J. (1998) *Nat. Prod. Rep.* 513-530.

Bullock, T. L., Breddam, K. & Remington, S. J. (1996) *J. Mol. Biol.* **255**, 714-725.

Burlingame, R. & Chapman, P. J. (1983) *J. Bacteriol.* **155**, 113-121.

Burlingame, R., Wyman, L. & Chapman, P. J. (1986) *J. Bacteriol.* **168**, 55-64.

Carson, R. L. (1962) 'Silent Spring', Houghton Mifflin, Boston, USA.

Carter, P. & Wells, J. A. (1987) *Science* **237**, 394-399.

Cavin, J.-F., Barthelmebs, L. & Divies, C. (1997) *Appl. Environ. Microbiol.* **63**, 1939-1944.

CCP4: Collaborative Computational Project, Number 4 (1994) *Acta Cryst.* **D50,** 760-763.

Chamberlain, E. M. & Dagley, S. (1968) *Biochem. J.* **110**, 755-763.

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

Chen, C.-L. & Chang, H.-M. (1985) Chemistry of lignin biodegradation. In *Biosynthesis and Biodegradation of Wood* (Higuchi, T., ed.) Academic, San Diego. pp 535-556.

Chen, J. C.-H., Miercke, L. J. W., Krucinski, J., Starr, J. R., Saenz, G., Wang, X., Spilburg, C. A., Lange, L. G., Ellsworth, J. L. & Stroud, R. M. (1998) *Biochemistry* **37**, 5107-5117.

Connors, W. M. & Stotz, E. (1949) *J. Biol. Chem.* **178**, 881-890.

Cooper, R. A. & Skinner, M. A. (1980) *J. Bacteriol.* **143**, 302-306.

Corey, D. R. & Craik, C. S. (1992) *J. Am. Chem. Soc.* **114**, 1784-1790.

Cousin, X., Hotelier, T., Giles, K., Toutant, J. P. & Chatonnet, A. (1998) *Nucleic Acids Res.* **26**, 226-228.

Craik, C. S., Roczniak, S., Largman, L. & Rutter, W. J. (1987) *Science* **237**, 909-913.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K. & Doctor, B. P. (1993) *Protein Sci.* **2**, 366-382.

Dagley, S. (1975) *Essays Biochem.* **11,** 81-138.

Dagley, S., Chapman, P. J. & Gibson, D. T. (1965) *Biochem. J.* **97**, 643-650.

David, F., Bernard, A.-M., Pierres, M. & Marguet, D. (1993) *J. Biol. Chem.* **268**, 17247-17252.

Davies, J. I. & Evans, W. C. (1964) *Biochem. J.* **91**, 251-261.

Davison, A. J. (1992) *Virology* **186**, 9-14.

Deane, S. M., Robb, F. T., Robb, S. M. & Woods, D. R. (1989) *Gene* **76**, 281-288.

De Simone, G., Galdiero, S., Manco, G., Lang, D., Rossi, M. & Pedone, C. (2000) *J. Mol. Biol.* **303**, 761-771.

De Simone, G., Menchise, V., Manco, G., Mandrich, L., Sorrentino, N., Lang, D., Rossi, M. & Pedone, C. (2001) *J. Mol. Biol.* **314**, 507-518.

Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. Z. & Derewenda, Z. S. (2000) *Structure* **8**, 1137-1146.

Diaz, E., Ferrandez, A. & Garcia, J. L. (1998) *J. Bacteriol.* **180**, 2915-2923.

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

Dixon, G. H., Go, S. & Neurath, H. (1956) *Biochim. Biophys. Acta* **19**, 193-200.

Doherty, D. G. (1955) *J. Am. Chem. Soc.* **77**, 4887-4892.

Dorn, E. & Knackmuss, H.-J. (1978) *Biochem. J.* **174**, 73-84.

Drenth, J. (1994) In *Principles of Protein X-ray Crystallography*, Springer-Verlag, N. Y. Inc.

Duggleby, C. J. & Williams, P. A. (1986) *J. Gen. Microbiol.* **132**, 717-726.

Dunn, G. (1998) Personal Communication.

Dunn, G. (1999) Personal Communication.

Dunn, G. (2000) PhD Thesis, University of Southampton (England).

Eaton, R. W. (1996) *J. Bacteriol.* **178**, 1351-1362.

Eaton, R. W. (1997) *J. Bacteriol.* **179**, 3171-3180.

Eaton, R. W., Selifonova, O. V. & Gedney, R. M. (1998) *Biodegradation* **9**, 119-132.

Egloff, M. P., Marguet, F., Buono, G., Verger, R., Cambillau, C. & van Tilbeurgh, H. (1995) *Biochemistry* **34**, 2751-2762.

Endo, F., Kubo, S., Awata, H., Kiwaki, K., Katoh, H., Kanegae, Y., Saito, I., Miyazaki, J.-I., Yamamoto, T., Jakobs, C., Hattori, S. & Matsuda, I. (1997) *J. Biol. Chem.* **272**, 24426-24432.

Evans, W. C., Smith, B. S. W., Fernley, H. N. & Davies, J. I. (1971) *Biochem. J.* **122**, 543-551.

Faix, O., Mozuch, M. D. & Kirk, T. K. (1985) *Holzforschung* **39**, 203-208.

Ferrandez, A., Garcia, J. L. & Díaz, E. (1997) *J. Bacteriol.* **179**, 2573-2581.

Flegel, T. W., Meevootisom, V. & Kiatapan, S. (1982) *J. Ferment. Technol.* **60**, 473-475.

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

Franken, S. M., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1991) *EMBO J.* **10**, 1297-1302.

Fraser, M. E., Strynadka, N. C., Bartlett, P. A., Hanson, J. E. & James, M. N. G. (1992) *Biochemistry* **31**, 5201-5214.

Fulop, V., Bocskei, Z. & Polgar, L. (1998) *Cell* **94**, 161-170.

Furukawa, K. & Miyazaki, T. (1986) *J. Bacteriol.* **166**, 392-398.

Fushinobu, S., Saku, T., Hidaka, M., Jun, S.-Y., Nojiri, H., Yumane, H., Shoun, H., Omori, T. & Wakagi, T. (2002) *Protein Sci.* **11**, 2184-2195.

Garavito, R. M., Rossman, M. G., Argos, P. & Eventoff, W. (1977) *Biochemistry* **16**, 5065-5070.

