SUBSTRATE SPECIFICITY OF PENICILLIN G ACYLASE

BY

COLIN EDWARD McVEY

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

AT THE
UNIVERSITY OF YORK

DEPARTMENT OF CHEMISTRY

FEBRUARY 1999

Abstract

The periplasmic enzyme penicillin G acylase (PGA) from *E. coli* is a heterodimer which consists of an A chain of 209 amino acids and B chain of 557 amino acids. It is produced by proteolytic cleavage of a single precursor protein to remove an N-terminal signal peptide and a 54 amino acid spacer peptide linking the two chains. It catalyses the cleavage of penicillin G to produce phenylacetic acid and 6-aminopenicillanic acid (6-APA). The enzyme is of immense importance to the pharmaceutical industry as 6-APA is the starting point for the synthesis of many semi-synthetic antibiotics. The object of this thesis was to study the structural basis of PGA substrate specificity using both wild type and mutant proteins.

Seven crystal structures have been determined to probe substrate specificity. Mutations at Ser B:1 (alanine or cysteine), Asn B:241 to alanine and Arg B:263 to isoleucine gave rise to two phenotypes (i) 'inactive processed' and (ii) 'inactive precursor'. Structural studies of the two inactive processed mutants Ser B:1 Cys and Asn B:241 Ala revealed that Ser B:1 Cys had lost the ability to bind substrates perhaps through oxidation of the cysteine but the Asn B:241 Ala mutant could still bind substrate. The structures of Asn B:241 Ala complexed with the substrate penicillin G and the analogue penicillin G sulphoxide, together with a wild type PGSO complex provided a clearer rationale for substrate specificity. These showed that residues Met A:142-Phe A:146 adopt two alternate conformations, and Phe A:146 is directly involved occupying a vital site between subsites S1 and S1'.

Structure determination of a mutant with altered substrate specificity revealed a conformational change within the active site which changes the complementarity of the S1' substrate subsite thus providing a structural rationale for previous kinetic results.

The role of Phe A:146 was explored further using a novel *in vivo* amber suppressor mutagenesis strategy to probe the influence of amino acid substitutions at this position. Preliminary kinetic results add further support to the role of this residue in substrate recognition and catalysis.

Acknowledgements

Firstly, I would like to express my sincere thanks to my supervisor Keith Wilson for his support, encouragement and the many helpful discussions spent in the library!

There are a whole host of people I would especially like to thank for both their friendship and willingness to spend time to give priceless advice and guidance. Firstly, I would like to thank Jim for his friendship and supervision in the wet lab. I would also like to thank those who have nursed me through crystallography, Ashley, Gideon, Jean (for the brilliant symmetry course!) Johan, Marek, Garib, Koen, Rich T, Rick, Zbyszek and last but by no means least Eleanor who we all owe so much to.

I am particularly indebted to my fellow Irishman Martin Walsh for his patience and help during my frequent synchrotron visits to Hamburg and for the occasional trip to Finnegans bar. I am also grateful to Heidi and Ash for help with all the complicated computer programs I find so troublesome to use. I mustn't forget Muffy for the lucky whiskers donated, voluntarily I might add!

A special thanks must go to Tom for developing X-FIT and all the extra goodies!, and to Garib for REFMAC, who not only provided such a great program but took time to explain its use.

A big thanks to Keith, Guy, Rick, Lol, Ash and Jim for proof reading and helpful discussions. Finally, I would like to thank all of those at York, present and past and last but by no means least to Denise, just for everything.

List of Abbreviations

β-ME	β-mercaptoethanol
DTT	Dithiothreitol
MES	2-(N-morpholino)-ethanesulphonic acid
CAPS	3-(Cyclohexylamino)-1-propanesulphonic acid
MOPS	3-(N-morpholino)-propanesulphonic acid
NIPAB	6-nitro-3-phenylacetamidobenzoic acid
PGA	Penicillin G acylase
SDM	Site-directed mutagenesis
PCR	Polymerase chain reaction
Pen G	Penicillin G
PGSO	Penicillin G sulphoxide
TRIS	2-amino-2-(hydroxymethyl)-1,3-propanediol
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
pac	The penicillin G acylase gene
MR	Molecular replacement
rmsd	Root mean squared deviation
F_o	Observed structure factor amplitude
F_c	Calculated structure factor amplitude

To Mum & Dad

Contents

List of Figures

List of Tables

Chapter One Introduction

Alexander Fleming's historic observation of the antibacterial properties of a culture of *Penicillium notatum* in 1928 was the beginning of a new stage in man's endeavour to combat bacterial infection. It began with the isolation of the active antibacterial component by Ernst Chain working in Howard Florey's laboratory at Oxford, utilising the relatively new techniques of column chromatography and freeze drying to obtain what we now know as penicillin. The curative effect of penicillin against bacterial infection in mice was then demonstrated. Finally, with great difficulty, sufficient material was prepared to treat a few human patients (Nayler, 1991a).

Large scale production of penicillin was a serious problem, which was overcome by the selection of a strain of *Penicillium chrysogenum* with greatly increased titres when combined with the use of large fermentors. Large scale fermentation also revealed that a mixture of chemically related penicillins was produced and it was demonstrated that the relative amounts of these naturally occurring penicillins were dependent on the composition of the fermentation medium (Vandamme and Voets, 1974). Subsequently, the selective fermentation of penicillin G by adding phenylacetic acid as a precursor to the fermentation broth was achieved. By similar methods, the biosynthetic penicillin V was obtained which although not superior to penicillin G in antibacterial activity, was shown to be more suitable for administration by mouth on account of its greater acid stability.

The efforts of chemists and X-ray crystallographers, with a vital contribution from Dorothy Hodgkin, who determined the X-ray structure of penicillin G (Crowf*oot et* al., 1949), established the existence of five penicillins in which different acyl side chains (RCH_2CO) were attached to a common bicyclic (penam) nucleus (Figure 1.1).

The search for more potent penicillin derivatives culminated in the discovery of an intermediate of penicillin biosynthesis, 6-aminopenicillanic acid (Batche*lor et* al., 1959). This discovery began the subsequent exploitation of 6-APA and the development of semi-synthetic approaches to novel penicillins (Nayler, 1991a; Nayler, 1991b).

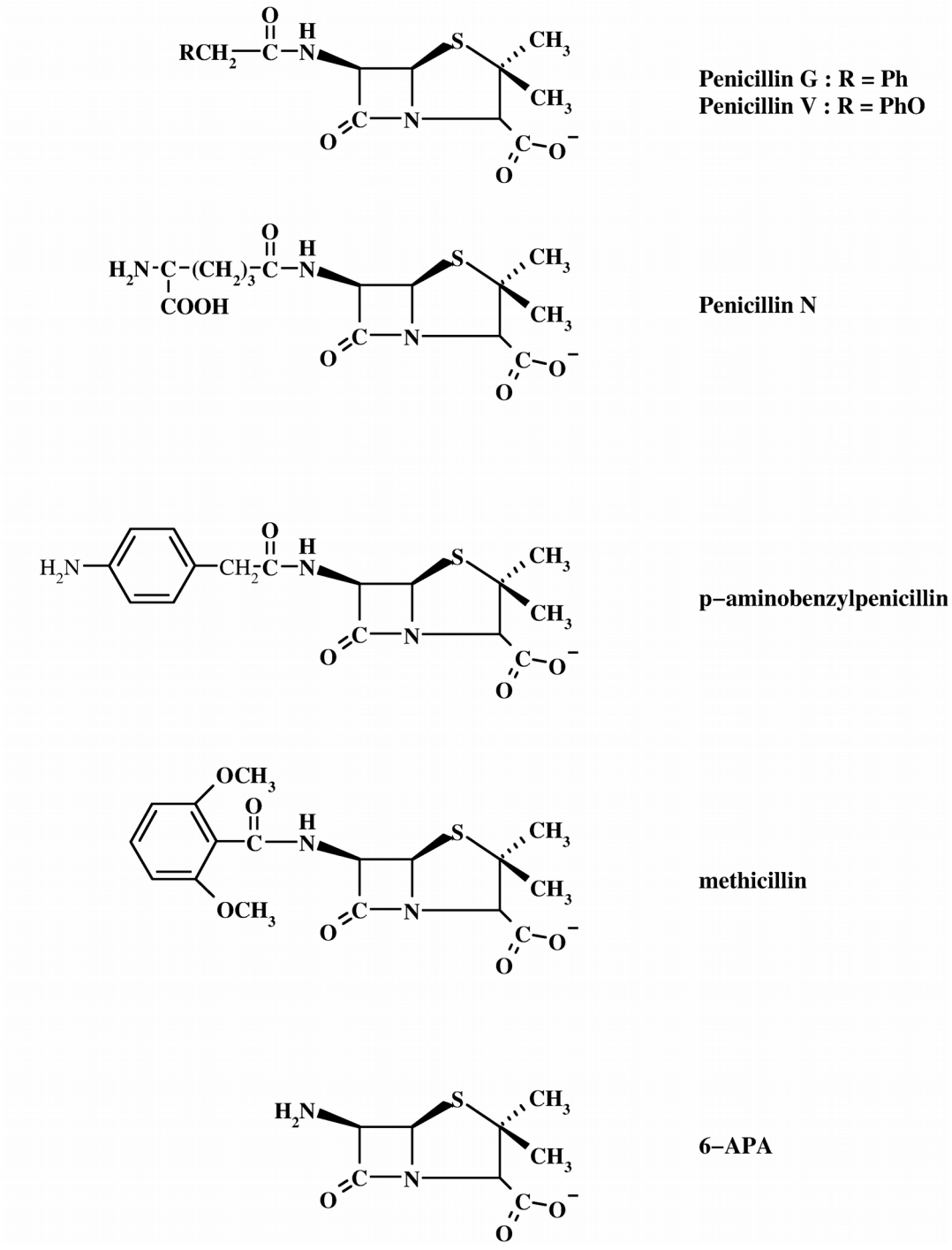


Figure 1.1 Structures of the early penicillins and 6-APA.

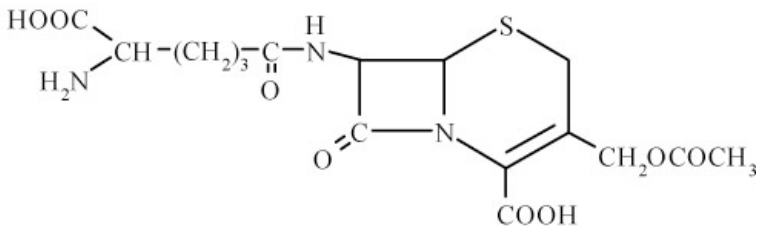
Meanwhile a sixth 'natural' penicillin was discovered by Newton and Abraham (Newton and Abraham, 1954) while investigating a trio of antibiotics from a strain of *Cephalosporium*. One component, cephalosporin C, became the progenitor of a second important series of β-lactam containing antibiotics, the cephalosporins (Figure 1.2a). Another component (originally called cephalosporin N) was found to be a penicillin and renamed penicillin N (Figure 1.1). It was less active than penicillin G against Gram-positive bacteria, but more active against certain Gram-negative bacteria such as *Salmonella*. This provided evidence that the antibacterial spectrum of penicillin was influenced significantly by the chemical structure of the side chain.

The penicillin molecule is susceptible to hydrolytic cleavage not only at the β-lactam bond, by the enzyme β-lactamase, but also at the amide bond (Figure 1.2b). The latter reaction can also be brought about enzymatically. The enzyme responsible has been given several different names, including penicillin amidase, penamidase, penicillin splitting and synthesising enzyme, acyltransferase and has the official nomenclature benzylpenicillin amidohydrolase (EC 3.5.1.11). In this thesis it will be referred to as penicillin G acylase (PGA).

1.1 Historical background to penicillin acylase

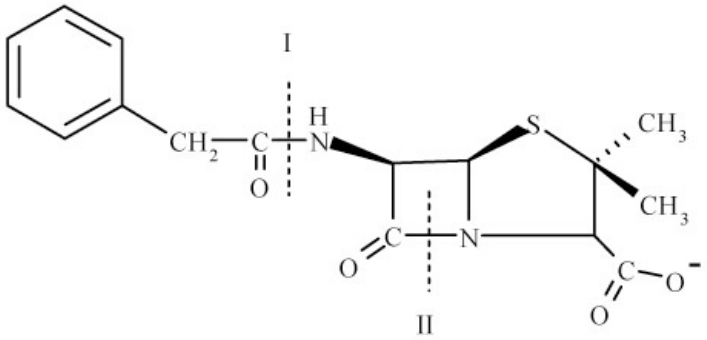
Acylase activity was first described in *Penicillium chrysogenum* Q 176 (Sakaguchi and Murao, 1950) showing that the mycelia of this strain were capable of converting penicillin G into phenylacetic acid and 6-aminopenicillanic acid. Penicillin acylases are widely distributed, among micro-organisms including both Gram-negative and -positive bacteria, yeast and fungi, but at the same time sparsely, e.g. only in W strain of *E. coli*. They are important industrial enzymes which have featured predominantly in the large scale synthesis of semi-synthetic penicillins, by producing 6-aminopenicillanic acid (6-APA), providing the penam nucleus from which semi-synthetic penicillins are derived.

(a)



Cephalosporin C

(b)



Penicillin G

Figure 1.2 (a) The chemical structure of cephalosporin C. (b) The penicillin G molecule showing the site of action of (I) penicillin acylase and (II) β-lactamase.

The initial classification was based on the source of the enzymes (Clari*dge et* al., 1960). Most acylases of bacterial origin were penicillin G specific, while those from fungi, yeast and *Streptomyces* were specific for penicillin V, F, or K (Clari*dge et* al., 1963). However it soon became clear that there were many exceptions to the rule. Today, the penicillin acylase group of enzymes is classified into three subtypes based on substrate preference: penicillin V acylases (PVA) which preferentially hydrolyse penicillin V (Type I), penicillin G acylases preferentially hydrolysing penicillin G (benzylpenicillin;

Type II) and ampicillin acylases (ApcA) which specifically hydrolyse ampicillin (Type III) (Hamilton-Miller, 1966; Va*lle et* al., 1991; Vandamme and Voets, 1974).

Rarely has a single enzyme been responsible for such a great revolution in therapeutic practice and had such an impact on human health. This was clearly illustrated in 1959 by the introduction of methicillin (Hamilton-Miller, 1966), the first commercially available semi-synthetic penicillin to be unaffected by staphylococcal β-lactamase, thereby changing the status of penicillin resistant hospital *Staphylococcus* literally overnight from 'public enemy number one' to merely a nuisance. However, semi-synthetic penicillins have become increasingly ineffective in combating bacteria which have acquired penicillin resistance. As a result, use of cephalosporin antibiotics are the preferred choice. Even third generation semi-synthetic cephalosporins are becoming inadequate against antibiotic resistant bacteria. An effective cephalosporin acylase would provide the means to make a wider range of semi-synthetic cephalosporins. Despite extensive screening, no natural enzyme does this on cephalosporin C, the most available starting material from fermentation.

1.2 Acylases

In view of PGA's ability to hydrolyse penicillin and its application in the production of novel antibiotics there has been extensive research into its production and properties, its isolation and cloning from various micro-organisms and its substrate specificity.

The diverse group of penicillin acylases are classified according to the substrates they prefer with Gram-negative such as *E. coli* PGA (Schumac*her et* al., 1986) a type II acylase of which a number have been sequenced and characterised (Table 1.1), including those from *Kluyvera citrophila* (Barb*ero et* al., 1986), *Providencia rettgeri* (Ljubijan*kic et* al., 1992), *Alcaligenes faecalis* (Verha*ert et* al., 1997), and from Gram-positive bacteria including *Bacillus megaterium* [Martin, 1995 #672] and *Arthrobacter viscosus* (Oha*shi et* al., 1988). In addition to having similar substrate preferences, these enzymes are evolutionarily related (Barb*ero et* al., 1986; Oha*shi et* al., 1988), and their peptide sequences (Figure 1.4) suggest that all type II PGA genes encode a polypeptide that is subject to processing.

Parameters	*E. coli*	*A. faecalis*	*K. citrophila*	*P. rettgeri*	*A. viscosus*	*B. megaterium*
% DNA identity of coding sequence	100	50	77	61.5	46	46
Molecular mass (kDa) mature protein	86.2	85.7	85.2	86.1	85.7	85.6
No. of signal sequence aa	26	26	26	23	26	26
α subunit						
% DNA identity	100	53	78	63	47	47
Molecular mass (kDa)	23.8	23.0	23.6	23.7	24.3	24.2
No. of aa	209	202	209	205	208	208
% aa identity	100	47	83	65	33	33
% aa similarity	100	67	85	71	46	47
No. of spacer amino acids	54	37	54	56	31	31
β subunit						
% DNA identity	100	49	77	62	62	46
Molecular mass (kDa)	62.4	62.7	61.7	62.2	61.4	61.4
No. of aa	557	551	555	553	537	537
% aa identity	100	41	87	64	31	32
% aa similarity	100	59	90	72	42	42

Table 1.1 Properties and sequence similarity of type II Penicillin G acylases.

Verhaert and co-workers (1997) have demonstrated the greater stability of the *A. faecalis* enzyme and showed that activity was not affected by incubation at 50 °C for 20 min, whereas more than 50% of the *E. coli* enzyme was irreversibly inactivated by the same treatment. A unique feature of the *A. faecalis* enzyme is the presence of two cysteines that form a disulphide bridge and improved thermostability is attributed to the presence of the disulphide bridge since stability of the *A. faecalis* enzyme was decreased by addition of a reductant (10 mM DTT).

1.2.1 Gene structure and organisation

E. coli pac is the best characterised gene of the PGA family and a complex regulatory network has been proposed (Roa and Garcia, 1997; Va*lle et* al., 1991). The expression of *pac* is subject to several regulatory controls, e.g. temperature, catabolite repression and induction by phenylacetic acid. The 5' flanking region of the *pac* gene contains several sequences that may be involved in the regulation of gene expression (Figure 1.3). Analysis of this region has identified two putative CRP-cAMP binding sites, three palindromic sequences proposed as recognition sites for regulatory proteins and a promoter sequence *(Oh et* al., 1987; Va*lle et* al., 1986).

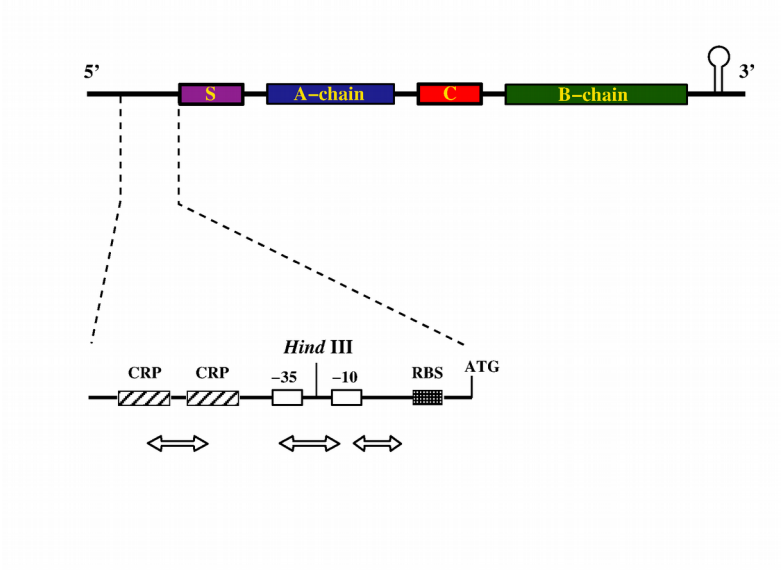
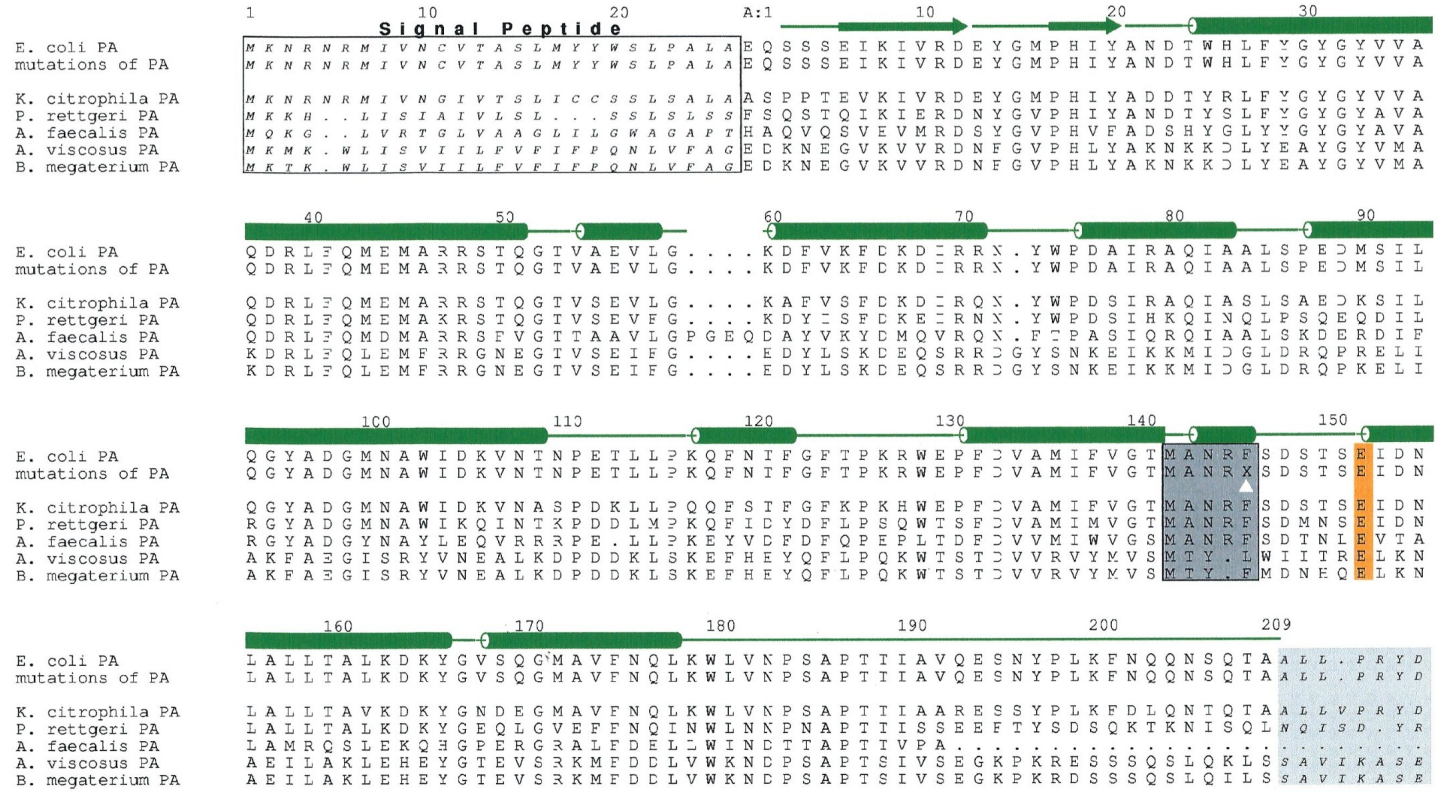


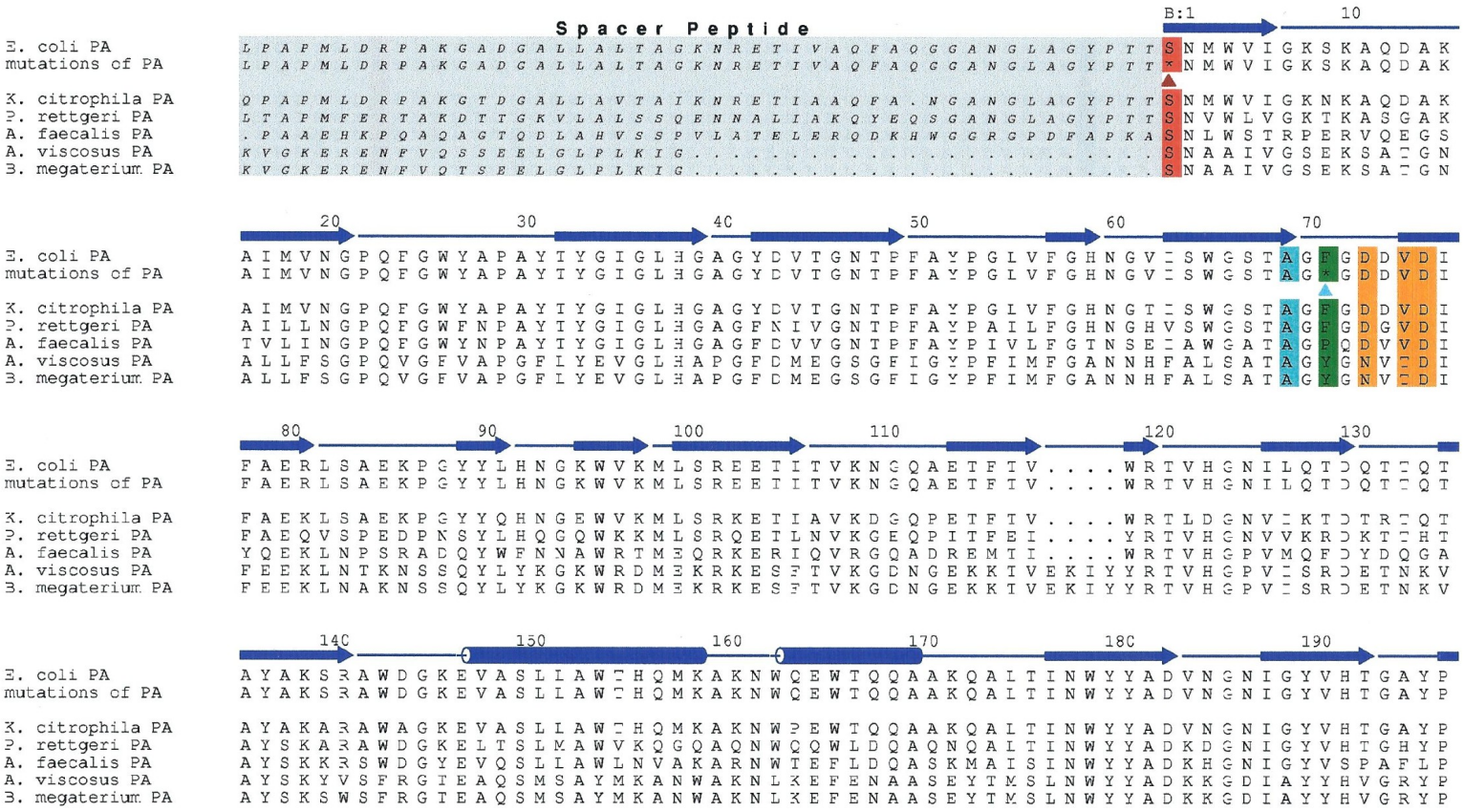
Figure 1.3 Organisation of the E. coli PGA gene and its 5' and 3' flanking regulatory regions. *Open arrows indicate imperfect palindromic sequences. CRP- cAMP receptor protein, RBS- ribosome binding site.*

1.3 Enzyme maturation

Mature *E. coli* PGA is a periplasmic 86 kDa heterodimer of two chains labelled A (23.8 kDa) and B (62.24 kDa) consisting of 209 and 557 amino acids respectively, synthesised as a single cytoplasmic precursor from which the active enzyme is generated by a sequential proteolytic processing pathway (Barb*ero et* al., 1986; B*ock et* al., 1983b; Schumac*her et* al., 1986). The PGA precursor polypeptide comprises four distinct regions, namely signal sequence, A subunit (A:1 to A:209), spacer peptide and B-subunit (B:1 to B:557). Each region plays a crucial role in the transport and processing mechanism distinct from the role of the A and B subunits in catalysis (Sizm*ann et* al., 1990). The signal sequence directs transport of the PGA precursor into the periplasmic space with the spacer peptide playing a putative role in the correct folding of the subunits before it is cleaved to release the mature enzyme (Da*umy et* al., 1985b).

Schumacher and co-workers (1986) showed that maturation of PGA in *E. coli* occurs in three steps: (i) removal of the signal sequence during transport to the periplasm, (ii) cleavage between the N-terminal serine of the B subunit and the spacer peptide, and (iii) removal of the spacer by cleavage at the C-terminal side of the A subunit. It has been proposed that this processing is autocatalytic (V. Kasche, submitted). Choi *et al*. proposed that the N-terminal serine is critical for processing as well as enzyme activity and also that there is a difference in structural requirements for the processing pathway and for enzymatic activity (C*hoi et* al., 1992). The autoproteolysis might help to ensure that the enzyme is correctly folded or active only at its correct cellular location in the periplasm.


Signal Peptide
E. coli PA
mutations of PA
K. citrophila PA
P. rettgeri PA
A. faecalis PA
A. viscosus PA
B. megaterium PA


Spacer Peptide
B:1
10
E. coli PA
mutations of PA
K. citrophila PA
P. rettgeri PA
A. faecalis PA
A. viscosus PA
B. megaterium PA
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190

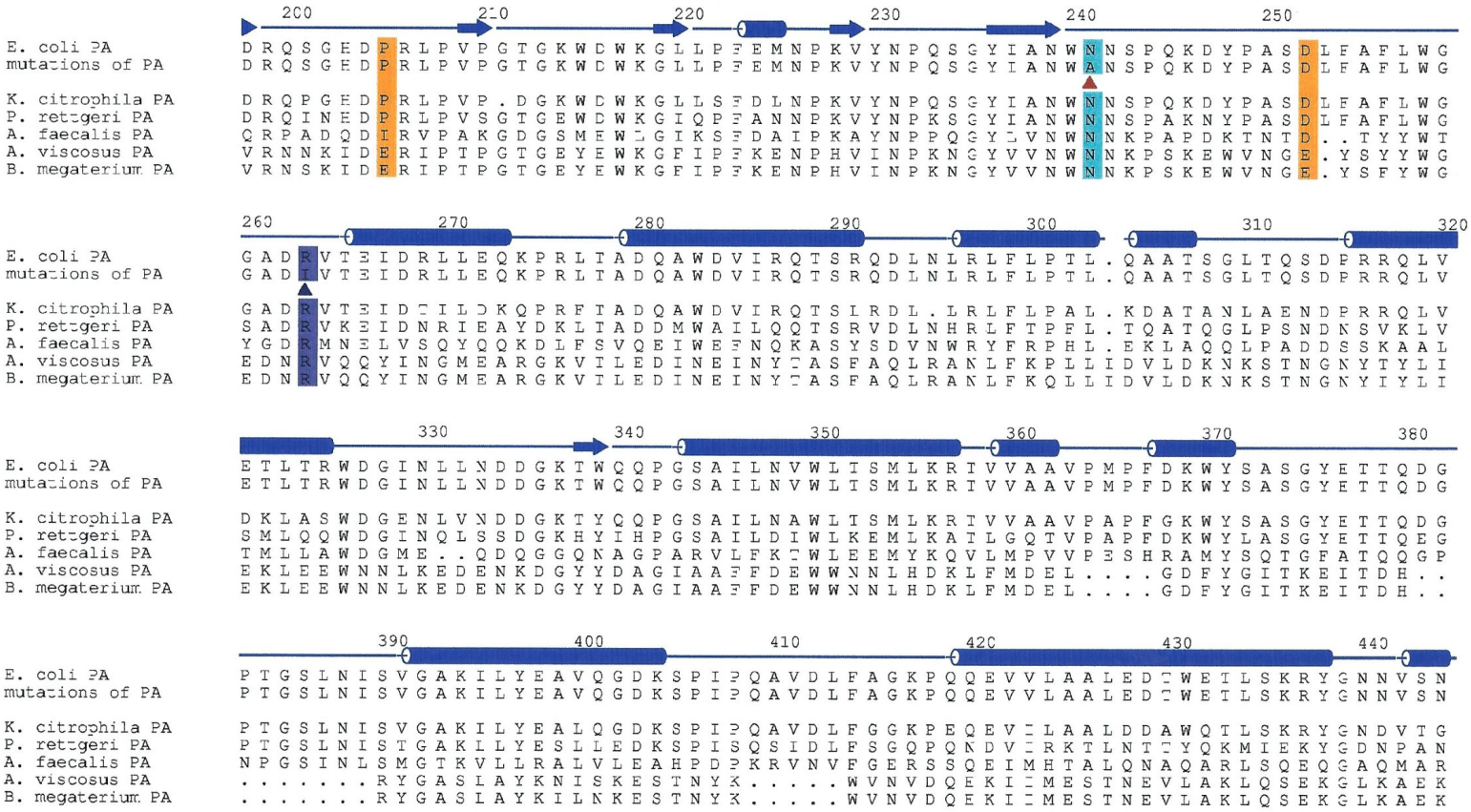

200
210
220
230
240
250
E. coli PA
mutations of PA
K. citrophila PA
P. rettgeri PA
A. faecalis PA
A. viscosus PA
B. megaterium PA
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440

Figure 1.4 Overlap of amino acid sequence and secondary structure.

1.4 The in vivo *role of PGA*

Studies on PGA's physiological role have been limited, and its *in vivo* function remains unclear. Expression of the enzyme is regulated by three major mechanisms: cAMP receptor protein (CRP), phenylacetic acid and thermoregulation (Va*lle et* al., 1991). Expression of PGA is induced by phenylacetic acid as the sole carbon source, giving rise to the hypothesis that the enzyme could be involved in the assimilation of aromatic compounds as carbon sources in the organism's free-living mode (Merino *et al*., 1992). Quantitative analysis of specific *pac* mRNA with a *lac*Z fusion to the 5'-terminal region of the *pac* gene demonstrated that both phenylacetic acid induction and catabolite repression by glucose regulate at the transcriptional level.

In *E. coli* the optimal temperature for PGA activity ranges from 25 to 30 °C and no activity is detectable above 37 °C (Bhattacha*rya et* al., 1993). The loss of enzyme activity is not due to heat inactivation of the enzyme, since the optimal temperature for activity is 42 °C (Szentirmai, 1965). The biochemical basis of this strict temperature-dependent formation has been investigated (Keilm*ann et* al., 1993), who showed that *pac* under the control of the *lacUV5* promoter exhibited the same temperature dependent expression as the chromosomally encoded gene transcribed from its own promoter. This indicated that a post-transcriptional level of *pac* mRNA, rather than the transcription of the gene, is temperature sensitive. It has been proposed that the effect of raised temperature is to interfere with the folding of the A subunit of PGA, thus preventing proper folding of the precursor (Lindsay and Pain, 1991). This may account for the basis of the temperature regulation of PGA production *in vivo*.

Regulation of *pac* from *Providencia rettgeri* (formerly *Proteus rettgeri*) is different from that of *E. coli*. Production of PGA in *P. rettgeri* is not subject to temperature regulation but is expressed constitutively, and is repressed by the C_4 dicarboxylic acid intermediates of the tricarboxylic acid cycle, such as succinate, fumarate, and malate (Da*umy et* al., 1982). Recently, it was noted that *pac* is located downstream of a gene cluster encoding the 4-hydroxyphenylacetic acid degradative pathway of *E. coli* W (Pri*eto et* al., 1996). The authors proposed that the presence of *pac* in the proximity of the 4-HPA cluster could mean that PGA was a recent acquisition to improve the ability of *E. coli* W to metabolise a wider range of substrates, enhancing its catabolic versatility

and further supports the role of PGA as a scavenger enzyme for phenylacetylated compounds (Mer*ino et* al., 1992).

1.5 Substrate specificity of penicillin acylase

PGA from *Escherichia coli W* ATCC 11105, is the most widely used and important of the type II acylases. The substrate specificity of this enzyme has been studied extensively (Clari*dge et* al., 1963; Cole, 1969a; Cole, 1969b; Hu*ang et* al., 1963; Kaufmann and Bauer, 1964; Margo*lin et* al., 1980; Plas*kie et* al., 1978; Šve*das et* al., 1996), and the main conclusions are that PGA is primarily active on penicillin G, but shows broader substrate specificity, capable of hydrolysing a diverse range of amides. PGA is highly specific to phenylacetyl compounds (Cole, 1969b) with the structure of the leaving group hardly affecting the hydrolysis rate (Cole, 1964) but with substitution of the α-carbon with an amino or hydroxyl group resulting in a fall in hydrolysis rate (Cole, 1969a).

The substrate specificity can be generalised as follows: for a substrate of the general formula R-CO-NH-R', the specificity is determined mainly by the acyl moiety, R, with particular preference for hydrophobic acyl moieties. In contrast, R' can be varied substantially, e.g. penam or cephem nucleus or a variety of amino acids (Margo*lin et* al., 1980; Plas*kie et* al., 1978).

Penicillin G is often claimed as the preferred substrate of PGA from *E. coli* (Hu*ang et* al., 1963; Shewale and Sivaraman, 1989), but according to Švedas and co-workers (Šve*das et* al., 1996) the most reactive substrates are all phenylacetylated compounds including penicillin G, but it is not the most reactive.