Garman, E. F. (1996) Crystallographic data collection. In *Methods in Molecular Biology - Crystallographic Methods and Protocols* (Jones, C., Mulloy, B. & Sanderson, M. R., Eds.) **56,** Humana Press Inc., N. J., USA.

Ghosh, D., Sawicki, M., Lala, P., Erman, M., Pangborn, W., Eyzaguirre, J., Gutierrez, R., Jornvall, H. & Thiel, D. J. (2001) *J. Biol. Chem.* **276**, 11159-11166.

Giacovazzo, C., Monaco, H. L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G. & Catti, M. (1992) In *Fundamentals of Crystallography* (Giacovazzo, C., Ed.) Oxford University Press, N. Y.

Gold, A. M. & Fahrney, D. (1964) *Biochemistry* **6**, 783-791.

Gold, M. H., Glenn, J. K., Mayfield, M. B., Morgan, M. A. & Kutsuki, H. (1983) Biochemical and genetic studies on lignin degradation by *Phanerochaete chrysosporium.* In *Recent Advances in Lignin Biodegradation Research* (Higuchi, T., Chang, H.-M. & Kirk, T. K., eds.) University, Tokyo. pp 219-232.

Habe, H., Kasuga, K., Nojiri, H., Yamane, H. & Omori, T. (1996) *Appl. Env. Microbiol.* **62**, 4471-4477.

Habe, H., Morii, K., Fushinobu, S., Nam, J.-W., Ayabe, Y., Yoshida, T., Wakagi, T., Yamane, H., Nojiri, H. & Omori, T. (2003) *Biochem. Biophys. Res. Commun.* **303**, 631-639.

Håkansson, K., Wang, A. H.-J. & Miller, C. G. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 14097-14102.

Hakulinen, N., Tenkanen, M. & Rouvinen, J. (2000) *J. Struct. Biol.* **132**, 180-190.

Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W. & Bolin, J. T. (1995) *Science* **270**, 976-980.

Harayama, S. & Rekik, M. (1993) *Mol. Gen. Genet.* **239**, 81-89.

Harrad, J. & Jones, K. (1992) *Chem. Brit.* **28**, 1110-1112.

Hartley, B. S. & Kilby, B. A. (1954) *Biochem. J.* **56**, 288-297.

Hayaishi, O. (1966) *Bacteriol. Rev.* **30**, 720-?

Hayaishi, O., Schimazono, H., Katagiri, M. & Saito, Y. (1956) *J. Am. Chem. Soc.* **78**, 5126-5127.

Hecht, H. J., Sobek, H., Haag, T., Pfeifer, O. & Van Pee, K.-H. (1994) *Nature Struct. Biol.* **1**, 532-537.

Heider, J. & Fuchs, G. (1997) *Eur. J. Biochem.* **243**, 577-596.

Heikinheimo, P., Goldman, A., Jeffries, C. & Ollis, D. L. (1999) *Structure* **7**, R141-146.

Helliwell, J. R. (1992) In *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press.

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

Henderson, R. (1970) *J. Mol. Biol.* **54**, 341-346.

Hendrickson, W. A. & Ogata, C. M. (1997) *Methods in Enz.* **276,** 494-523.

Hofer, B., Backhaus, J. & Timmis, K. N. (1994) *Gene* **144**, 9-16.

Hofer, B., Eltis, L. D., Dowling, D. N. & Timmis, K. N. (1993) *Gene* **130**, 47-55.

Horn, J. M., Harayama, S. & Timmis, K. N. (1991) *Mol. Microbiol.* **5**, 2459-2474.

Hsiang, H. H., Sim, S. S., Mahuran, D. J. & Schmidt, D. E., Jr. (1972) *Biochemistry* **11**, 2098-2102.

Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114-122.

Huisman, G. W., Wonink, E., Meima, R., Kazemier, B., Terpstra, P. & Witholt, B. (1991) *J. Biol. Chem.* **266**, 2191-2198.

Janssen, D. B., Pries, F., van der Ploeg, J., Kazemier, B., Terpstra, P. & Witholt, B. (1989) *J. Bacteriol.* **171**, 6791-6799.

Jany, K.-D., Lederer, G. & Mayer, B. (1986) *FEBS Lett.* **199**, 139-144.

Johnson, M. S. (1990) Modelling and Biocomputing Group, Centre for Biotechnology, PO Box 123, FIN-20521, Turku, Finland. MALIGN: © 1990 Kramsku, Finland.

Jorquera, R. & Tanguay, R. M. (1997) *Biochim. Biophys. Res. Commun.* **232**, 42-48.

Kabisch, M. & Fortnagel, P. (1990) *Nucleic Acids Res.* **18**, 3405-3405.

Kahne, D. & Still, W. C. (1988) *J. Am. Chem. Soc.* **110**, 7529-7534.

Kasuga, K., Nojiri, H., Yamane, H., Kodama, T. & Omori, T. (1997) *J. Ferment. Bioeng.* **84**, 387-399.

Kern, H. W. (1983) *Holzforschung* **37**, 109-115.

Kikuchi, Y., Yasukochi, Y., Nagata, Y., Fukuda, M. & Takagi, M. (1994) *J. Bacteriol.* **176**, 4269-4276.

Kim, K. K., Song, H. K., Shin, D. H., Hwang, K. Y., Choe, S., Yoo, O. J. & Suh, S. W. (1997) *Structure* **5**, 1571-1584.

Kimbara, K., Hashimoto, T., Fukuda, M., Koana, T., Takagi, M., Oishi, M. & Yano, K. (1989) *J. Bacteriol.* **171**, 2740-2747.

Kirk, T. K. (1984) Degradation of lignin. In *Microbial Degradation of Organic Compounds* (Gibson, D. T., ed.) Dekker, New York.

Kirk, T. K. & Adler, E. (1970) *Acta. Chem. Scand.* **24**, 3379-3390.

Kirk, T. K. & Cowling, E. B. (1984) *Adv. Chem. Ser.* **207**, 455-487.

Kirk, T. K. & Farrell, R. L. (1987) *Ann. Rev. Microbiol.* **41**, 465-505.

Kirk, T. K. & Shimada, M. (1985) Lignin biodegradation: The microorganisms involved and the physiology and biochemistry of degradation by white-rot fungi. In *Biosynthesis and Biodegradation of Wood* (Higuchi, T., ed.) Academic, San Diego. pp 579-605.

Kirk, T. K. & Tien, M. (1983) Biochemistry of lignin degradation by *Phanerochaete chrysosporium*: Investigations with non-phenolic model compounds. In *Recent Advances in Lignin Biodegradation Research* (Higuchi, T., Chang, H.-M. & Kirk, T. K., eds.) University, Tokyo. pp 233-245.