A detailed study has been carried out on ligand complexes of PGA with phenylacetic acid derivatives to investigate the specificity of PGA towards the phenylacetyl side chain, using X-ray structures and kinetic data (Done *et al*., 1998).

1.5.1 Stereospecificity

The *E. coli* enzyme is primarily specific for the acyl structure but the nature of the amino group that is acylated is also of importance, particularly its stereochemical configuration. Stereospecificity appears to be virtually absolute for deacylation of α amino acids, the amino acid side chain must be in the L-configuration for hydrolysis to

occur. Asymmetry in the acyl group has only a slight effect, compounds with the α-carbon atom in the D-configuration being slightly more readily deacylated (Cole, 1969a; Cole, 1969b).

1.6 Structure of penicillin acylase

The crystal structure of PGA from *E. coli* was previously determined to 1.9 Å resolution (Duggl*eby et* al., 1995). The structure is pyramidal in cross section with a deep cone-shaped depression harbouring the active site. The two chains of the heterodimer are closely intertwined, with no obvious discrete domains (Figure 1.5), the A chain is predominantly α-helical with the B chain organised into three regions of which the largest is a β-sandwich motif flanked by α-helices. One side of the sandwich is essentially a flat sheet formed by six anti-parallel strands, with two α-helices packed against it; the other side is a twisted sheet of seven anti-parallel β-strands with a second pair of flanking α-helices. The second domain (B:294-B:439) in the B-chain is α-helical where six helices are arranged in two layers, three in the upper and three in the lower, whilst the third domain (B:74-B:142) is a seven stranded anti-parallel β-barrel.

1.6.1 Chemical modification and related functional studies

Much effort was devoted to identifying residues having a role in the active centre of PGA before determination of its crystal structure. These studies focused on PGA from *E. coli* and *K. citrophila* which share high homology (C*hoi et* al., 1992; Mahajan and Borkar, 1983; Márq*uez et* al., 1988; Prabhune and Sivaraman, 1990; Sl*ade et* al., 1991). The results implicated a number of residues and residue types important for enzyme function, the identification ranging from wild speculation based on sequence information and random chemical modification to more precise and intuitive mutagenesis experiments.

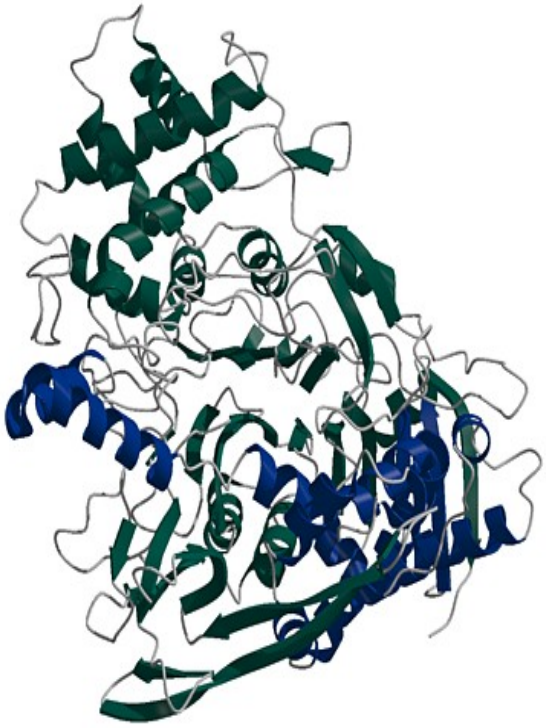


Figure 1.5 A protein cartoon of the secondary structure of PGA. *The A chain is coloured blue and the B chain coloured green.*

Majahan and Borkar carried out chemical modifications of the *E. coli* PGA using 2-nitrophenyl-sulphenyl chloride as well as 2-hydroxy-5-nitrobenzyl bromide which are specific for tryptophan, postulating its involvement at the catalytic site (Mahajan and Borkar, 1983). They showed that specific tryptophan modification inactivated the enzyme but the competitive inhibitor phenylacetic acid (PAA) protected the enzyme from inactivation but not modification, although this was to a lesser extent. The crystal structure reveals the presence of two tryptophan residues (Trp B:154 and Trp B:179) at the back of the pocket where PAA binds. Similarly, chemical modification provided evidence for involvement of an arginine residue at the catalytic site of *E. coli* PGA (Prabhune and Sivaraman, 1990). Incubation with phenylglyoxal or 2,3-butanedione resulted in enzyme inactivation, while both penicillin G and phenylacetate protect the enzyme against inactivation. The authors proposed the presence of arginine at or near the catalytic site and inactivation kinetics indicate the presence of a single essential arginine moiety. The crystal structure suggests that the likely arginine residue is Arg

B:263. Several studies suggested that the N-terminal serine is responsible for catalysis [Slade, 1991 #637; Choi, 1992 #675; Martin, 1991 #642] and the implications of their findings are discussed later in Chapter 4.

1.7 Ligand complexes of PGA

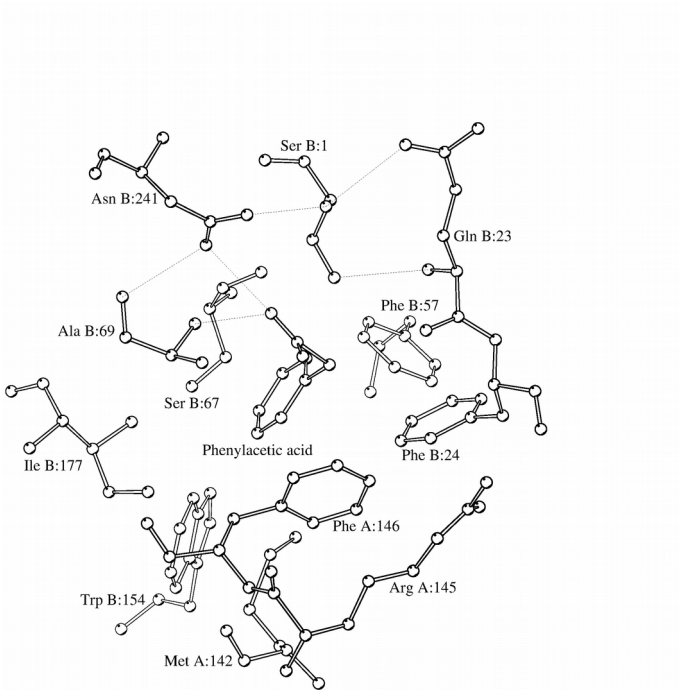
The structures of the enzyme complexed with phenylacetic acid and phenylmethylsulphonyl fluoride (PMSF) (Figure 1.6) locate the binding site for the side chain of the substrate and the residues that form the oxyanion hole (Dugg*leby et* al., 1995).

The structure of the complex with the competitive inhibitor phenylacetic acid revealed a pocket containing the phenyl ring of the inhibitor which presents the carboxylate group of the inhibitor to the active site serine B:1. The pocket is lined with aromatic and hydrophobic side chains and the complementary fit explains the specificity of the enzyme towards the phenyl moiety of a broad range of substrates (Cole, 1969b).

Done investigated the specificity of PGA towards the phenylacetyl moiety using a series of substituted phenylacetic acids, this work provided the molecular basis for binding of phenylacetic acid using analysis of kinetic data complemented by the corresponding crystal structures (Figure 1.7) and explains why deviation from this structure decrease the rate of hydrolysis (D*one et* al., 1998). The resulting complexes fall into two subsets which arise from two distinct energetically favourable conformations depending on the class of ligand bound. Weaker binding ligands, e.g. *m*-nitrophenylacetic acid, result in a major conformational change of Arg A:145 and Phe A:146 and unwinding of the active site helix (Figure 1.8), which will be discussed in a different context, in Chapter 5.

Ligands that cause a conformational change are shifted out of the active site pocket and make fewer hydrogen bonds to the protein. Done demonstrates that those ligands that do not cause a conformational change, e.g. *p*-hydroxyphenylacetic acid, have a lower K_i. Done also confirmed the residues involved in stabilisation of the transition state using benzylboronic acid, a transition state mimic, which is the most potent PGA inhibitor known to date (K_i = 1.6 nM) (Done, 1996).

(a)



(b)

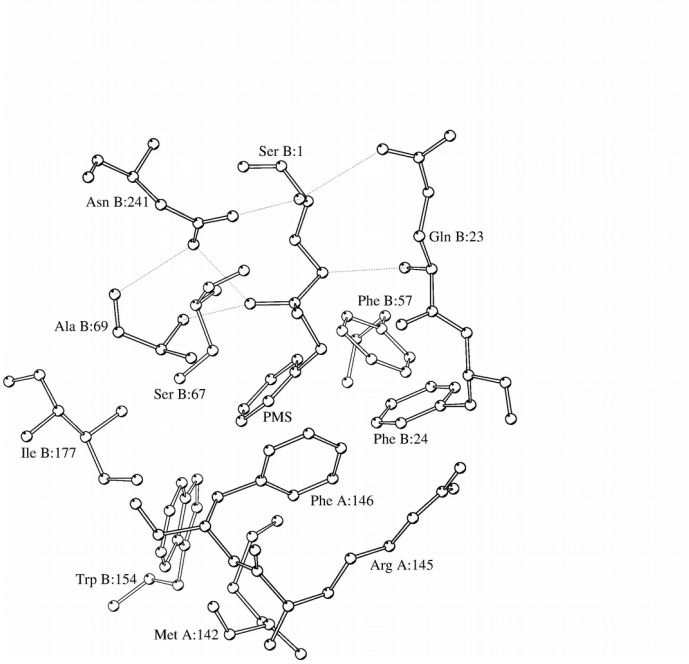


Figure 1.6 The structure of (a) PAA bound in the active site S1 specificity pocket and (b) the serine protease inhibitor PMSF covalently bound in the active site of PGA, labelled PMS.

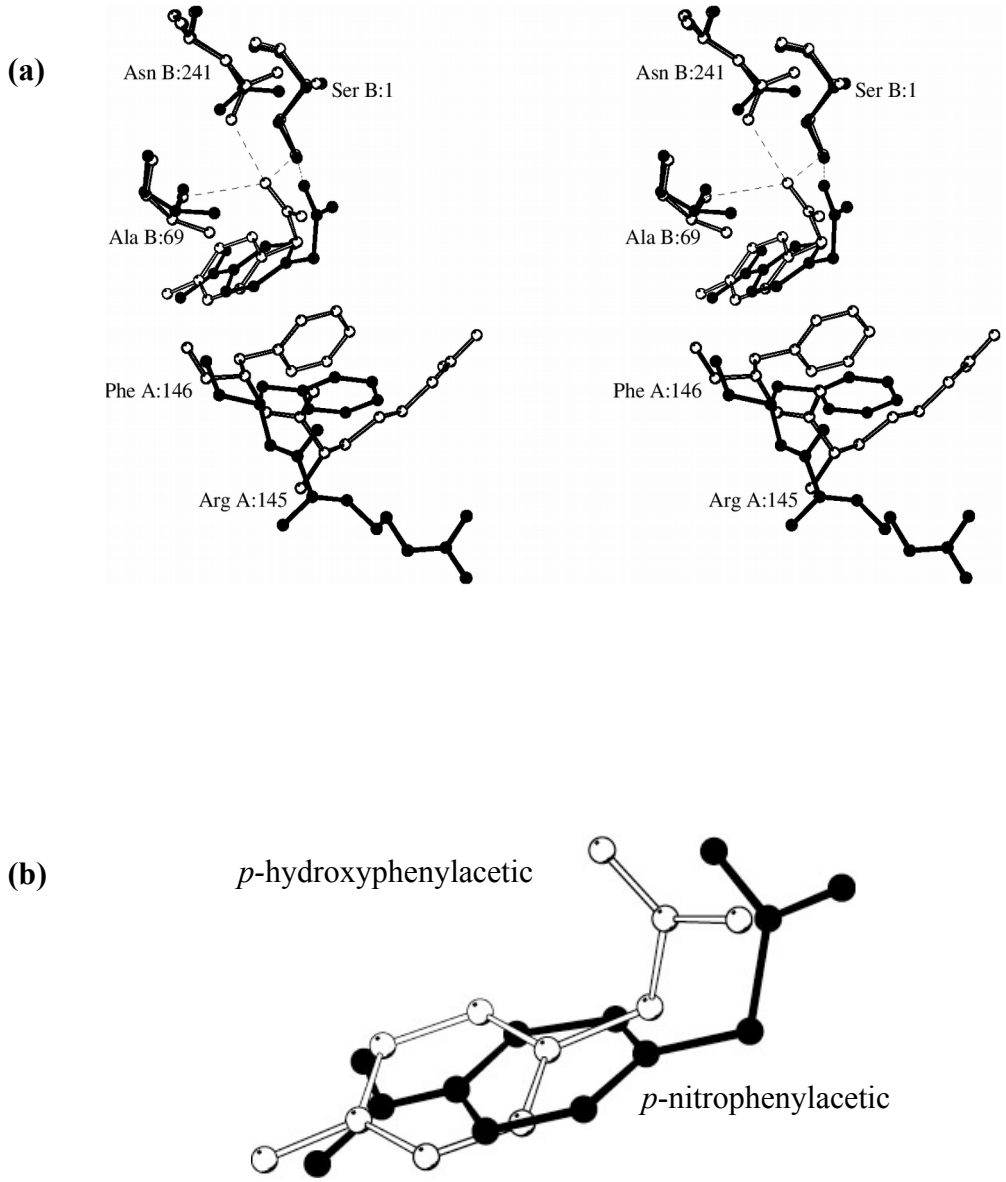


Figure 1.7 The two observed conformational states of the active site of PGA. *(a) stereo representation of p-hydroxyphenylacetic from subset 1 (white) and p-nitrophenylacetic from subset 2 (black) structures and the interactions made with each ligand. (b) close up of the overlaid ligand positions (Done et al., 1998).*

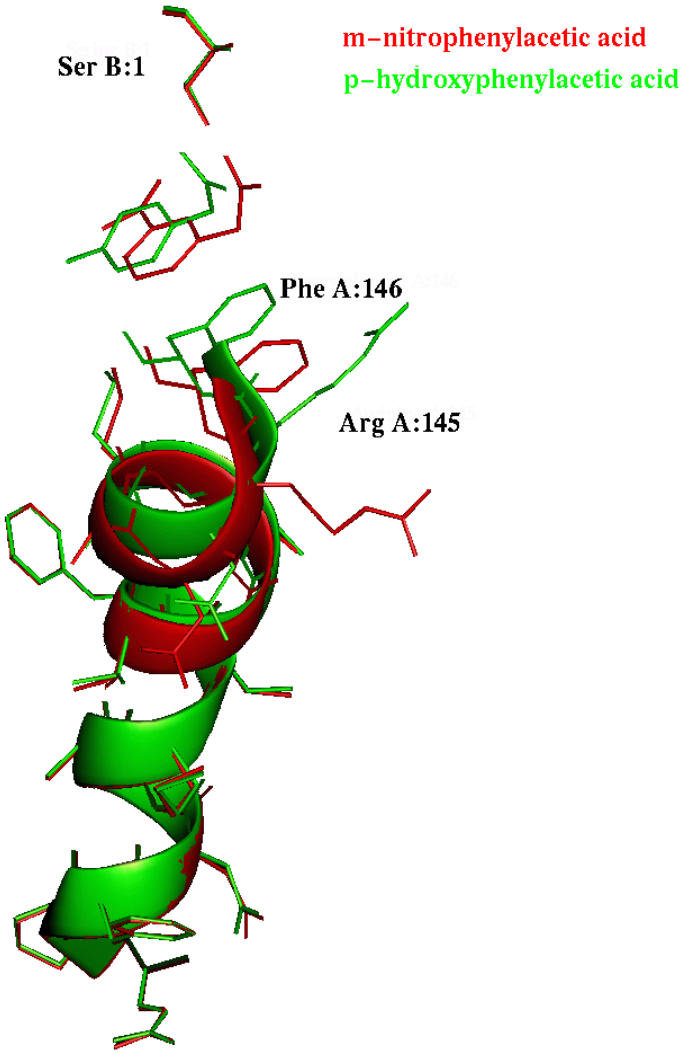


Figure 1.8 Helix unwinding as a result of ligand displacement from the binding pocket.

1.7.1 Calcium binding in PGA

There had been no report of a metal ion requirement for PGA, yet the X-ray structure determination of PGA from *E. coli* revealed the presence of a calcium ion. Examination of the published structure (1PNK.pdb) shows the metal binding site is typical of Ca^{2+} coordination with an octahedral arrangement (Duggl*eby et* al., 1995), and is completely buried within the interface between the A and B chains of the protein. The calcium ion

is therefore not easily exchangeable and may explain the absence of any reported calcium involvement in PGA function.

The calcium ion appears to perform a structural role, holding the two chains of the enzyme together. There is an interesting parallel in that calcium has also been found to have a structural role in some serine proteinases amongst others (McPha*len et* al., 1991). The metal binding within PGA has implications for folding of the protein. Proteins that result from proteolytic processing are characteristically difficult, if not impossible, to refold. Lindsay and Pain (1990) have investigated the refolding of PGA and the different factors that influence the correct folding, using unfolded PGA as well as unfolded A and B subunits under the influence of pH, ionic strength, and temperature. The levels of recovery were estimated by enzyme activity with the chromogenic substrate NIPAB (Kutzbach and Rauenbusch, 1974). Yields of 60% refolded protein were obtained based on recovery of enzyme activity. Since ligands for the calcium ion come from both chains, the presence of calcium ions is a likely condition for successful reassociation to yield active enzyme. The partial recovery of enzyme activity may be representative of trace amounts of calcium, but this cannot be confirmed as no details are given of the precise contents of the refolding buffers and the importance of metal ion binding was not studied. A more detailed description of the calcium binding site is given in Chapter Three.

1.8 Enzyme mechanism

Linus Pauling postulated that enzymatic catalysis is due to preferential binding of substrates to the enzyme in a configuration that resembles the transition state complex (Pauling, 1946). By doing so, the enzyme lowers the activation energy barrier of a reaction, resulting in an increase in reaction rate. This simple but elegant idea continues to serve as the single explanation for enzyme catalysis and has received experimental and theoretical support especially from the investigations into the serine proteinases.

It was proposed that the mechanism of PGA catalysis is similar to that of the serine proteinases and involved nucleophilic attack by the hydroxyl of the B-chain N-terminal Serine (Ser B:1) (Sl*ade et* al., 1991). The serine reagent, phenylmethylsulphonyl fluoride (PMSF), structurally resembles the side chain of penicillin G, inactivating PGA both from *E. coli* (Kutzbach and Rauenbusch, 1974) and from *P. rettgeri* (Da*umy et* al.,

1985a) highlighting the important role of serine B:1 in catalysis. The N-terminal serine has been changed to a cysteine both by chemical means (Sl*ade et* al., 1991) and by site-directed mutagenesis (C*hoi et* al., 1992), and this completely abolished enzyme activity. Chemical modification of the N-terminal serine from *Kluyvera citrophila* PGA led to the proposal that the residue was a likely candidate for the nucleophile [Martin, 1991 #642], and that the catalytic mechanism progresses through an acyl-enzyme intermediate (Kone*cny et* al., 1983).

Interestingly, the crystal structure reveals no adjacent base (histidine or lysine) nor an accompanying carboxylate nearby to enhance the nucleophilic capacity of the serine, raising the question of why there is no catalytic triad equivalent to that of the serine proteinases. The nucleophilic character of the active serine must be enhanced by other means. Probably a base adjacent to the serine is necessary to enhance the nucleophilic character of the hydroxyl function and the X-ray structure suggests that the likely candidate is the N-terminal α-amino group of Serine B:1 mediated by a bridging water molecule (Dugg*leby et* al., 1995). Following nucleophilic attack the resulting oxyanion tetrahedral intermediate is thought to be stabilised by interactions with the main chain amides of B:23, B:69 and with the Nδ of Asn B:241, creating an oxyanion hole equivalent to that in the serine proteinases. The oxyanion hole can be envisaged as a positively polarised binding pocket that accommodates the carbonyl oxygen of a scissile peptide bond (Henderson, 1970). It acts to orient and polarise the carbonyl bond leaving the carbonyl carbon atom with partial positive charge and thus more susceptible to nucleophilic attack, and to preferentially bind the transition state, putting detail to Pauling's hypothesis of 1946. An outline of the proposed catalytic mechanism of PGA is shown in Figure 1.9.

The formation of the covalent acyl-enzyme intermediate is the rate limiting step in the hydrolysis of the chromogenic substrate NIPAB (6-nitro-3-phenylacetamido-benzoic acid). However, D_2O isotope effect studies showed that it is the activation of the catalytic serine residue via a water molecule that is the actual rate-limiting step in the formation of the covalent intermediate (Jans*sen et* al., 1998).

Recent developments have revealed that serine proteinases can function without the landmark Ser/His/Asp catalytic triad and novel serine proteinases have been discovered that use hydroxyl/ε-amine or hydroxyl/α-amine 'catalytic dyads' as their reactive centres (Paetzel and Dalbey, 1997).

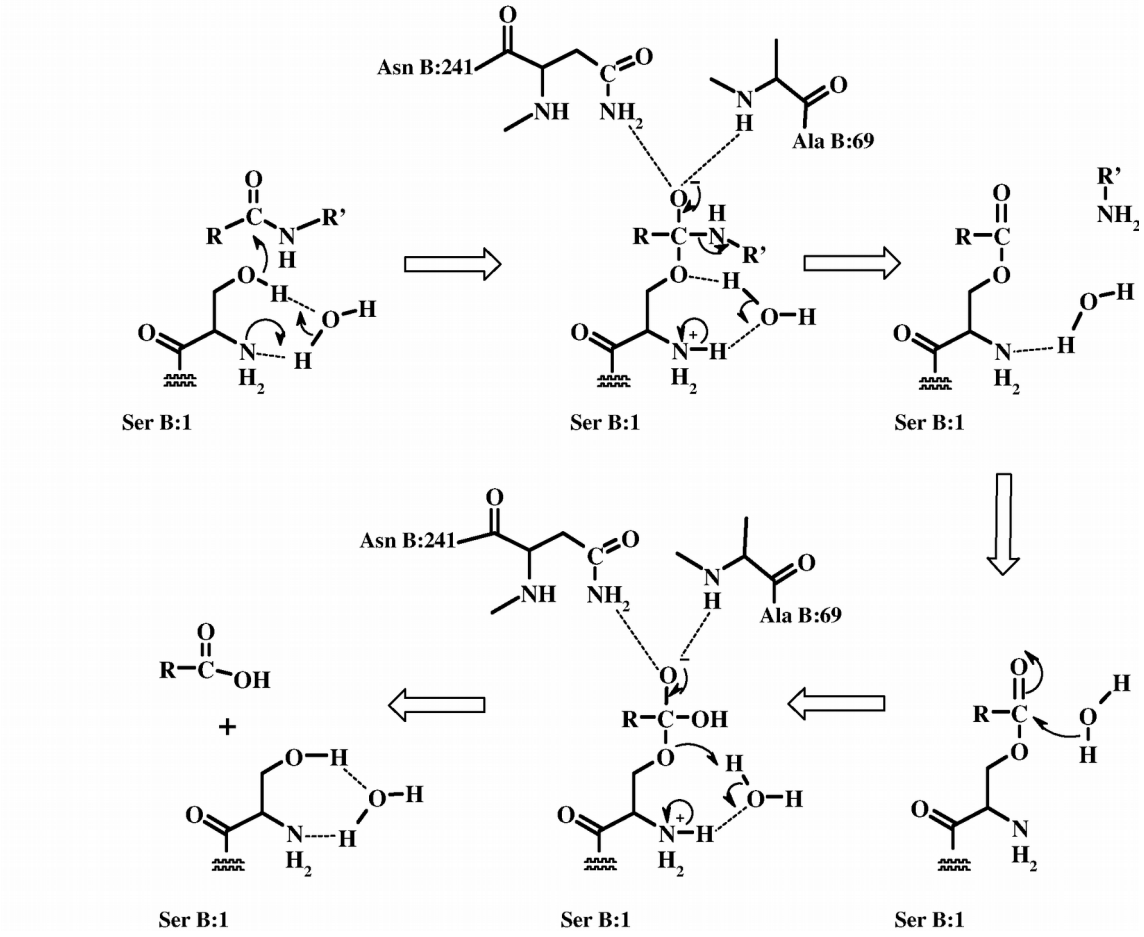


Figure 1.9 Catalytic mechanism of penicillin acylase in which a water molecule is used as a virtual base in the catalytic attack (Modified from Duggleby et al., 1995).

1.9 N-terminal nucleophile (Ntn) hydrolases

The traditional classification of proteinases and other amidohydrolases has been based on the residue that provides the reactive centre nucleophile. With the increasing number of protein structures being solved it is not unsurprising that this dogma has already been challenged by the discovery of the α/β-hydrolases that share a common fold in which the position of the nucleophilic residue is conserved, but the residue itself is variable. Another exception to the rule was the observation of the chymotrypsin fold in viral 3C proteinases that have a cysteine rather than serine as the nucleophile. It has been proposed that a proper classification should be based on the protein fold and the locations of the catalytic group (Murzin, 1996).

Recently PGA has been assigned to a new structural superfamily of enzymes, the N-terminal nucleophile (Ntn) hydrolases which includes glutamine PRPP amidotransferase (Sm*ith et* al., 1994), the proteasome (L*öwe et* al., 1995), aspartylglucosaminidase (Oino*nen et* al., 1995) and glucosamine-6-phosphate synthase (Isu*pov et* al., 1996). They are characterised by an N-terminal nucleophile which forms part of a four layer α + β structure with two anti-parallel β-sheets (I and II; figure 1.10), one essentially flat, the other slightly twisted (Brannig*an et* al., 1995).

The Ntn hydrolases share three common features characterised by a common fold (Figure 1.10) in which the nucleophile and other catalytic groups occupy equivalent sites, namely,

1) a β-sheet sandwich flanked on either side by a pair of α-helices
2) an N-terminal catalytic residue with the potential for nucleophilic attack
3) the capacity to autocatalytically activate itself from a precursor to unmask the N-terminal nucleophile.

The Ntn fold which is embedded in the β-subunit of PGA provides the framework in which the important catalytic residues sit (Figure 1.11). The absence of sequence similarity between these enzymes hides the evolutionary relationship which has only being unveiled by their structure determination. The common connectivity of sequential elements of secondary structure in these enzymes suggests that the Ntn hydrolases have diverged from a common ancestor (Brannig*an et* al., 1995).

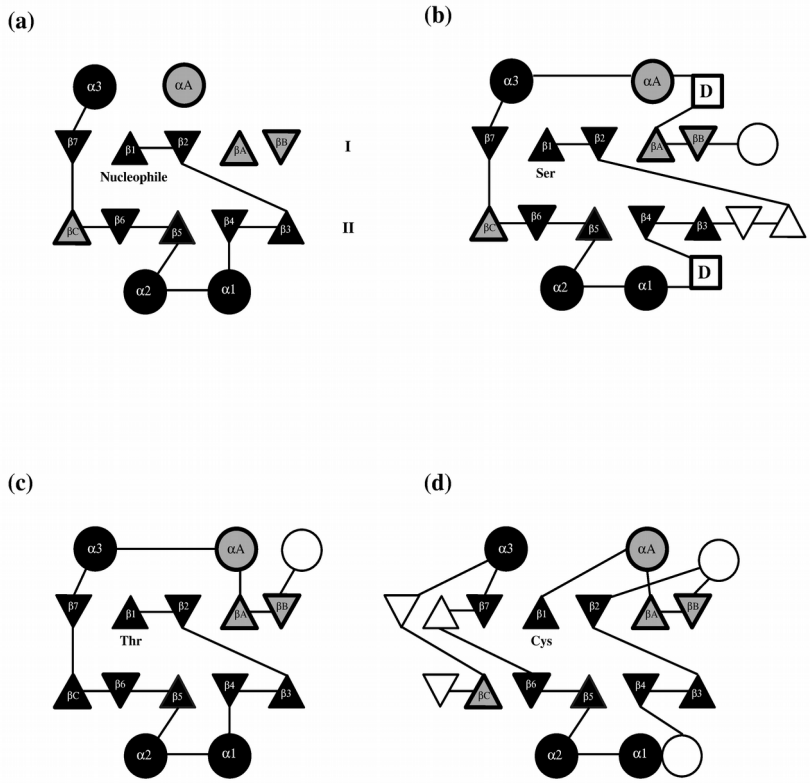


Figure 1.10 Topology diagrams for Ntn hydrolases. *α-helices are represented by circles and β-strands as triangles. Structural elements with common connectivity are coloured in black while those that are spatially equivalent but different connectivity are coloured grey. Structural elements that are spatially equivalent to the common core are depicted as open circles. (a) Ntn topology, (b) Penicillin acylase, D represents excursions from the common fold which form separate domains (c) 20S Proteasome (d) glutamine PRPP amidotransferase.*

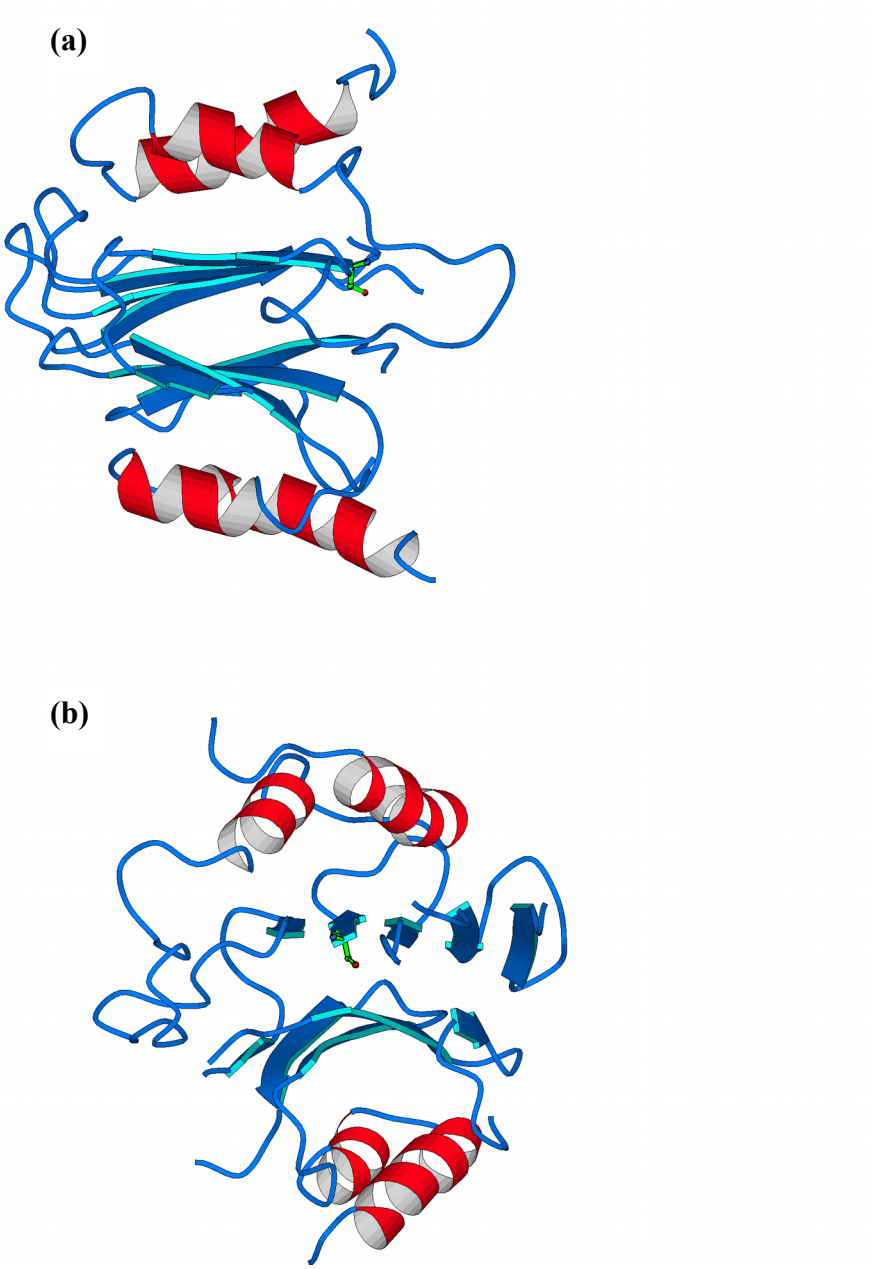


Figure 1.11 Organisation of the secondary structure of the Ntn fold of PGA. *(a) the β-sheet structure of the Ntn fold, (b) a perpendicular view of the fold*

1.10 Applications of PGA

In the 1950's it had been shown that penicillin acylase could hydrolyse penicillin G (Murao, 1955; Sakaguchi and Murao, 1950). Later Rolinson *et al*. demonstrated the practical importance of this enzymatic reaction, in which 6-APA was formed through the hydrolysis of penicillin G, providing an elementary starting point for the production of semi-synthetic penicillins (Rolin*son et* al., 1960). Indeed the naming of penicillin acylase has been a reflection of this reaction and its industrial application in the synthesis of antibiotics.

The main industrial use of PGA is to catalyse the hydrolysis of penicillin G (Pen G) to phenylacetic acid (PAA) and 6-aminopenicillanic acid (6-APA). It is also used for the hydrolysis of cephalosporin G, obtained from penicillin G through chemical ring expansion, to 7-aminodesacetoxycephalosporanic acid (7-ADCA) (Bald*aro et* al., 1988). Subsequently 6-APA and 7-ADCA are chemically acylated at the amino group with different side chains to yield a wide variety of semi-synthetic antibiotics. In comparison with other proteolytic or hydrolytic enzymes, the potential of PGAs toward "unnatural" substrates have been less fully explored, and their uses mainly confined to the industrial acylation-deacylation processes. The enzymatic cleavage of natural penicillins that was initially performed used cell free extracts and was not found to be commercially competitive with the more efficient chemical deacylation process.

The introduction of immobilised enzymes, which allows multiple use of the catalyst is now the method of choice. PGA represents one of the first successful applications of immobilised enzymes in the fine chemicals industry. However, new applications of PGA have been reported in recent years (Waldmann and Sebastian, 1994), that take advantage of its chemoselectivity in the hydrolysis of the phenylacetyl moiety and of its stereospecificity toward chiral carbon atoms. The broad substrate specificity of penicillin acylase has been exploited for a wide variety of industrial applications including the catalysed deprotection of peptides during their synthesis, resolution of racemic mixtures, and protection and deprotection of amino acids and sugars to expand the range of substrates which the enzyme will accept and to increase its catalytic efficiency towards these substrates. The chemoselectivity of PGA for phenylacetyl amides and esters has been utilised for the design of new protecting groups for amino-,

hydroxyl- and carboxyl- functionalities in peptide, carbohydrate and β-lactam chemistry (Bald*aro et* al., 1988; Waldmann, 1988; Waldm*ann et* al., 1992).

1.11 Protein engineering of penicillin acylase

Protein engineering has been used to alter the physico-chemical properties of PGA before its 3-dimensional structure was known. A wealth of sequence and biochemical information has been built. An important strategic goal of the protein engineering strategy for PGA has been to develop acylases that can efficiently hydrolyse natural β-lactam antibiotics other than penicillin G, such as cephalosporin C. Various research groups (Forney and Wong, 1989b; For*ney et* al., 1989a; Niersb*ach et* al., 1995a; *Roa et* al., 1994) have used both genetic engineering and classical mutagenic techniques to try to achieve this goal, but as yet no one has reached this objective, the possible reasons for this will be discussed later.

1.12 Aims of the project

Structural and mechanistic information, sequence comparisons, and site-directed mutagenesis continue to provide a basis for the rational design of new protein functions and the alteration of existing functions. Random mutagenesis and 'directed evolution' approaches, however, have proved highly successful for the fine tuning of properties such as enzyme substrate specificity. The rational redesign of an enzyme's catalytic site requires extensive knowledge of the catalytic mechanism and enzyme structure. For this reason, it is helpful to study structures of its complexes with various ligands and mutants with altered characteristics.

Penicillin acylase exhibits broad substrate specificity for the amide side chain of the substrate but in contrast show very precise recognition of the phenylacetyl side chain. A primary objective of this thesis was to investigate the interaction of penicillin acylase with the amide linked side chain of its substrates. This was to be undertaken by combining the techniques of mutagenesis and X-ray crystallography. A series of PGA mutants, characterised by either altered substrate specificity or by enzyme inactivity, were studied. The inactive mutants were the basis for probing substrate binding and to understand the molecular basis for enzyme activity. Mutants were used for kinetic studies with the aim of correlating structural data with their biochemical characteristics.

It is hoped that the contributions of the work described in this thesis would aid in the long term goal of engineering an enzyme with specificity towards the cephalosporin family of antibiotics.