Kleywegt, G. J. & Jones, T. A. (1994) *Acta Cryst.* **D52**, 826-828.

Kossiakoff, A. A. & Spencer, S. A. (1981) *Biochemistry* **20**, 6462-6473.

Kratzschmar, J., Haendler, B., Langer, G., Boidol, W., Bringmann, P., Alagon, A., Donner, P. & Schleuning, W.-D. (1991) *Gene* **105**, 229-237.

Krem, M. M. & Di Cera, E. (2001) *EMBO J.* **20**, 3036-3045.

Lack, L. (1961) *J. Biol. Chem.* **236**, 2835-2840.

Lajoie, C. A., Chen, S.-Y., Oh, K.-C. & Strom, P. F. (1992) *Appl. Environ. Microbiol.* **58**, 655-663.

Lajoie, C. A., Zylstra, G. J., DeFlaun, M. F. & Strom, P. F. (1993) *Appl. Environ. Microbiol.* **59**, 1735-1741.

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

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

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Cryst.* **33**, 491-497.

Lau, P. C. K., Garnon, J., Labbe, D. & Ying, W. (1996) *Gene* **171**, 53-57.

Leisola, M., Ulmer, D., Haltmeier, T. & Fiechter, A. (1983) *Appl. Microbiol. Biotechnol.* **17**, 117-120.

Lenz, H., Wunderwald, P. & Eggerer, H. (1976) *Eur. J. Biochem.* **65**, 225-236.

Leslie, A. G. W. (1992) In *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* **26,** Daresbury Laboratory, Warrington, U. K.

Leslie, A. G. W. (1994) MOSFLM User Guide.

Levitt, M. & Perutz, M. F. (1988) *J. Mol. Biol.* **201**, 751-754.

Levitzki, A., Stallcup, W. B. & Koshland, D. E., Jr. (1971) *Biochemistry* **10**, 3371-3378.

Lhongi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997) *J. Mol. Biol.* **268**, 779-799.

Liao, D., Breddam, K., Sweet, R. M., Bullock, T. & Remington, S. J. (1992) *Biochemistry* **31**, 9796-9812.

Liao, D. & Remington, S. J. (1990) *J. Biol. Chem.* **265**, 6528-6531.

Lindblad, B., Lindstedt, S. & Steen, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4641-4645.

Lineweaver, H. & Burk, D. (1934) *J. Am. Chem. Soc.* **56,** 658-666.

Ljungcratz, P., Carlsson, H., Mansson, M.-O., Buckel, P., Mosbach, K. & Biilow, L. (1989) *Biochemistry* **28**, 8786-8792.

Longenecker, J. B. & Snell, E. E. (1955) *J. Biol. Chem.* **213**, 229-235.

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

Markley, J. L. (1979) Catalytic groups of serine proteases: NMR investigations. In *Biological Applications of Magnetic Resonance* (Shulman, R. G., ed.) Academic Press, New York. pp. 397-461.

Masai, E., Sugiyama, K., Iwashita, N., Shimizu, S., Hauschild, J. E., Hatta, T., Kimbara, K., Yano, K. & Fukuda, M. (1997) *Gene* **187**, 141-149.

Mason, J. R. & Cammack, R. (1992) *Annu. Rev. Microbiol.* **46**, 277-305.

Massey, V. (1994) *J. Biol. Chem.* **269**, 22459-22462.

Masui, Y., Coleman, J. & Inouye, M. (1983) Multipurpose expression cloning vehicles in Escherichia coli. In *Experimental Manipulation of Gene Expression* (Inouye, M., ed.) Academic Press, New York. pp. 15-32.

Matheis, G. & Whitaker, J. R. (1984) *Int. J. Biochem.* **16**, 867-873.

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

Matthews, B. W. (1972) *Macromolecules* **5**, 818-819.

Matthews, B. W. (1977) X-ray structure of proteins. In *The Proteins, Vol. 3* (Neurath, H. & Hill, R. L., eds.) Academic Press, New York. pp. 404-590.

Maymo-Gayell, X., Chien, Y., Gossett, J. M. & Zinder, S. H. (1997) *Science* **276**, 1568-1571.

McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. (1995) *Nature* **375**, 549-554.

McGrath, M. E., Osawa, A. E., Barnes, M. G., Clark, J. M., Mortara, K. D. & Schmidt, B. F. (1997) *FEBS Lett.* **413**, 486-488.

Medynski, D. (1992) *Bio/Technology* **10**, 1002-1006.

Meister, A. & Greenstein, J. P. (1948) *J. Biol. Chem.* **175**, 573-588.

Menn, F. M., Zylstra, G. J. & Gibson, D. T. (1991) *Gene* **104**, 91-94.

Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. (1988) *Acta Crystallogr.* **B44**, 26-38.

Mohammed, F. (1999) Personal Communication.

Mohammed, F. (2001) PhD Thesis, University of Southampton (England).

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

Nagainis, M. P., Pu, W., Cheng, B., Taylor, K. E. & Schmidt, D. E. (1981) *Biochim. Biophys. Acta* **657**, 203-211.

Nakayama, K. (1994) *Methods Enzymol.* **244**, 167-175.

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

Nandhagopal, N., Yamada, A., Hatta, T., Masai, E., Fukuda, M., Mitsui, Y. & Senda, T. (2001) *J. Mol. Biol.* **309**, 1139-1151.

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

Neurath, H. (1984) *Science* **224**, 350-357.

Nicholls, A., Sharp, K. & Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296.

Nicolas, A., Egmond, M., Verrips, C. T., de Vlieg, J., Longhi, S., Cambillau, C. & Martinez, C. (1996) *Biochemistry* **35**, 398-410.

Nordland, Z. & Shingler, V. (1990) *Biochim. Biophys. Acta* **1049**, 227-230.

Ogata, C. M. (1998) *Nat. Struct. Biol.* **5,** 638-640.

Ohlendorf, D. H., Lipscomb, J. D. & Weber, P. C. (1988) *Nature* **336**, 403-405.

Ohlendorf, D. H., Orville, A. M. & Lipscomb, J. D. (1994) *J. Mol. Biol.* **244**, 586-608.

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

Orville, A. M., Elango, N., Lipscomb, J. D. & Ohlendorf, D. H. (1997) *Biochemistry* **36**, 10039-10051.

Orville, A. M., Lipscomb, J. D. & Ohlendorf, D. H. (1997) *Biochemistry* **36**, 10052-10066.