Chapter Two Principles and Methods of Protein Design

A wealth of information can be gained regarding enzyme function using the combined approaches of mutagenesis, structural determination, and biochemical characterisation. This chapter describes the methods employed to study the substrate specificity of penicillin G acylase (PGA), combining both the techniques of genetic engineering and X-ray crystallography. Rational approaches to engineering proteins are usually organised as design cycles (Figure 2.1), where repeated cycles of the strategy aid in a further understanding of the specificity and function of PGA.

The cycle is usually entered by obtaining protein, usually by cloning of the gene followed by expression and purification of the recombinant protein. Following biochemical characterisation of the protein, its 3-dimensional structure can be determined by X-ray analysis if the protein can be crystallised. PGA has been both extensively characterised in solution and its crystal structure determined. The structure has provided the platform to pursue more incisive and focused experiments, and at the same time allowing the interpretation of past scientific findings.

Once the structure has been defined, a more speculative and uncertain step is required in which a mutation is introduced to test hypotheses about parameters such as specificity, binding affinity and enzyme function. Knowledge of the active site architecture provides a better starting point to identify possible candidates for mutagenic studies. It is at this stage the main body of the work described in this thesis begins. The next step in the cycle is to test the design hypothesis by site-directed mutagenesis followed by expression, biochemical characterisation and structural determination of the mutant protein. The ability to modify protein molecules by site-directed mutagenesis plays a major role in addressing many of the questions posed. The 3-dimensional structure of a protein-ligand complex can help unravel the key interactions and mechanism of enzyme catalysis and specificity. The methods involved in X-ray analysis and structure determination will be described in more detail later.

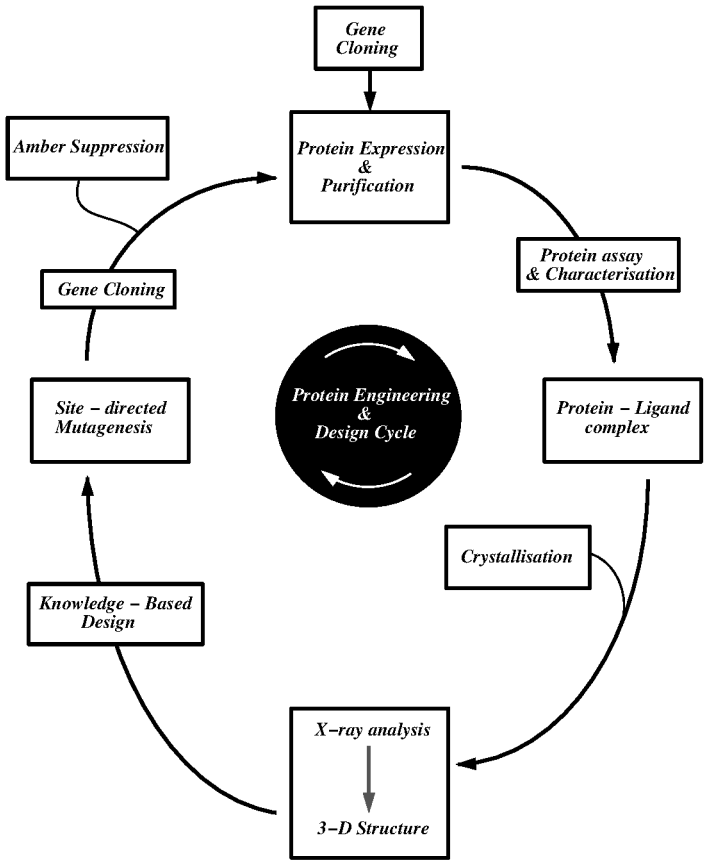


Figure 2.1 A schematic diagram outlining the protein engineering and design cycle*. The cycle encompasses the methods and principles employed to study the specificity of penicillin acylase.*

2.1 Purification and cloning of DNA

The advance of recombinant DNA techniques has been breath-taking over the last decade and has been instrumental in the development of protein engineering techniques. The manipulation of DNA is a prerequisite to the expression and study of recombinant proteins. Equally, the use of DNA modifying enzymes is a key requirement in carrying out these tasks. Nucleic acid modifying enzymes provide the foundation for many techniques in molecular biology. This group of enzymes is used to synthesise, degrade, join or remove portions of DNA in a controlled and defined manner. Combined with the use of restriction enzymes it enables the cloning of DNA into plasmid vectors which provide a vehicle for subsequent manipulations and protein expression.

The extensive manipulations of DNA molecules rely on their isolation and purification. The isolation of recombinant DNA constructs such as plasmids forms the basis of any molecular cloning experiment. Applications, such as cloning of DNA, frequently require the purification of DNA fragments from agarose gels or amplification reactions. This is achieved using silica-based resin which eliminates the need for organic solvent extractions. These resins can be tailored to specifically bind plasmid DNA or linear DNA fragments. The method is based on selective adsorption of DNA to the resin in the presence of high concentrations of chaotropic salts and allows for elution of the DNA, after the removal of contaminants, with low ionic strength solutions or simply into sterile water which avoids complications of inappropriate buffers when carrying out subsequent DNA manipulations.

2.1.1 Purification of plasmid DNA

Plasmid DNA can be isolated from a bacterial culture (2-3ml) that has been transformed with the plasmid in which antibiotic selection is used to maintain the plasmid in the cell. Purification of plasmid DNA was carried out using the Wizard™ miniprep DNA purification system (Promega). Purification of plasmid DNA isolated from bacterial strains carrying the *end*A1 genotype (product of *end*A1 gene is endonuclease I which can use dsDNA as a substrate) used the Wizard™ SV miniprep DNA purification system (Promega). Both systems were used according to the manufacturers' instructions.

2.1.2 Purification of DNA fragments from agarose gels

Purification of DNA from agarose gels has a number of advantages where the isolation of specific DNA fragments is an effective method to selectively clone specific sequences of DNA. The DNA fragment can be generated through specific excision of the DNA using restriction enzymes. DNA fragments were purified from agarose gels using either the Wizard™ PCR preps DNA purification system (Promega) or, if DNA yield was low the Prep-A-Gene kit (Biorad). Both systems were used according to the manufacturers' instructions.

2.1.3 Determination of DNA concentration

DNA concentration was determined spectrophotometrically at λ = 260 nm. An optical density (OD) of one absorbance unit for dsDNA was approximated to be equivalent to 50 μg/ml. For ssDNA 1 OD unit was approximated to 37 μg/ml.

2.1.4 DNA cloning

All mutants were subcloned from the mutagenesis vector Bluescript KS (-) (Figure 2.2) using the appropriate restriction enzymes. The fragment of the *pac* gene containing the respective mutation was cloned into the wild type expression clone pA1 (*pac* in pACYC184) replacing the existing non-mutated fragment. Positive clones were identified by restriction analysis of unique restriction sites incorporated in the mutations. All restriction enzymes were used according to the manufacturers instructions. A typical restriction enzyme digest was set up as follows, plasmid DNA (500 ng) in restriction buffer containing acetylated bovine serum albumin (0.1 mg/ml) and restriction enzyme (5 U) and made up to a final volume of 15 μl. The reaction was incubated at 37 °C for 1-2 h and analysed by agarose gel electrophoresis. Cloning into a single restriction site entailed additional modification of the plasmid vector to prevent self-ligation by dephosphorylating the 5'-phosphate. Following a restriction digest with a single enzyme the 5'-protruding end of the plasmid DNA was dephosphorylated with shrimp alkaline phosphatase (0.1U; USB) in 20 mM Tris-HCl pH 8.0, containing 10 mM $MgCl_2$ for 1 h at 37 °C followed by heat inactivation at 65 °C for 15 min.

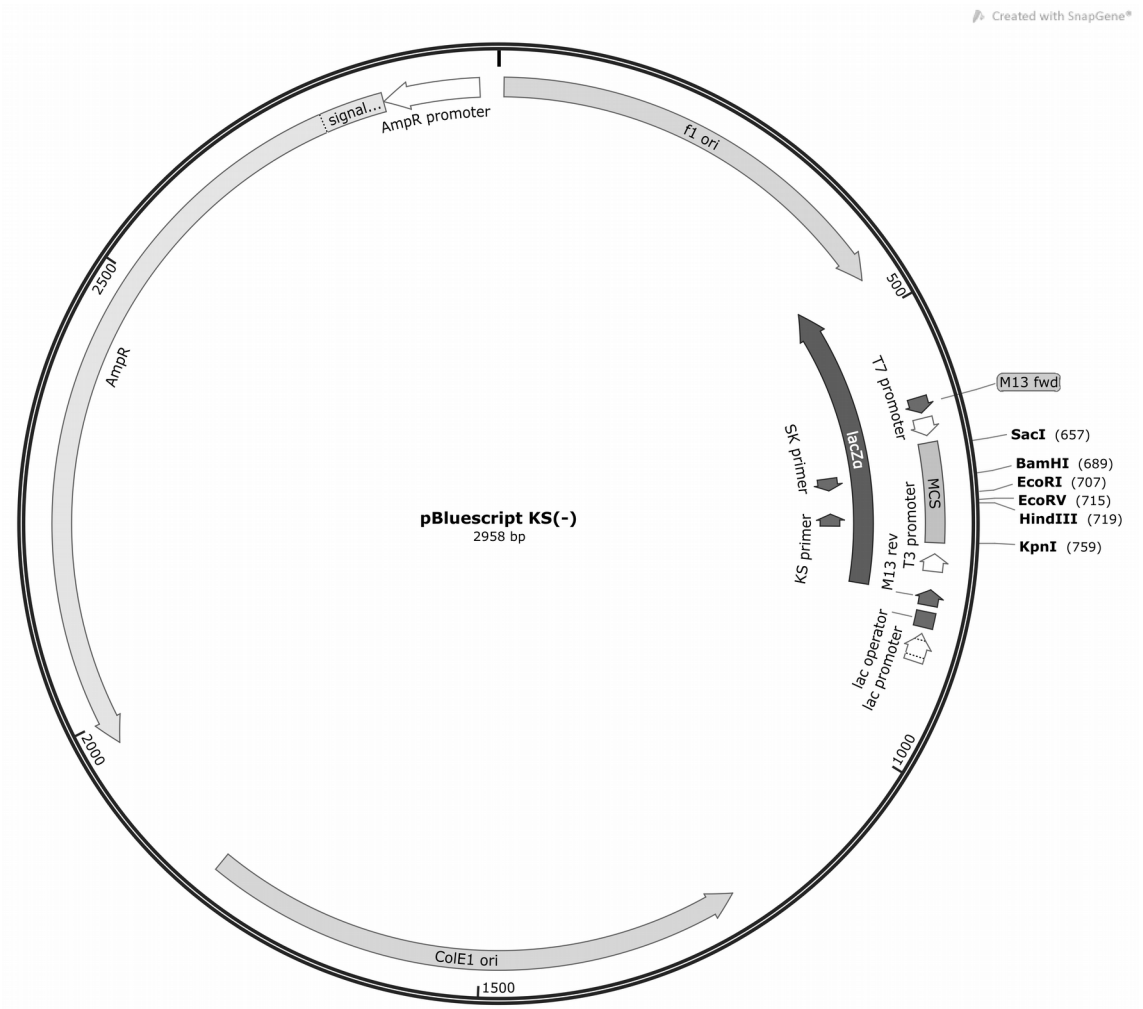


Figure 2.2 A plasmid map for the cloning vector pBluescript KS (-). *The 3.24 kb pac gene was cloned into the BamHI restriction site.*

2.1.5 DNA ligation

Following restriction and purification of DNA fragments, vector DNA and insert DNA were ligated using the enzyme T4 DNA ligase (Promega). Vector DNA (50 ng) was ligated to insert DNA (100 ng) in a molar ratio of 1 : 3 in ligase buffer (Promega) and T4 DNA ligase (5U) in a final volume of 10 μl. The reaction was incubated at 14 °C overnight followed by incubation at 65 °C for 10 min to inactivate the ligase prior to transformation.

2.1.6 DNA agarose gel electrophoresis

DNA was analysed by agarose gel electrophoresis. Briefly, DNA samples (5 μl) were mixed in (x6) DNA sample buffer (1 μl) and loaded onto a 0.8% agarose gel (0.8g agarose; Gibco BRL, in 100 ml of TAE buffer) and electrophoresed in TAE buffer at 120 V for 1 h using a DNA gel electrophoresis apparatus GNA-100 (Pharmacia). The gel was stained in TAE buffer containing ethidium bromide (15 μg/ml) for 10 min followed by destaining in TAE buffer for 30 min. The gel was visualised using the ImageStore 7500 gel documentation system (UVP).

2.2 Mutagenesis

Site-directed mutagenesis is a technique by which defined mutations can be made *in vitro* in a cloned plasmid DNA gene. The desire to perform site-directed mutagenesis is fuelled by a need to manipulate genes and evaluate the effect of the proposed mutation. Oligonucleotide-directed mutagenesis is by far the most commonly used method. An oligonucleotide encoding the desired mutation(s) is annealed to one strand of the DNA of interest and serves as a primer for initiation of DNA synthesis. In this manner, the mutagenic oligonucleotide is incorporated into the newly synthesised strand. In theory, this should result in a 50% efficiency of mutagenesis due to the semi-conservative nature of replication. Observed efficiencies, however, are much lower due to factors such as incomplete strand synthesis, random priming of DNA synthesis, oligonucleotide primer displacement and host repair mechanisms.

A number of methods introduce genetic tricks to improve mutagenesis efficiency. These selection and enrichment methods include use of a uracil-containing DNA strand

which can be selectively degraded *in vivo*, and dNTP analogue incorporation which can render one strand of heteroduplex DNA impervious to digestion. The preparation of DNA template to which the mutant oligonucleotide is hybridised can take two forms, either using dsDNA template which is denatured prior to hybridisation or the synthesis of ssDNA by phage infection. The preparation of ssDNA was the method of choice employed for mutagenesis experiments mainly because of the greater hybridisation efficiency.

2.2.1 Preparation of single-stranded DNA

The pac gene cloned into the plasmid vector Bluescript KS (-) (Stratagene) which contains a F' origin of replication (Figure 2.2) was transformed into the *E. coli* strain XL-1 Blue MRF' (Stratagene). A single transformed colony was then grown overnight in LB media (5 ml) containing ampicillin (100 μg/ml) at 37 °C in an orbital shaker (170 rpm). The inoculum (200 μl) was added to 2TY media (20 ml) containing ampicillin (100 μg/ml) and was grown for 1 h at 37 °C in an orbital shaker (200 rpm). After the addition of R408 helper phage (80 μl, 2 x 10^{11} pfu/ml; Promega) the culture was shaken for a further 6 h and with the addition of ampicillin (100 μg/ml) was left for a further 14 h or overnight The culture was centrifuged (6,000 x g, 10 min, 4 °C) using a Sorvall RC 5B plus centrifuge and the supernatant transferred to SS34 tubes.

Phage particles were precipitated by the addition of PEG/NaCl solution (0.2 vol) and incubated on ice for 1 h. The phage was pelleted by centrifugation (6,000 x g, 20 min) and the supernatant removed. Centrifugation was repeated for a further 5 min and any remaining supernatant removed. The phage pellet was resuspended in TE (500 μl) by brief vortexing at intervals over 1 h on ice. The resuspension was transferred to a 1.5 ml eppendorf tube and centrifuged (12,000 x g, 30 sec) to remove any remaining cells. The supernatant was transferred to a new eppendorf tube and PEG/NaCl solution (200 μl) added and incubated at room temperature for 15 min. The phage was pelleted by centrifugation (12,000 x g, 5 min) and the supernatant removed. This was repeated several times to ensure complete removal of PEG/NaCl solution.

The pellet was resuspended in TE (500 μl). An equal volume of TE saturated chloroform was added to lyse the phage and vortexed for 2 min and centrifuged in a microcentrifuge (12,000 x g, 5 min). This step also removes excess PEG/NaCl soln.

The upper aqueous phase (containing phagemid DNA) was transferred to a fresh tube leaving the interface behind and an equal volume of TE saturated phenol added. The solution was vortexed for 2 min and centrifuged as before. The aqueous phase was transferred to a fresh tube and the chloroform extraction repeated as before to remove excess phenol. To the aqueous phase 0.1 vol of 3 M sodium acetate pH 4.6 and 2.5 vol of ice cold ethanol were added. The phagemid DNA was precipitated at -70 °C for 1 h and pelleted by centrifugation (12,000 x g, 10 min). The supernatant was removed and the pellet carefully rinsed with 70% ethanol and centrifuged for a further 2 min. The pellet was dried briefly at 37 °C and resuspended in sterile deionised water (100 μl). The concentration of ssDNA was determined as described.

2.2.2 Oligonucleotide design and phosphorylation

Oligonucleotides were synthesised (40 nm scale) on an ABI 392 DNA/RNA synthesiser. The oligonucleotides were purified using a NAP-10 desalting column (Pharmacia) to remove the ammonia by-product and eluted in sterile deionised water.

Oligonucleotide concentration was determined by measuring the absorbance at $\lambda = 260$ nm. Using the Beer-Lambert law and the molar absorption coefficient ε, (A = 15.4, T = 9.7, G = 11.4, C = 9.2 M^{-1} cm^{-1}) determined by the summary of contributions of each nucleotide base in the oligonucleotide, the absorbance can be converted into pmoles/μl.

Preparation of ssDNA from Bluescript KS (-) produced the coding strand of the *pac* gene. The TAG 146 oligonucleotide was designed complementary to this sequence apart from its mismatch mutation site. The proper design of the mutagenic oligonucleotide is critical to the efficiency of the reaction. The following guidelines were taken into account when designing the oligonucleotide. For efficient hybridisation, there needs to be 100% base pairing at either end of the target sequence without secondary structure formation. For small substitutions, 10-15 bases hybridising on either side of the mismatch are usually sufficient. The composition of the 3'-end of the primer is particularly important as polymerases do not typically extend from a mismatched or poorly hybridised 3'-end.

The oligonucleotide sequence used for the amber mutagenesis experiment is shown below.

TAG-146 non-coding sequence

5'-ATC AAT TTC ACT AGT GCT ATC AGA CTA GCG GTT TGC CAT G-3'

***Spe*I** amber stop codon

To increase the efficiency of incorporation of the mutant oligonucleotide, the 5'-end of the oligonucleotide was phosphorylated using T4 polynucleotide kinase. The 5'-phosphorylation of oligonucleotide (100 pmol) was carried out in kinase buffer (76 mM Tris-HCl, pH 7.6; 10 mM $MgCl_2$ and 5 mM DTT) with the addition of 1 mM ATP and T4 polynucleotide kinase (5 U) in a final volume of 25 μl. The reaction was incubated for 30 min at 37 °C followed by inactivation of T4 polynucleotide kinase by incubation at 70 °C for 10 min.

2.2.3 GeneEditor mutagenesis system

The overall strategy for mutagenesis with the use of mutant enrichment by antibiotic selection is outlined in Figure 2.3. Mutagenesis of the *pac* gene was carried out using the GeneEditor™ *in vitro* site-directed mutagenesis system (Promega) which uses positive selection for antibiotic resistance to allow high efficiency oligo-directed mutagenesis. The selection involves altering the substrate specificity of TEM-1 β-lactamase, the enzyme responsible for bacterial resistance to β-lactam antibiotics. During the mutagenesis reaction, the ampicillin resistance gene is mutated to give resistance to the GeneEditor™ antibiotic selection mix. The desired mutations from the gene of interest are also made in the same mutagenesis reaction and both mutagenic oligonucleotides must hybridise to the same DNA strand. Taking advantage of the degeneracy of the genetic code, the incorporation of a unique restriction site (*Spe*I in TAG-146) in the mutant oligonucleotide without altering the amino acid sequence can serve as a quick means to confirm the identity of clones containing the desired mutation in the mutagenesis experiment.

Protocol (Figure 2.3)

1. Preparation of ssDNA template.
2. Synthesise the new mutant strand, with the desired mutagenic oligonucleotide and the new antibiotic resistance selection oligonucleotide.

3. Transform into competent BMH 71-18 *mut*S in the presence of GeneEditor™ antibiotic selection mix, where only mutants will grow.

4. Miniprep DNA and transform into competent JM109. Plate out on LB plates with GeneEditor™ antibiotic selection mix.

5. Screen mutants by restriction analysis to identify clones carrying desired mutation.

The protocols described above followed the manufacturers' instructions with the exception of the antibiotic selection mix where the recommended amount added (100 μl) was decreased by 50% as suggested in Promega Notes (Kopps, 1998) since bacteria containing the altered β-lactamase mutants also exhibit some sensitivity to the antibiotic selection mix at high concentrations.

2.3 Bacterial transformation

Bacterial transformation is the incorporation of plasmid DNA into a bacterial host cell which permits the transfer of a gene of interest for the purposes of protein expression and also the propagation of plasmid DNA for gene cloning and DNA manipulations. The plasmid construct pA1 containing the 3.24 kb *pac* gene cloned into the plasmid pACYC184 (New England Biolabs) was provided by Dr. Jim Brannigan (Figure 2.4). For expression and subsequent purification of the enzyme the plasmid pA1 was transformed into the *E. coli* strain BL21 (DE3) (Novagen). Preparation of competent cells was carried out by the $CaCl_2$ method (see below).

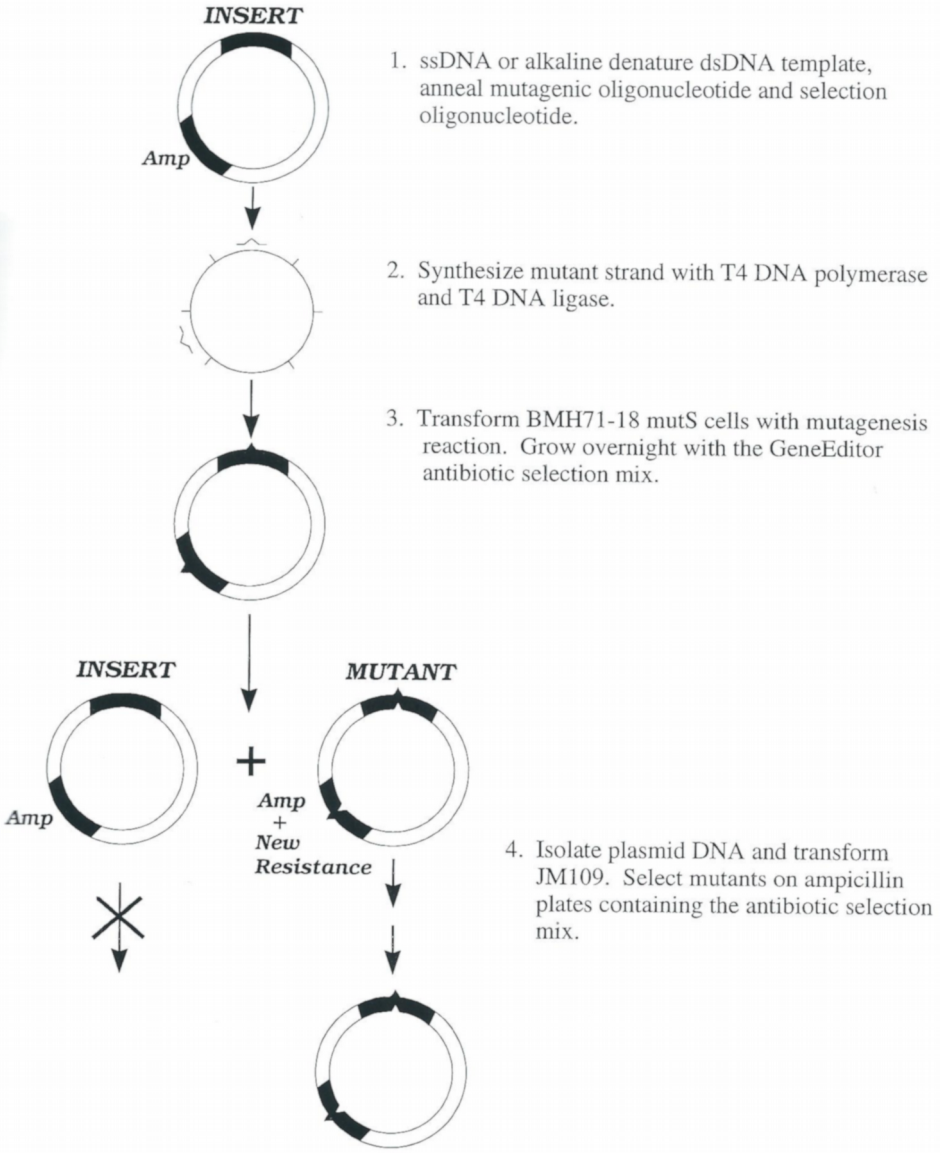


Figure 2.3 A schematic diagram describing the procedure for the GeneEditor™ in vitro site-directed mutagenesis system. *Amp: Ampicillin resistance.*

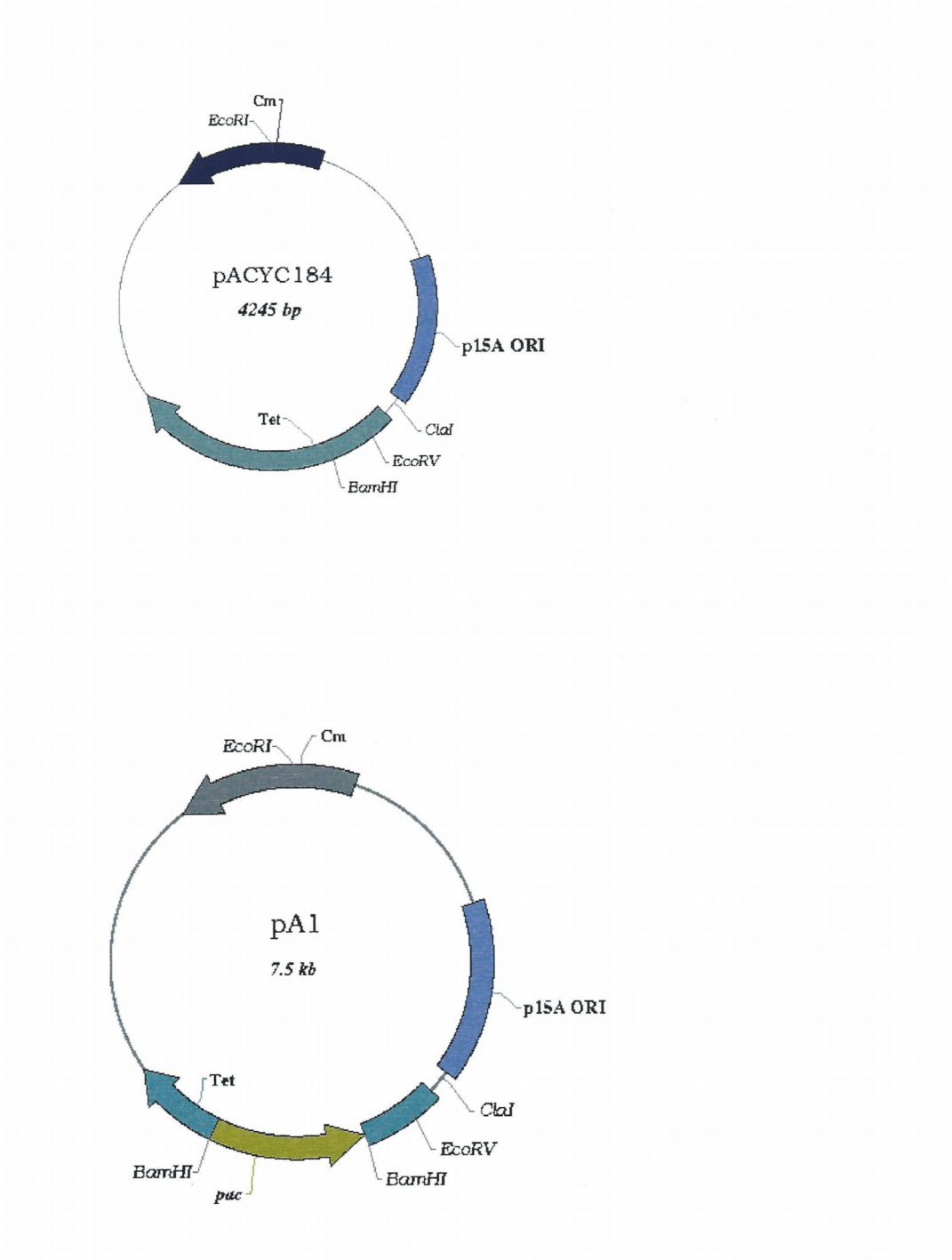


Figure 2.4 A plasmid map of the expression vector pACYC184. *The pac gene is cloned into the BamHI restriction site to give the Cm^r Tc^s plasmid pA1. Cm^r, chloramphenicol resistant. Tc^s, tetracycline sensitive.*

2.3.1 Preparation of competent *E. coli* cells

An overnight culture (5 ml) of host cells was grown at 37 °C. 100 ml of sterile LB media were inoculated with 1 ml of the overnight culture and grown at 37 °C until the cells reached the mid-log phase, usually between an O.D of 0.2 – 0.5. This took usually 1 h 45 min for BL21 (DE3) and 2 h for XL-1 Blue MRF' host cells. The cells were centrifuged gently (3000 x g, 5 min, 4 °C) and after removal of all the supernatant the cell pellet was resuspended in ice-cold sterile 100 mM $CaCl_2$ (0.5 vol). The cells were incubated on ice for 1 h. The cells were again centrifuged and the pellet resuspended gently in $1/50^{th}$ volume of ice-cold 100 mM $CaCl_2$ and left on ice for at least 2 h.

100 μl aliquot of competent cells was used for each transformation. Plasmid DNA (50 ng) or 10 μl of a ligation reaction were added to the cells and mixed gently. The cells were left on an ice/ethanol bath for 30 min followed by heat shock at 42 °C for 1 min 30 sec and left on ice for a further 2 min. The cells were plated on LB agar plates containing the appropriate selection antibiotic and incubated overnight at 37 °C.

2.3.2 Colony PCR screening

Colony-PCR provides an efficient method to screen a library of bacterial colonies without the need of isolating the plasmid DNA. PCR is a powerful technique that uses DNA oligonucleotides (primers) to identify complementary DNA sequences In this case two primers (PA146-seq and WRT-bio primer sequences are given in the Appendix) complementary to the *pac* gene produces a 900 bp DNA product and is used to discriminate between clones which contain the gene and those which do not. The colony PCR was carried out on an *E. coli* B strain which does not contain an intrinsic copy of the *pac* gene.

Identification of positive clones was carried out by colony PCR as described in the pET system manual (#TB055, Novagen). Briefly, a bacterial colony was selected and transferred to a 1.5ml eppendorf containing sterile deionised water (50 μl) and vortexed to disperse the cells. The sample was heated to 95 °C for 5 min to lyse the cells and denature DNases. The sample was centrifuged (12,000 x g, 1 min) to remove cell debris and the solution (10 μl) transferred to the PCR reaction. The PCR profile is outlined below (Table 2.1):

PCR reaction protocol	(μl)
DNA sample	10
ddH_2O	21
x10 PCR buffer (Promega)	5
$MgCl_2$ (25 mM, Promega)	8
PA146-seq (12 pmoles/μl)	2.5
WRT-bio (15 pmoles/μl)	2
dNTPs (10 mM, Sigma)	1
Taq DNA polymerase (Promega)	0.5
Final volume	50

Table 2.1 PCR reaction protocol.

The reaction is overlaid with 1-2 drops (20-40 μl) of nuclease-free mineral oil (Sigma) to prevent condensation and evaporation.

The primer sequences are the main consideration in determining the temperature of the annealing step of the PCR amplification cycle. The annealing temperature is dependent upon the primer with the lowest melting temperature and is based on the formula below.

$$T_M = 81.5 + 16.6 (\log_{10}[Na^+]) + 0.41(\%G+C) - 675/n$$

where $[Na^+]$ is the molar salt concentration: $[K^+] = [Na^+]$ and n = number of bases in the oligonucleotide. In general, one minute is allowed for every 1kb of DNA during the extension step.

PCR cycle parameters		
Step	time (min)	temperature (°C)
denaturation	1	94
primer annealing	1	60
primer extension	1	72

Table 2.2 PCR cycle parameters

The PCR reaction was heated to 95 °C for 5 min and placed in a Perkin Elmer Cetus DNA thermal cycler for 25 cycles (Table 2.2) followed by a final extension of 10 min at 72 °C. The purification of PCR product used the Wizard™ PCR preps DNA purification system (Promega) following the manufacturers' instructions.

2.4 Expression and purification of Penicillin acylase

PGA is a periplasmic protein and optimal expression of the protein occurs between 25 to 30 °C (Bhattacha*rya et* al., 1993). To analyse the expression of PGA and its mutants, initial expression trials were performed on a small scale. Once conditions had been determined and mutants characterised, protein was expressed and purified on a large scale. Since PGA is a periplasmic protein, this has enabled isolation of the periplasmic fraction, which reduces background protein levels from cytoplasmic proteins, and hence the number of steps required for purification of the protein to homogeneity.

2.4.1 Small-scale periplasmic extraction

Isolation of periplasmic protein was carried out by osmotic shock extraction. 5ml cell culture was grown for two days at 28 ºC. 1.8 ml of culture was added to a 2 ml eppendorf containing 180 μl of 1 M Tris-HCl pH 8 and incubated at room temperature for 10 min. The cells were centrifuged (14,000 x g, 2 min) and the pellet resuspended in 300 μl protoplast media (40 % sucrose, 2 mM EDTA in 50 mM Tris-HCl pH 8) and mixed constantly for 10 min. The sample was centrifuged (14,000 x g, 3 min) and the

supernatant removed. Centrifugation was repeated to ensure all the supernatant was removed. The pellet was resuspended in ice-cold sterile deionised water (ddH_2O) and left on ice for 10 min. The cell debris was removed by centrifugation and the supernatant transferred to a fresh 1.5 ml eppendorf. A sample (24 μl) was added to SDS loading buffer (8 μl) and incubated at 95 ºC for 3 min and analysed by SDS-PAGE.

2.4.2 Large-scale purification of PGA

To simplify the purification process, PGA was extracted from the periplasmic space of the cells by osmotic shock using a modification of published protocols (Rodríguez *et al*., 1992)(Barb*ero et* al., 1986). Wild-type and mutant acylases with the exception of the mutant His A:146 were purified from the periplasmic cell fraction by FPLC. Briefly, cells were grown at 28 ºC in six litres of Luria-bertani broth (LB, see Appendix) containing chloramphenicol (35 μg/ml) and $CaCl_2$ (250 μM) for at least 48 h. Prior to harvesting of cells, the cell culture was incubated in 0.1 M Tris-HCl pH 8 for 20 min which causes the cell wall to become brittle and allows easier separation of the periplasm fraction by osmotic shock.

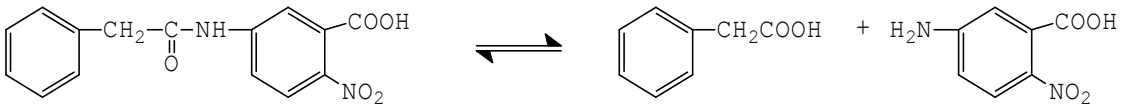
Cells were harvested by centrifugation (8,000 x g, 15 min, 4 ºC) and the cell pellet resuspended in 900 ml of protoplast media (40% sucrose, 2 mM EDTA in 50 mM Tris-HCl pH 8) and stirred for 1 h. The cells were again centrifuged (10,000 x g, 30 min, 4 ºC) and the pellet subjected to osmotic shock by the addition of ice-cold water (1 L). The cell suspension was stirred at 8 ºC for 1 h to release the periplasmic fraction. Cell debris was removed by centrifugation (10,000 x g, 20 min, 4 ºC).

The extracted enzyme supernatant was further purified by ammonium sulphate fractionation at room temperature. The protein that precipitated between 50 and 70% ammonium sulphate was resuspended in 50 mM Tris-HCl pH 7.5 containing 1.1 M ammonium sulphate and loaded onto a Hiload 26/10 Phenyl Sepharose HP column (Pharmacia) equilibrated in the same buffer. PGA protein was eluted from the Phenyl-Sepharose column with 50 mM Tris-HCl pH 7.5 and a linear gradient (1.1 M to 0) of ammonium sulphate. Fractions of 10 ml were collected and the elution profile monitored using a Pharmacia UV-1 monitor. The main peaks containing PGA were identified by SDS-PAGE and activity with NIPAB substrate. The protein fractions

were pooled and dialysed against 100 mM Tris-HCl pH 8.5 for wild-type, pH 8.9 for mutants as they characteristically do not bind at the lower pH, and applied to an anion-exchange Hiload 26/10 Q-Sepharose HP column (Pharmacia) equilibrated with the same buffer. Unretained protein was eluted with 100 mM Tris-HCl pH 8.5 and retained protein eluted with a linear gradient (0 to 0.5 M) of NaCl. Fractions of 5 ml were collected and the elution profile monitored. Fractions were assayed and/or analysed by SDS-PAGE.