Ouchiyama, N., Miyachi, S. & Omori, T. (1998) *J. Gen. Appl. Microbiol.* **44**, 57-63.

Pathak, D. & Ollis, D. L. (1990) *J. Mol. Biol.* **214**, 497-525.

Perona, J. J. & Craik, C. S. (1995) *Protein Sci.* **4**, 337-360.

Philips, R. S. & Dua, R. K. (1991) *J. Am. Chem. Soc.* **113**, 7385-7388.

Phillips, M., Kaplan, A., Rutter, W. & Bartlett, P. A. (1992) *Biochemistry* **31**, 959-963.

Platt, A., Shingler, V., Taylor, S. C. & Williams, P. A. (1995) *Microbiology* **141**, 2223-2233.

Pokorny, D., Steiner, W. & Ribbons, D. W. (1997) *TIBTech* **15**, 291-296.

Polgar, L. (1992) *FEBS Lett.* **311**, 281-284.

Porter, R. & Clark, S. (1985) *Enzymes in Organic Synthesis,* Ciba Foundation Symposium III, Pittman.

Prieto, M. A., Diaz, E. & Garcia, J. L. (1996) *J. Bacteriol.* **178**, 111-120.

Rambaud, M., Bakasse, M., Duguay, G. & Villieras, J. (1988) *Synthesis* **7,** 564-566.

Ravdin, R. G. & Crandall, D. I. (1951) *J. Biol. Chem.* **184**, 137-149.

Raven, M. L., Harel, M., Pang, Y. P., Silman, I., Kozikowski, A. P. & Sussman, J. L. (1997) *Nat. Struct. Biol.* **4**, 57-63.

Rawlings, N. D. & Barrett, A. J. (1994) *Methods Enzymol.* **244**, 19-61.

Read, R. (2000) http://www-structmed.cimr.cam.ac.uk/Course/MolRep/molrep.html

Rees, D. C., Lewis, M. & Lipscomb, W. (1983) *J. Mol. Biol.* **168**, 367-387.

Reineke, W. & Knackmuss, H. J. (1988) *Annu. Rev. Microbiol.* **42**, 263-287.

Reinhammer, B. (1984) Laccase. In *Copper Proteins and Copper Enzymes* (Lontie, R., ed.) CRC, Boca Raton, Fla. pp 1-35.

Remington, S. J. & Breddam, K. (1994) *Methods Enzymol.* **244**, 231-249.

Rhodes, G. (2000) In *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models* 2nd Edition, Academic Press, Inc.

Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2619-2623.

Robertson, T. A. (1999) Personal Communication.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J. Powers, J. C. & Wilcox, P. E. (1972a) *Biochemistry* **11**, 2439-2449.

Robertus, J. D., Kraut, J., Alden, R. A. & Birktoft, J. J. (1972b) *Biochemistry* **11**, 4293-4303.

Robinson, A., Edwards, K. J., Carr, P. D., Barton, J. D., Ewart, G. D. & Ollis, D. L. (2000) *Acta Cryst.* **D56**, 1376-1384.

Roget, J. & Calvet, F. (1962) *An. Fis. Quim.* **58**, 357-?

Rossman, M. G. & Blow, D. M. (1962) *Acta Cryst.* **15,** 24-31.

Saku, T., Fushinobu, S., Jun, S.-Y., Ikeda, N., Nojiri, H., Yamane, H., Omori, T. & Wakagi, T. (2002) *J. Biosci. Bioeng.* **93**, 568-574.

Sanvoisin, J., Langley, G. J. & Bugg, T. D. H. (1995) *J. Am. Chem. Soc.* **117**, 7836-7837.

Schrag, J. D. & Cygler, M. (1993) *J. Mol. Biol.* **230**, 575-591.

Schrag, J. D. & Cygler, M. (1997) *Methods Enzymol.* **284**, 85-107.

Schrag, J. D., Li, Y., Wu, S. & Cygler, M. (1991) *Nature* **351**, 761-764.

Seidah, N. G. & Chretien, M. (1994) *Methods Enzymol.* **244**, 175-188.

Senda, T., Sugiyama, S., Narita, H., Yamamoto, T., Kimbara, K., Fukuda, M., Sato, M., Yano, K. & Mitsui, Y. (1996) *J. Mol. Biol.* **255**, 735-752.

Shanley, M. S., Harrison, A., Parales, R. E., Kowalchuk, G., Mitchell, D. J. & Ornston, L. N. (1994) *Gene* **138**, 59-65.

Shaw, E., Mares-Guia, M. & Cohen, W. (1965) *Biochemistry* **4**, 2219-2226.

Shimada, Y., Sugihara, A., Izumi, T. & Tominaga, Y. (1990) *J. Biochem.* **107**, 703-707.

Shingler, V., Powlowski, J. & Marklund, U. (1992) *J. Bacteriol.* **174**, 711-724.

Simpson, T. (1995) http://enviro.mond.org/9511/951104.html

Slabas, A. R., Windust, J. & Sidebottom, C. M. (1990) *B. J. Lett.* **269**, 279-280.

St.-Louis, M. & Tanguay, R. M. (1997) *Hum. Mutat.* **9**, 291-299.

Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T. & Wilson, I. A. (1993) *Structure* **1**, 83-93.

Steiert, J. G. & Crawford, R. L. (1985) *Trends Biotechnol.* **3**, 300-305.

Steitz, T. A. & Shulman, R. G. (1982) *Annu. Rev. Biophys. Bioeng.* **11**, 419-444.

Steller, I., Bolotovsky, R. & Rossman, M. G. (1997) *J. Appl. Cryst.* **30,** 1036-1040.

Streyer, L. (1994) *Biochemistry*, W. H. Freeman.

Strickland, S. & Massey, V. (1973) *J. Biol. Chem.* **248**, 2944-2952.

Suen, W. C. & Spain, J. C. (1993) *J. Bacteriol.* **175**, 1831-1837.

Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. & Silman, I. (1991) *Science* **253**, 872-879.

Sutherland, J. B., Crawford, D. L. & Speedie, M. K. (1982) *Mycologia* **74**, 511-513.

Tanguay, R. M., Jorquera, R., Poudrier, J. & St.-Louis, M. (1996) *Acta Biochim. Pol.* **43**, 209-216.

Thomas, T., Cooper, A., Bussey, H. & Thomas, G. J. (1990) *J. Biol. Chem.* **265**, 10821-10824.

Timm, D. E., Mueller, H. A., Bhanumoorthy, P., Harp, J. M. & Bunick, G. J. (1999) *Structure Fold. Des.* **7**, 1023-1033.