2.5 Assay of enzyme activity

To analyse and correlate the effects of mutations on protein function in wild type PGA, a simple assay was employed using a chromogenic substrate. Enzyme activity was determined using the substrate 6-nitro-3-phenylacetamidobenzoic acid (NIPAB; Boehringer Mannheim/Sigma) following the spectrophotometric release of 3-amino-6-nitrobenzoic acid (ANB) which has a molar absorption coefficient $\varepsilon = 9090\ M^{-1}\ cm^{-1}$ at $\lambda = 405$ nm (Kutzbach and Rauenbusch, 1974).



NIPAB PAA ANB

2.6 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

PGA was analysed on a SDS-12 % polyacrylamide gel, according to Laemmli [1970 #898]. The gel was prepared using the Biorad SDS-PAGE apparatus with 1 mm spacers and the samples electrophoresed alongside Biorad low molecular markers at 200 V for 1 h. Protein bands were visualised by staining the gel with 0.25% coomassie blue R in 10% (v/v) glacial acetic acid and 25% isopropanol (v/v). The destaining solution contained 7% (v/v) glacial acetic acid and 5% (v/v) isopropanol. PGA protein expressed in *E. coli* was signified by a M_r = 23,500 band and a M_r = 62,000 band appearing in the SDS-PAGE gel corresponding to the A subunit and B subunit respectively.

2.7 Determination of protein concentration

Spectroscopic determination of protein concentration was performed using ε_{280} = 197,000 M^{-1} cm^{-1} (Sl*ade et* al., 1991), assuming a molecular weight of 86 kDa for the mature protein.

2.8 Crystallographic methods

The methods of protein crystallography are graphically summarised in Figure 2.5. The methods employed in proceeding from crystal to final model will be briefly discussed, emphasising those pertaining to the methods used in this thesis, including crystallisation, cryocrystallography, data collection, molecular replacement and refinement.

The limiting factors to advancing through the many hurdles of structure determination include

1) Crystal growth
2) Crystals quality
    a) disorder
    b) resolution limit
3) Phase problem

Lack of phase information is an inherent problem associated with X-ray data since the intensities are measured but not their phases. However for this work it is not an obstacle since there is a structure of PGA and initial phase information can be derived by molecular replacement.

4) Large systems

Working with large systems, such as proteins, in the past has provided problems both for computational and crystallographic methods. However, the growing power of

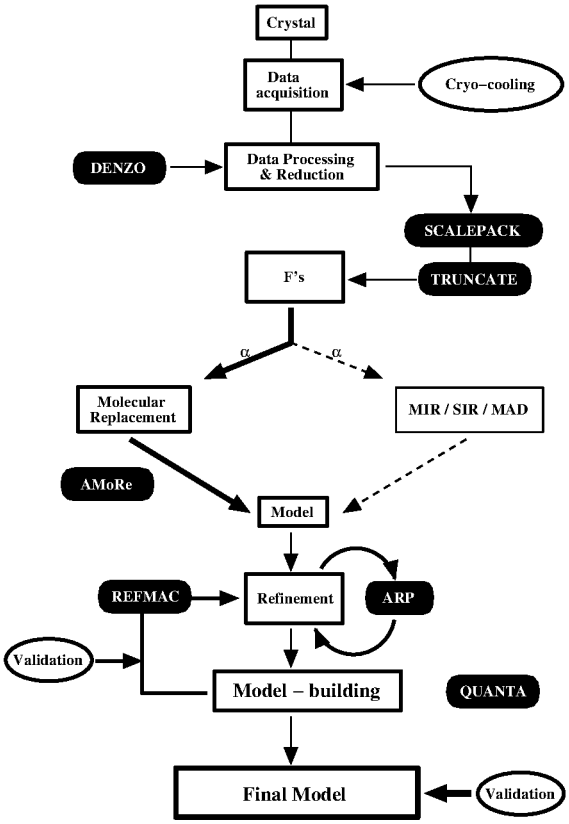


Figure 2.5 Flow chart of crystallographic methods and principles.

computing, aided by improvements in computational and crystallographic methods has made tractable work with very large proteins such as PGA (86 kDa). Areas relevant to this thesis include

- Crystal growth
- Data quality and the advantages of synchrotron radiation sources and 2-D detectors
- Phase problem and Molecular replacement
- Refinement, with maximum likelihood and automated refinement procedures
- Validation

2.8.1 Crystal growth

In the crystallisation of proteins two critical steps, nucleation of the initial seed and the enlargement of this seed, determine the quality of the final crystal (Weber, 1997), Figure 2.6. However, the degree of supersaturation required to nucleate crystals is often higher than the optimal concentration necessary for enlargement. Given that ideal conditions for nucleation and growth differ, a logical crystallisation strategy involves the independent optimisation of these processes. This can be accomplished by seeding, a technique where crystals are transferred from nucleation conditions to those that will support only growth. The technique of streak seeding as described by Stura and Wilson (Stura and Wilson, 1991), uses a probe to transfer seeds from crystals to a drop of non-nucleated protein. The probe consists of a rabbit or cat whisker mounted with wax to a pipette tip. The whisker is used to touch an existing crystal and dislodge seeds from it. Some of the seeds remain attached to the whisker as it is drawn out of the solution, and these can be introduced to a pre-equilibrated drop which may support only growth. Seeding can be employed in situations were crystals are small and of poor quality, where they take too long to grow and when the crystallisation is not very reproducible. Streak seeding also provides a powerful tool for cross-seeding between macromolecular species, such as genetic variants (Stura and Wilson, 1990). Crystallisation of penicillin acylase combines the techniques of hanging drop vapour diffusion and streak seeding resulting to give two crystal forms of PGA. This was determined by the presence or absence of ethylene glycol, a commonly used cryoprotectant.

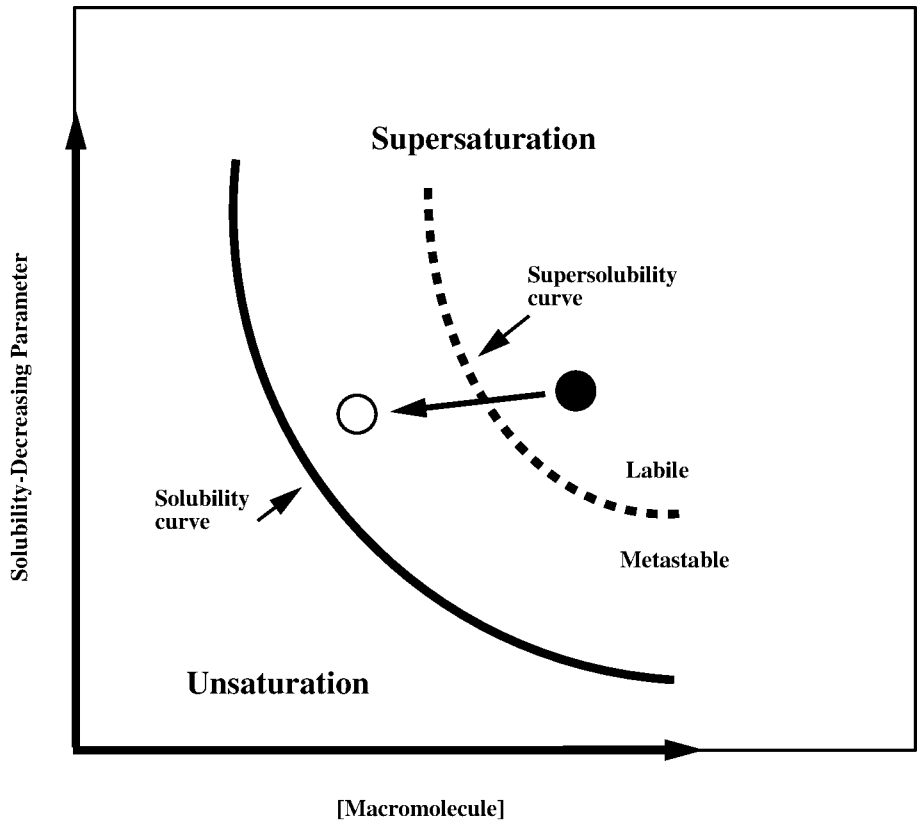


Figure 2.6 Phase diagram showing a schematic representation of crystallisation by seeding. *The solid circle represents conditions of supersaturation where nucleation and initial growth occur. The open circle is sufficiently supersaturated to support only growth. Figure reproduced from Weber (1997).*

2.8.2 Data quality

The general improvement in the quality of diffraction data available in recent years can be attributed to the use of synchrotron radiation sources, and the application of two-dimensional detectors, combined with the development of cryogenic freezing techniques.

Native structures, which cannot themselves prove the inherent function of a protein, do at least provide a 'road map' for rational mutagenesis experiments that can link structure to function. To obtain the most from an electron density map, the details

should be clear and interpretable, and so the importance of obtaining as a high a resolution as possible for structural determination has a significant bearing on the rational design and engineering of proteins to better understand the molecular basis for their function. The problem of collecting high resolution data for macromolecules is their inherent disorder. Their size means that they can no longer close pack in the crystal. As a result, the interstices are filled with disordered solvent, the lattice forces are weak and often the surface residues of the protein show both static and thermal disorder. This exacerbates the weakness of the data at high resolution.

A knock-on effect of the solvent content of protein crystals is the sensitivity to radiation damage where direct or primary radiation damage to the crystal can have a secondary effect caused by the diffusion of the resulting radicals and ions, from both protein and solvent atoms, through the solvent channels. However, use of synchrotron facilities, coupled with the development of cryogenic freezing techniques can alleviate some of the problems. The opportunity to reach the goal of atomic resolution has been realised for an increasing number of proteins. A general definition of atomic resolution according to Sheldrick (Sheldrick, 1990) is that the data should extend to at least 1.2 Å, and at least 50% of data in the outer resolution shell should have intensities of $>2\sigma$, which roughly corresponds to a merging R factor of ~25 %.

Obtaining a structure at atomic resolution offers many advantages, revealing the finer details of protein structure. The information that can be unveiled varies with the limiting resolution, Table 2.3. The need for an accurate structure is particularly important and provides a standard by which meaningful comparisons with the structures of mutants and substrate complexes could be made with a high degree of certainty. It is hoped that clear interpretation of structural differences can then be made, both large and, more importantly, subtle changes within the active site.

Similarly, X-ray detectors play an important role in protein crystallography and a high performance is demanded especially when the molecular weight of specimens become larger and/or crystal sizes available become smaller. The development of the X-ray storage-phosphor imaging-plate detector (IP) has enabled protein crystallographers to obtain more accurate data with a reduced X-ray dosage and a shortened exposure time (Amemiya, 1997). This reduction minimises the radiation damage to the crystals and their instability during X-ray exposure and gives a better signal-to-noise ratio due to a

high detective quantum efficiency (DQE). A combination of synchrotron radiation and imaging-plates has mutually enhanced their capabilities. The advent of more intense X-rays from the use of insertion devices at synchrotron radiation sources is complemented by the high DQE of the image plate.

Resolution (Å)	Protocol	Features identifiable
1.0	Full-matrix, anisotropic atoms	Fully resolved atoms
1.5	Border between isotropic/anisotropic	Hydrogens, disorder, ordered atoms distinguished
2.0	Isotropic atoms, stereochemical restraints	Some disorder
2.5	Isotropic model starts to break down	Shape of small groups
3.0	Rigid groups, some constraints	Shape of fragments, e.g. helices
6.0	Complete domains as rigid bodies	Globular protein

Table 2.3 The effect of resolution in the refinement of protein structures. Reproduced from Dauter, 1997.

2.8.3 Cryocrystallography

In recent years there have been many advances and improvements in crystallisation methods and data collection techniques, none more so than in the field of cryocrystallography. Data are now routinely collected from macromolecular crystals at cryogenic temperatures. The crystals are prepared and rapidly flash cooled to prevent ice lattice formation in the aqueous solvent in and on the surface of the crystal. Instead of ice, a glass forms, encasing the crystal with little or no damage. Crystals are then maintained at a temperature of between 100-120K during data collection. This has given crystals 'immortality' in the X-ray beam and enabled better quality data with the

capability for long-term storage and reuse of crystals and often extending the limiting resolution obtainable (Gamblin and Rodgers, 1993; Garman and Schneider, 1997; Rodgers, 1994; Watenpaugh, 1991).

Intrinsic limitations due to diffraction quality, can be dramatically improved by cryotechniques. Many factors contribute to improvements in data quality, including reduced thermal vibrations, decreased background scatter and absorption due to absence of glass capillary and excessive solvent during crystal mounting, enhancing the signal-to-noise ratio, reduced conformational disorder and potentially higher limiting resolution. Another important effect is the suppression of radiation damage, permitting complete data to be collected from a single crystal which in turn eliminates errors from merging and scaling of multiple data subsets from several crystals. In combination these improvements lead to enhanced contrast and sharper detail in electron density maps.

Data collected at cryogenic temperatures are superior to data recorded at room temperature in many respects and facilitate smoother progress in structure solution and refinement. However, the use of cryocooling and cryoprotectants create an environment very different from that of its normal habitat and may induce some changes that can lead to complications later in structure solution and refinement (Garman and Schneider, 1997). The work described by Gregoriou (Gregor*iou et* al., 1998) showed that the cryoprotectant glycerol was found to be a competitive substrate in inhibitor-binding studies on glycogen phosphorylase *b*. These must be taken into account when results obtained at cryogenic temperatures are interpreted in terms of structure and function at ambient temperatures. Some changes are induced by the binding of cryoprotectant molecules, especially when binding takes place in regions of particular interest. It should also be noted that, when data to only relatively low resolution are available, cryoprotectant molecules might not be identified as such in electron density maps and thus might give rise to misinterpretation of the protein structure.

The introduction by Teng (Teng, 1990) of a loop mounting technique for flash cooling of crystals was a major advance for crystallography. Loops are made of fine fibres, e.g. rayon, in which the crystal is scooped from the cryosolution and held within the loop suspended by a thin film of the solution. The loop is supported by a pin, which is itself

attached to a steel base used for placing the assembly on a magnetic cap on the goniometer (Figure 2.7).

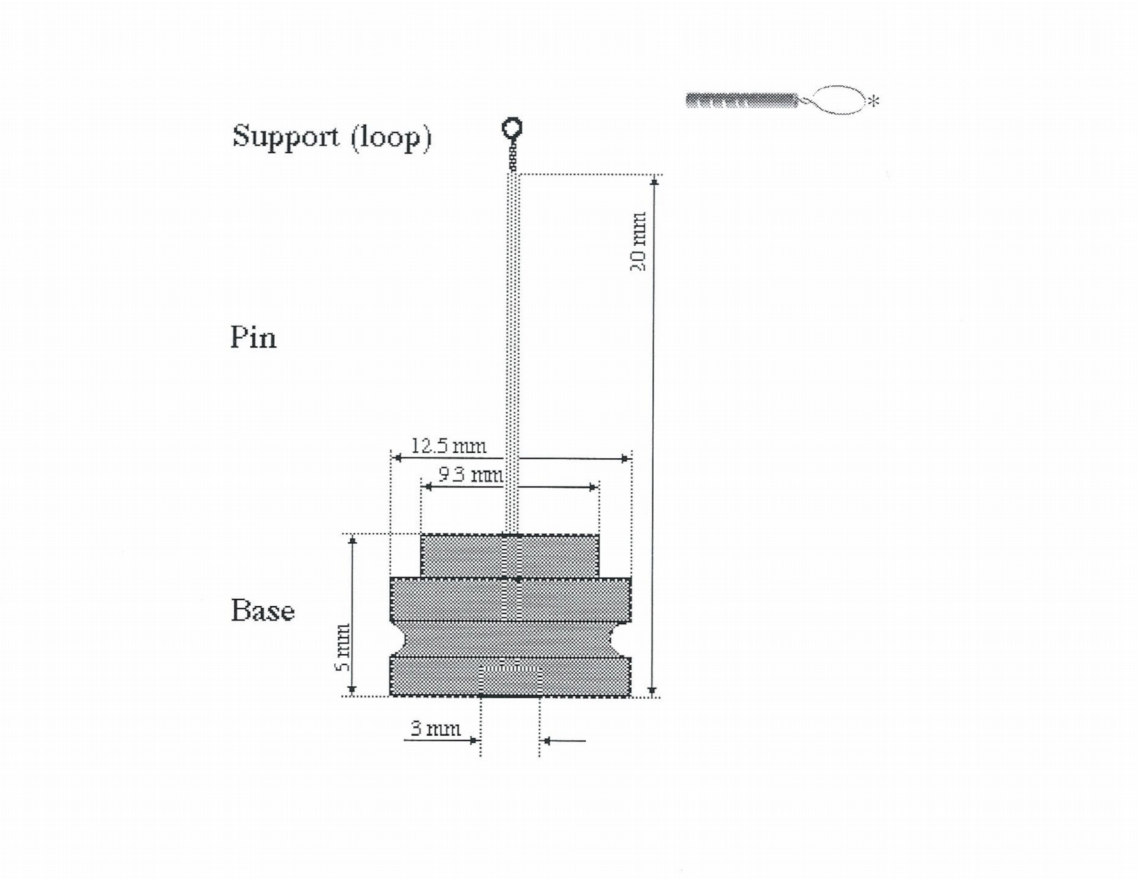


Figure 2.7 Loop assembly for cryogenic cooling. *The loop assembly is attached to a small pin inserted into a metal base which can easily transferred to the magnetic mount on the goniometer head. *The orientation of the loop relative to the spindle to ensure the crystal is mounted with b along the spindle axis.*

2.8.4 Data collection

The best strategy during X-ray diffraction data collection depends on qualitative factors, such as crystal quality and availability, type of X-ray source and detector, and time available, and quantitative ones, such as cell parameters, resolution limit, and crystal symmetry. Data were collected using the rotation method employing a ϕ range of crystal rotation based on resolution, crystal orientation and mosaicity and processed using the *HKL* suite (Otwinowski and Minor, 1997). The analysis and reduction of the diffraction data consists of the following steps:

1) indexing of the diffraction pattern and determination of the crystal orientation
2) refinement of the crystal and detector parameters
3) integration of the diffraction intensities
4) refining the relative scale factors between equivalent measurements
5) precise refinement of crystal parameters using all the data
6) merging and statistical analysis of the measurements related by space group symmetry

Steps 1 to 3 are carried out by the program *DENZO*, while steps 4 to 6 are performed by the program, *SCALEPACK*.