Timmis, K. N., Steffan, R. J. & Unterman, R. (1994) *Annu. Rev. Microbiol.* **48**, 525-557.

Tramper, J., von der Plas, H. C. & Linko, P. (1985) *Biocatalysis in Organic Synthesis*, Elsevier.

Tsai, S. C., Miercke, L. J. W., Krucinski, J., Gokhale, R., Chen, J. C. H., Foster, P. G., Cane, D. E., Khosla, C. & Stroud, R. M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14808-14813.

Uppenberg, J., Ohrner, N., Norin, M., Hult, K., Kleywegt, G. J., Patkar, S., Waagen, V., Anthonsen, T. & Jones, T. A. (1995) *Biochemistry* **34**, 16838-16851.

Vagin, A. & Teplyakov, A. (1997) *J. Appl. Cryst.* **30,** 1022-1025.

Van de Ven, W. J., Roebroek, A. J. & Van Duijnhoven, H. L. (1993) *Crit. Rev. Oncogen.* **4**, 115-136.

Van der Meer, J. R., DeVos, W. M., Harayama, S. & Zehnder, A. J. B. (1992) *Microbiol. Rev.* **56**, 677-694.

Veerapandian, B., Cooper, J. B., Sali, A. & Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, B. D. & Hoover, D. J. (1992) *Protein Sci.* **1**, 322-328.

Verschueren, K. H., Franken, S. M. Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1993) *J. Mol. Biol.* **232**, 856-872.

Voet, D. & Voet, J. G. (1995) *Biochemistry, 2nd edition*, Wiley & Sons, Inc.

Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I. & Joachimiak, A. (1999) *Acta Cryst.* **D55,** 1726-1732.

Walsh, T. A. & Ballou, D. P. (1983) *J. Biol. Chem.* **258**, 4413-4421.

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

Wang, Y., Rawlings, M., Gibson, D. T., Labbe, D., Bergeron, H., Brousseau, R. & Lau, P. C. K. (1995) *Mol. Gen. Genet.* **246**, 570-579.

Warshel, A., Naray-Szabo, G., Sussman, F. & Hwang, J. K. (1989) *Biochemistry* **28**, 3629-3637.

Watanabe, T., Inoue, R., Kimura, N. & Furukawa, K. (2000) *J. Biol. Chem.* **275**, 31016-31023.

Webb, E. C. (1992) *IUBMB Enzyme Nomenclature.* Academic Press, New York.

Weeks, C. M. & Miller, R. (1999) *Acta Cryst.* **D55**, 492-500.

Wei, Y. Y., Contreras, J. A., Sheffield, P., Osterlund, T., Derewenda, U., Kneusel, R. E., Matern, U., Holm, C. & Derewenda, Z. S. (1999) *Nat. Struct. Biol.* **6**, 340-345.

Westbrook, E. M. & Naday, I. (1997) *Methods in Enz.* **276,** 244-268.

Whitman, C. P., Aird, B. A., Gillespie, W. R. & Stolowich, N. J. (1991) *J. Am. Chem. Soc.* **113**, 3154-3162.

Williams, P. A. & Sayers, J. R. (1994) *Biodegradation* **5**, 195-217.

Williams, P. A. & Shaw, L. E. (1997) *J. Bacteriol.* **179**, 5935-5942.

Winkler, F. K., D'Arcy, A. & Hunziker, W. (1990) *Nature* **343**, 771-774.

Wright, C. S., Alden, R. A. & Kraut, J. (1969) *Nature* **221**, 235-242.

Yamada, A., Kishi, H., Sugiyama, K., Hatta, T., Nakamura, K., Masai, E. & Fukuda, M. (1998) *Appl. Environ. Microbiol.* **64**, 2006-2012.

Yang, H.-H., Effland, M. & Kirk, T. K. (1980) *Biotechnol. Bioeng.* **22**, 65-77.

Young, C. L., Barker, W. C., Tomaselli, C. M. & Dayhoff, M. O. (1978) In *Atlas of Protein Sequence and Structure Vol. 5 Suppl. 3* (Dayhoff, M. O., ed.) National Biomedical Research Foundation, Washington D. C. pp. 73.

Zeikus, J. G., Wellstein, A. L. & Kirk, T. K. (1982) *FEMS Microbiol. Lett.* **15**, 193-197.

Zhu, X., Larsen, N. A., Basran, A., Bruce, N. C. & Wilson, I. A. (2003) *J. Biol. Chem.* **278**, 2008-2014.

APPENDIX A

ANALYSIS OF MhpC/DKNDA CO-CRYSTAL SYMMETRY

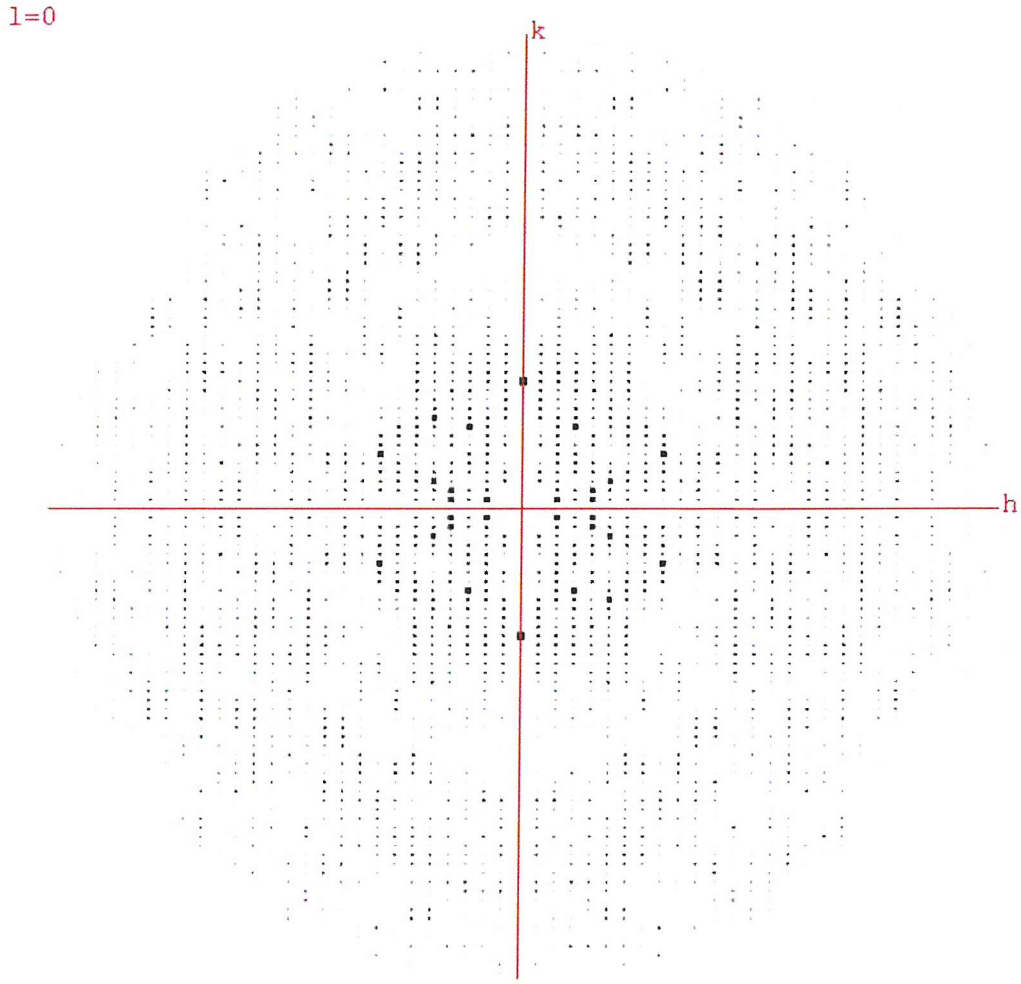


Figure 94: Pseudo-precession picture from HKLVIEW of the *hk0* layer showing two mirror planes (mm symmetry).

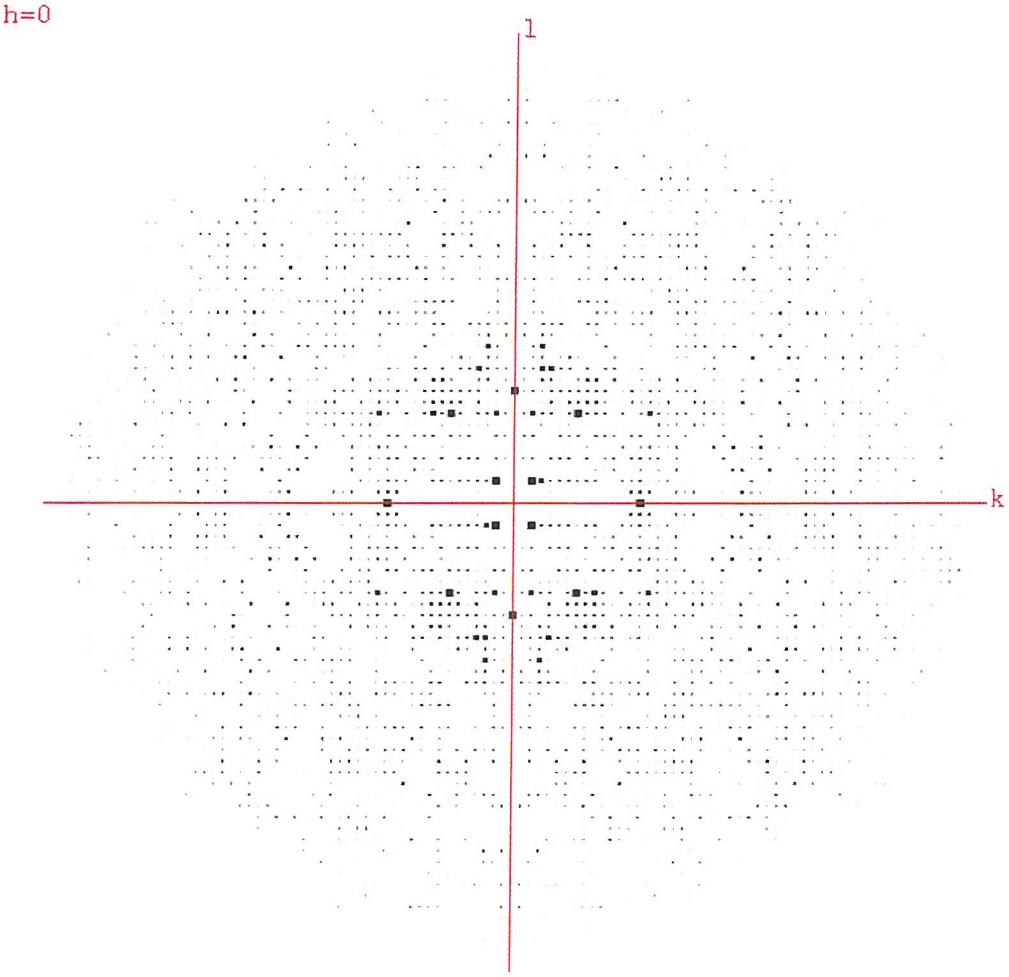


Figure 95: Pseudo-precession picture output from HKLVIEW of the *0kl* layer also showing mm symmetry.

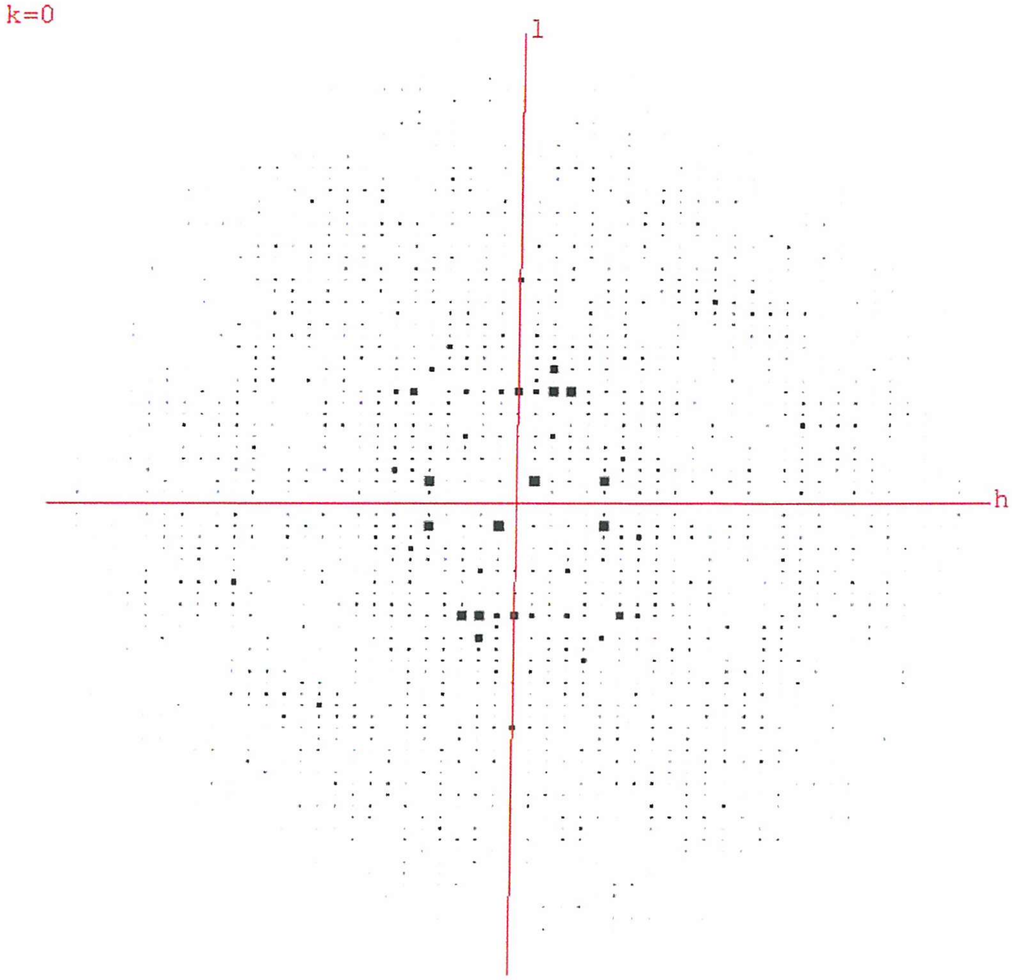


Figure 96: Pseudo-precession picture from HKLVIEW of the *h0l* layer showing no mirror planes and only inversion symmetry. This is indicative of point group 2 and not point group 222. An orthorhombic point group would have mm symmetry on all three zeroth layers.

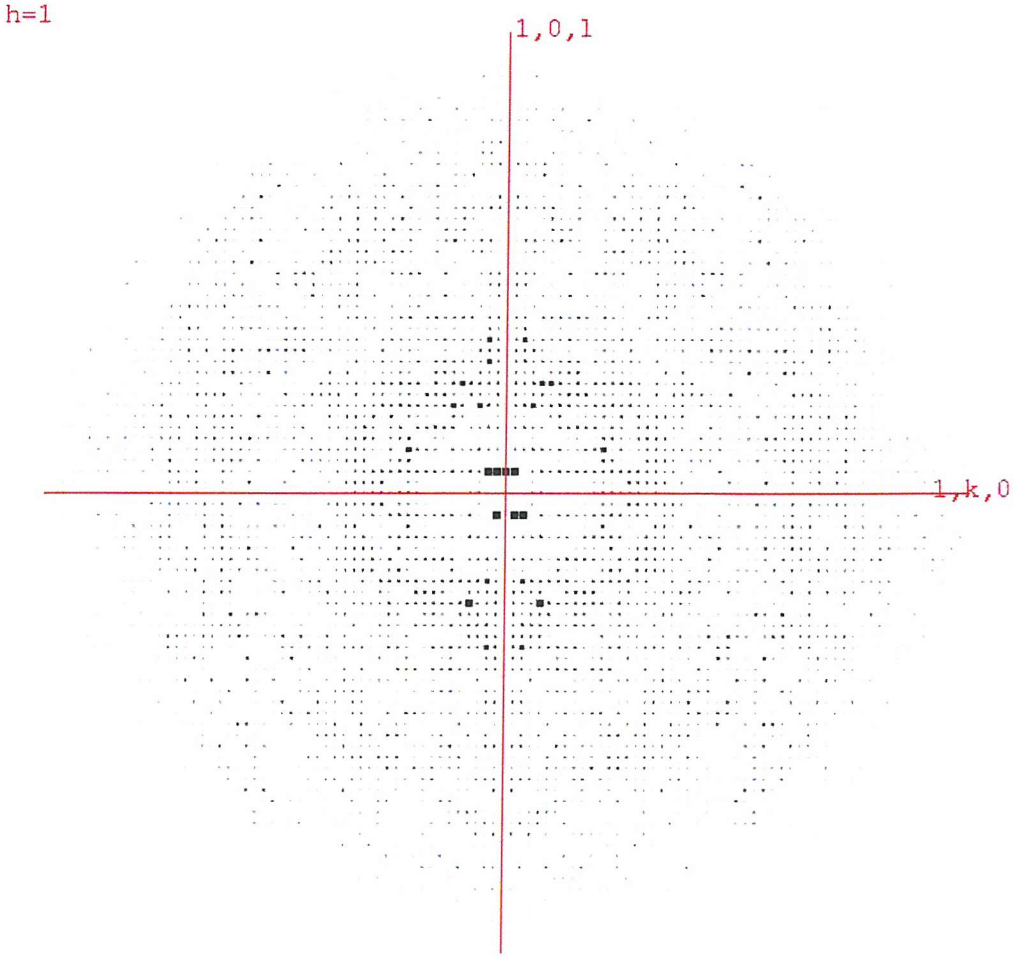


Figure 97: Pseudo-precession picture from HKLVIEW of the *1kl* layer showing only one mirror plane, perpendicular to the *k* axis. This is indicative of a monoclinic point group (ie. point group 2) and not an orthorhombic point group (ie. point group 222), as the latter would still have two mirror planes even at higher layers.

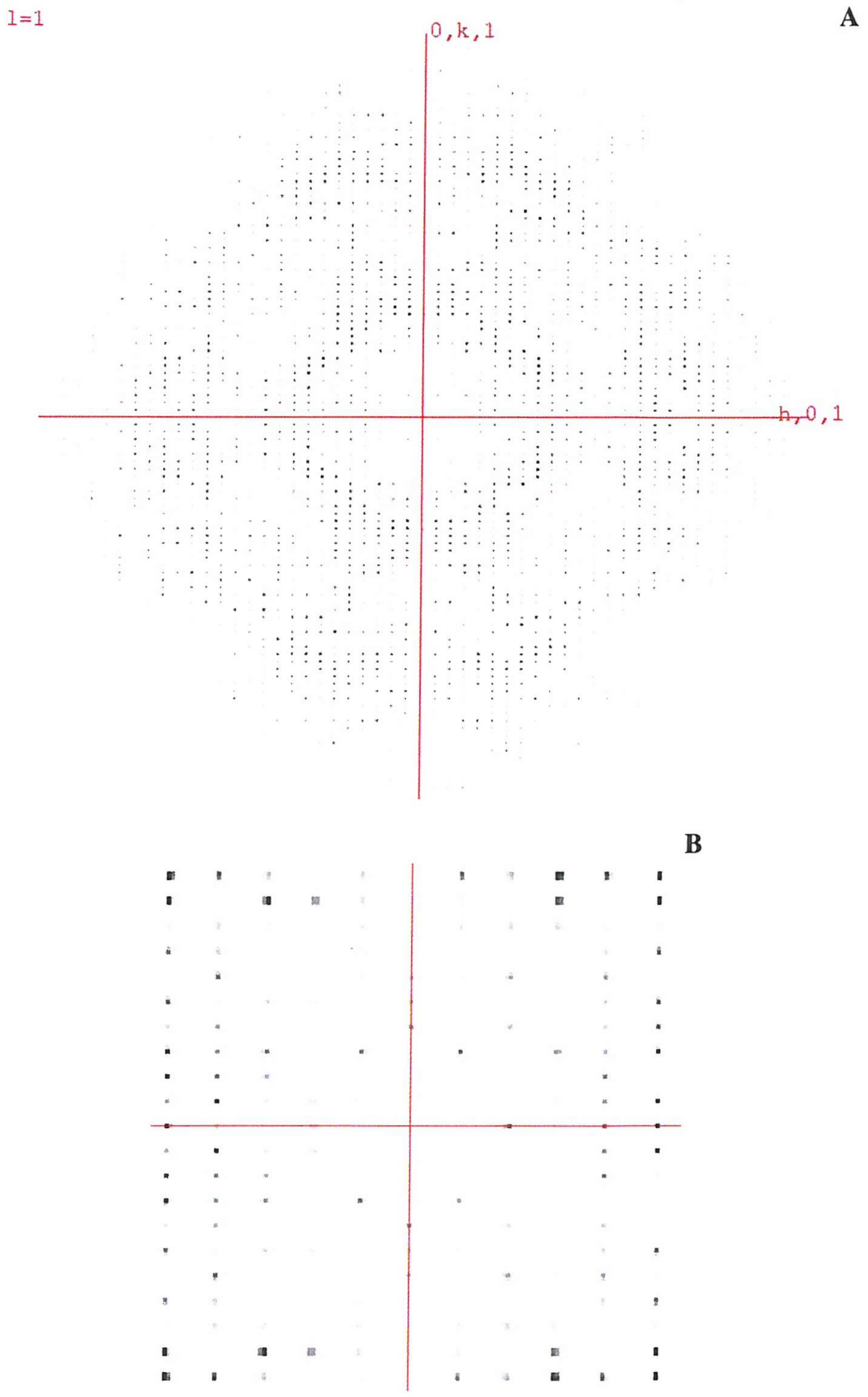


Figure 98: (**A**) Pseudo-precession picture from HKLVIEW of the *hk1* layer. It is difficult to discern the symmetry from this image, but from a blow-up of the central portion (**B**) it is possible to see m symmetry perpendicular to the *k* axis.

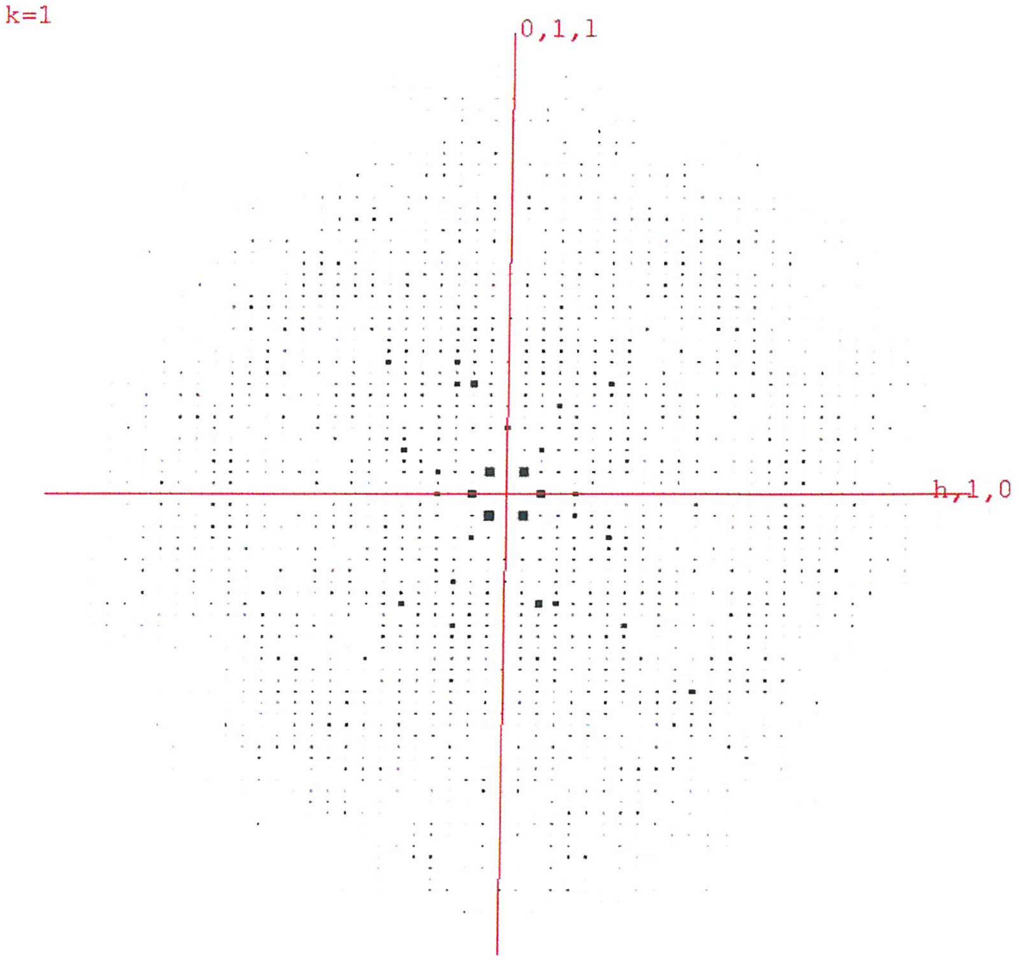


Figure 99: Pseudo-precession picture from HKLVIEW of the *h1l* layer showing no mm or m symmetry but only inversion symmetry. As before this indicates that the crystal is monoclinic in nature and not orthorhombic.