One of the least well-defined criterion in data collection is the resolution limit of diffraction (Dauter, 1997). A useful guide is to restrict the resolution to the point below which more than half of the intensities are higher than 2σ. This assumes the errors (σ) of the measured intensities are measured correctly. σ's can be modified during scaling and merging from the distribution of intensities of equivalent reflections. The R_{merge} (also called R_{sym}) is the most widespread statistic used to indicate data quality for macromolecular crystallography and is defined as

$$R_{merge} = 100 \sum_{hkl} |I_{hkl} - \langle I \rangle| / \sum_{hkl} I_{hkl} \qquad [1]$$

R_{merge} is commonly used to guide decisions during data reduction, such as determining to what resolution data are reliable. Diederichs and Karplus (Diederichs and Karplus, 1997) have shown that R_{merge} is seriously flawed, because it has an implicit dependence on the redundancy of the data. The R_{merge} is less appropriate than I/σ as a criterion to judge the resolution limit because of its dependency on the multiplicity of the data and

on the symmetry of the crystal. A decision on where to apply a cut-off varies for each individual case, but as a general rule of thumb a value up to 30% is generally accepted as a cut off for good data. Once criteria have been set for what defines an acceptable standard for data quality, then time restraints must also be considered. This is important for data collection in general, and especially so at a synchrotron where time is often a limiting factor.

The desired crystal rotation per image is an important consideration during data collection. To prevent overlaps from successive zones, the crystal rotation per image, $\Delta\phi$, depends on the resolution limit and cell dimension along the beam direction and can be estimated from:

$$\Delta\phi = (180/\pi)(d/a) \qquad [2]$$

where d is the resolution limit and a is the cell dimension along the beam in the given crystal orientation during exposure. The calculated rotation range should be reduced to allow for the crystal mosaicity and beam divergence. The cell dimension along the spindle axis will never come into play, since the crystal is rotated around this axis and will never lie along the beam assuming the crystal is perfectly orientated with the beam.

2.8.5 Phase Problem: Molecular Replacement

Having accumulated X-ray data which contain information on the intensity of the diffraction but not on the relative phase, we are then faced with obtaining phase information. Since a model was available in a different spacegroup, molecular replacement was used to obtain initial phases for the monoclinic data. The basic problem of molecular replacement is to determine the positions of the molecules within the crystal cell. This is achieved by generating model crystal structures with molecular models placed at all possible orientations and positions, and then selecting those that give an optimal agreement between the calculated structure factors and the observed ones using a program such as *AMoRe* (Navaza and Saludjian, 1997).

Each molecular model is treated as a rigid body: its position in the crystal cell, with respect to an initial reference one , is defined by three rotational and three translational parameters (Blow, 1985). The rotational parameters are first determined in terms of Eulerian angles from the rotation function; then, given the orientation, the translational

parameters are determined from the translation function. This approach is a practical method of reducing a six parameter search into two three parameter searches which is computationally less demanding.

The derivative model determined by multiple isomorphous replacement (MIR) at 1.9 Å (Duggl*eby et* al., 1995), provided a good search model for molecular replacement avoiding the phase problem. The molecular replacement solution was determined using the program *AMoRe* (Navaza and Saludjian, 1997). The model was positioned using *LSQKAB (Collaborative Computational Project, 1994)*.

2.8.6 Refinement

Refinement is the process of adjusting the parameters of the model to find a closer agreement between the calculated and observed structure factors. The adjustment of the model consists of changing the positional parameters and the temperature factors of the atoms whilst maintaining sensible stereochemical restraints in the structure. The numbers of atoms in proteins can be very large and hence also the number of parameters to be refined. For PGA with ~7000 atoms, the refinement of three positional parameters (x, y, and z) and one isotropic temperature factor parameter (B) makes a total of 28,000 parameters. Additional observations are incorporated into the refinement in the form of stereochemical data obtained from small molecular structures of amino acids and peptides in which the bond lengths and angles have been determined with high precision (Engh and Huber, 1991). The stereochemical information can be applied in two different ways. For constrained refinement the geometry is considered as rigid which reduces the number of parameters. In contrast, when the stereochemical parameters are allowed to vary around a standard value, the refinement is restrained in which appropriate weights are applied between geometric and X-ray terms. Restraints on bond lengths, bond angles, torsion angles, and van der Waals contacts add to the number of observations.

2.8.6.1 REFMAC: Maximum Likelihood Refinement

Refinement was carried out using restrained maximum likelihood refinement implemented in the program *REFMAC* (Murshu*dov et* al., 1997) using the Engh and Huber dictionary for restraints (Engh and Huber, 1991). Each cycle of the program performs two steps:

- Estimates the overall parameters of likelihood (σ_A's) using the free set of reflections
- Uses these parameters to build the likelihood function and refine the atomic parameters (working reflections are used)

Maximum likelihood refinement is performed by calculating the first derivative and making an approximation of the second derivative of the likelihood function with respect to parameters that can be refined and then estimates the shifts to be added to the parameters.

The values of sigma are given for the geometric restraints used during refinement. These values are the input estimated standard deviations that determine the relative weights of the corresponding restraints. The restraints used during refinement by *REFMAC* include distance, bond angle, torsion angles, chiral volumes, van der Waals repulsion and B value.

Model rebuilding was carried out with the *X-FIT* module in *QUANTA* and manual addition of waters with *X-SOLVATE*.

2.8.7 Validation and Structure quality

2.8.7.1 Global indicators

The correctness and precision of the atomic parameters in a structure need to be assessed thoroughly, both during and after refinement. An initial indication of the reliability of a protein crystal structure is the resolution of the data against which the structure has been refined. The conventional *R*-factor (R_{cryst}) is an indicator of the general quality of a crystal structure. However it is not a good independent validator, since it can be manipulated by excluding or overfitting the data. A better assessment of the fit between observed and calculated structure factors is the free *R*-factor (Brunger, 1992). This is based on the general statistical principle of cross validation where the

model is required not only to reproduce those experimental data included in its estimation but also a set of excluded data. This is particularly valuable as it aids bias-free refinement by indicating overfitting. However, *R*-factors give only a global indication of structure quality and cannot point out local error.

2.8.7.2 Estimate of coordinate precision

Isotropic temperature factors, described by equation 1, are related to the thermal vibrations of an atom and therefore to some extent to the precision with which its atomic position can be measured. It has been demonstrated that atomic precision in proteins depends strongly on B values (Eqn [3]) (Dao*pin et* al., 1994),

$$B = 8\pi^2 \langle u^2 \rangle \qquad [3]$$

where u is the atomic displacement parameter.

Luzzati plots are widely but incorrectly used to assign an overall average error for atomic coordinates (Luzzati, 1952). However, recent literature (Dod*son et* al., 1996) suggests that the Luzzati plot, which was designed to give a measure of how far a refinement had to go to completion, is not an appropriate measure of coordinate error for use at the end of refinement. A formula has been suggested for useful comparison of the relative precision of different structures (Cruickshank, 1996), dependent on *R* factor, number of refined parameters, resolution and completeness of data, by comparing atoms with $B = B_W$ (Eqn [4]),

$$\sigma_w(x) = (N_i/p)^{1/2} C^{-1/3} d_{min} R \qquad [4]$$

where, N_i is the total number of atoms in a structure

p is n_{obs}-n_{params}

C is the fractional completeness of the data to d_{min}

d_{min} is the resolution to which the data extend

R is the final R-factor

The smaller d_{min} and the smaller R the higher the precision of the structure. It provides a rough guide for the diffraction-data-only error contribution for an atom with temperature factor equal to the Wilson B (B_W) for the structure. It is known as the 'Diffraction-data Precision Indicator' (DPI) and contains none of the restraint data.

Cruickshank stresses that the DPI is only a rough formula for diffraction-data contribution to coordinate precision but shows that it gives useful order of magnitude results for the global average precision of structures (Cruickshank, 1996).

The program *PROCHECK* (Laskow*ski et* al., 1993) was used to examine the stereochemistry of the final models. Hydrogen bonds were assigned according to the criteria for polar groups: distance cut off of 3.3 Å, angles O • • • H-N > 120° and source atom • • •O-C > 90° for other hydrogen bonds.

Chapter Three A New Crystal form

3.1 Introduction

The benefits of high resolution data include greater precision of the structure in which one can hope to identify fully resolved atoms and in particular give a clearer definition of the active site of enzymes. A high resolution structure of the active site of PGA is important since a primary objective of this study is the substrate specificity of the enzyme. High resolution would enable clear distinctions to be made between conformations of active site residues in wild type PGA and with substrates in complex both with mutants and the wild type enzyme.

This chapter discusses the experiments used to obtain a very high resolution structure of PGA. The requirement for high resolution, and the ensuing lengthy data collection time, means that the data collection strategy used would require cryogenic temperatures. Previously, X-ray data for PGA has been collected at room temperature (Don*e et al*., 1998; Duggleb*y et al*., 1995), where data from several crystals was merged together, however the final data was still incomplete. As mentioned in the previous chapter, low temperature data collection has obvious benefits and often leads to an improvement in the resolution limits of the data sometimes by 0.5 Å or more (Daute*r et al*., 1995). A high resolution structure of PGA could also clear up uncertainties in the model of PGA, in particular poorly ordered loop regions and in general provide a more precise description of the native enzyme. Comparison between the resulting low temperature structure and the room temperature structure of PGA (Duggleb*y et al*., 1995) highlights the benefits obtained from data collection at cryogenic temperatures. A search for suitable crystallisation conditions to enable cryogenic freezing of the crystals was undertaken.

3.2 Experimental and computational procedures

3.2.1 Crystallisation

3.2.1.1 The triclinic crystal form of PGA

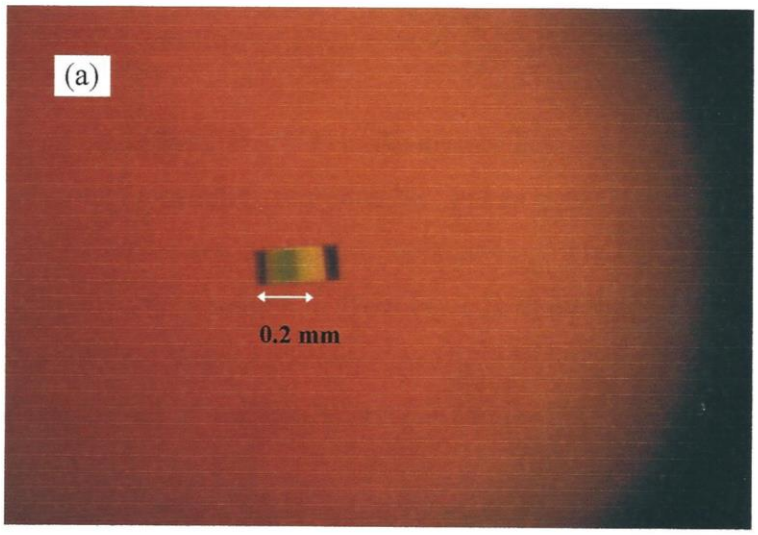
The original PGA crystals were grown by batch methods which proved unreliable and required large amounts of protein (Hun*t et al*., 1990), as other methods of crystallisation such as hanging drop and microdialysis were reported to be unsuccessful. Crystals were grown from protein (12 mg/ml) equilibrated in 50 mM MOPS pH 7.2 with 12–15% polyethylene glycol 8000 (PEG 8K) as precipitant. These crystals appeared within two to three weeks. A more reliable supply of crystals was obtained using macroseeding. The triclinic batch crystals had cell dimensions a = 52.12 Å, b = 65.08 Å, c = 76.30 Å, α = 100.20°, β = 95.58° and γ = 105.81 and solvent content 55%.

3.2.1.2 An early monoclinic crystal form of PGA

Applying novel precipitants for crystallisation (Brzozowski and Tolley, 1994) a new crystal form was characterised. Again using the batch method, crystals with a variety of morphologies were obtained using 12–14% polyethylene glycol 2000 monomethyl ether (MME PEG 2K) as precipitant (S. Tolley, personal communication). One of the best crystals allowed collection of X-ray data to 2.5 Å resolution at room temperature. The space group was $P2_1$ with unit cell dimensions a = 117.95 Å, b = 103.46 Å, c = 139.13 Å and β = 95.58° with an overall R_{merge} of 11% and the data were 96% complete (Done, 1996; McVe*y et al*., 1997). There were four molecules in the asymmetric unit. Further work on this crystal form was abandoned due to the irreproducibility of the crystallisation method and the poor diffraction. However, the results prompted further investigation of crystal growth conditions and the development of alternate protocols in the hope of producing crystals with better diffraction properties.

3.2.1.3 A new monoclinic crystal form

An initial objective of the research in this thesis was to develop a suitable crystallisation procedure to produce crystals amenable to flash cooling for cryogenic data collection. The approach used involved an additive to the crystallisation buffer and provided a novel crystal form of PGA using the precipitant MME PEG 2K and ethylene glycol (McVe*y et al*., 1997), exploiting a combination of hanging drop vapour diffusion and streak seeding. 4 µl hanging drops [2 µl protein (10 mg/ml in 50 mM MOPS pH 7.2) + 2 µl reservoir (50 mM MOPS pH 7.2, 10–12% MME PEG 2K, and 20–25% ethylene glycol)] were equilibrated at 18 °C for 2 days. Initially using a stock of batch crystals provided by Dr. S. Tolley, 3-6 consecutive hanging drops were streak seeded using a mounted cat's whisker, the progressive seeding giving a smaller number of seeds in the later drops. This produced single crystals within 2-3 days with dimensions about 0.2 x 0.2 x 0.1 mm^3 (Figure 3.1a) in the monoclinic space group $P2_1$ with one molecule in the asymmetric unit. The aim of adding the cryoprotectant, ethylene glycol, to the crystallisation buffer was to provide crystals in mother liquor that was mildly or fully cryoprotecting, minimising the need for a potentially damaging transfer to harvest buffer and serial transfer to cryoprotectant buffer (Gamblin and Rodgers, 1993). This approach was successfully applied later to crystallisation of PGA mutants by cross seeding with wild type crystals.



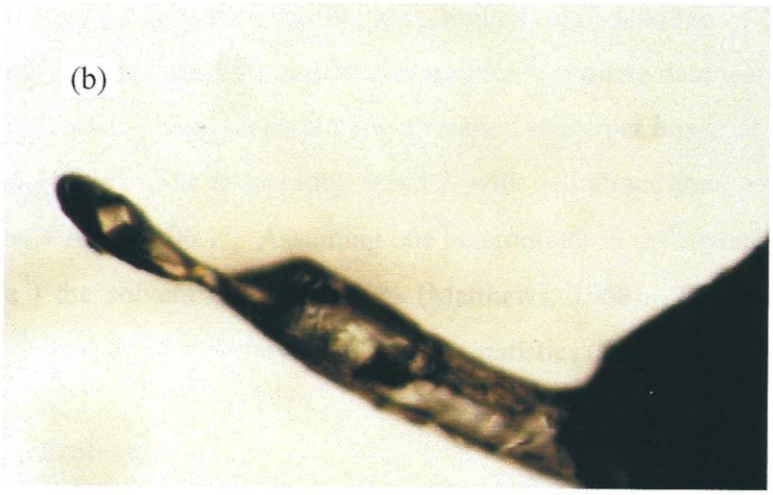


Figure 3.1 (a) A typical monoclinic crystal of PGA, (b) crystal mounted in loop.

3.2.2 Data collection

3.2.2.1 In-house data

The ethylene glycol concentration (25%) in the mother liquor proved insufficient to form a protective glass upon cryocooling reproducibly, and so crystals were transferred briefly to a higher concentration of cryoprotectant (30% ethylene glycol, 12% MME PEG 2K, in 50 mM MOPS pH 7.2). Attempts to grow crystals in mother liquor containing 30% ethylene glycol produced crystals of lower quality. Crystals were mounted in rayon loops (Hampton Research) of similar dimensions to the crystal, and quickly transferred to the liquid N_2 stream.

X-ray diffraction data were collected at 120 K with Cu Kα radiation (λ = 1.5418 Å) using a 30 cm MAR research image plate mounted on a Rigaku rotating anode generator operated at 50 kV and 100 mA. The mosaic spread was 0.42° and whilst the crystal diffracted to at least 1.9 Å, measurements were limited to a resolution of 2.73 Å owing to the large unit-cell dimensions and time constraints. Complete data were collected to 2.73 Å using the rotation method, employing a rotation range per image of 1° and a total crystal rotation of 126°. The spacegroup was $P2_1$ with cell dimensions a = 51.5 Å, b = 131.9 Å, c = 64.4 Å, β = 106.1°. Assuming one heterodimer in the asymmetric unit (V_m = 2.45 $Å^3$ Da^{-1}) the solvent content is 49% (Matthews, 1968). Data reduction with *ROTAVATA* and *AGROVATA* gave good quality statistics (Table 3.1).

3.2.2.2 Synchrotron data

The quality of the in-house diffraction data suggested that monoclinic crystals would diffract to higher resolution at a synchrotron source. Therefore, the in-house data were not used for structure refinement. The requirements for collecting high resolution data for PGA involved optimising the crystal orientation and limiting the mosaicity of the crystal. For PGA, where one cell dimension is much longer that the two others, it is favourable to have it aligned close to the spindle axis. Crystals often grow slowest along the direction of the longest axis and usually it is their thinnest crystal dimension. The crystal orientation was an important deciding factor for

efficient data collection and it was necessary to devise a set-up that would be conducive to high resolution data collection with a suitable freezing strategy that would minimise the mosaic spread and damage to the crystal quality.

Data statistics for wild type PGA			
*EMBL, DESY	In-house	*BW7A	*BW7B
Hamburg beam line	λ = 1.54 Å	λ = 0.92 Å	λ = 0.89 Å
Space group	$P2_1$	$P2_1$	$P2_1$
Cell parameters (Å)	a = 51.5	a = 50.8	a = 51.3
	b = 131.9	b = 131.4	b = 131.6
	c = 64.4	c = 64.7	c = 63.9
	β =106.1°	β =105.8°	β =105.9°
Resolution (Å)	20 – 2.73	20 – 1.65	20 – 1.31
Mosaicity (°)	0.42	0.58	0.2
Temperature (K)	120	110	100
Reflections measured	60235	723673	855106
Unique reflections	23555	88311	190229
$\langle I \rangle / \langle \sigma(I) \rangle$	18.3 (14.3)[a]	13.1 (2.1)[b]	20.5 (3.4)[c]
Completeness (%)	99.1 (99.9)	91 (83)	95.5 (82)
†R_{merge}	3.0 (4.2)	7.6 (44.3)	7.5 (24.4)

[abc] Numbers in parentheses refer to the outermost resolution shell [a](2.73-2.66 Å) [b](1.67-1.65 Å) [c](1.31-1.30 Å).

†$R_{merge} = 100 \sum_{hkl} |I_{hkl} - \langle I \rangle| / \sum_{hkl} I_{hkl}$

Table 3.1 Data reduction statistics for wild type PGA.

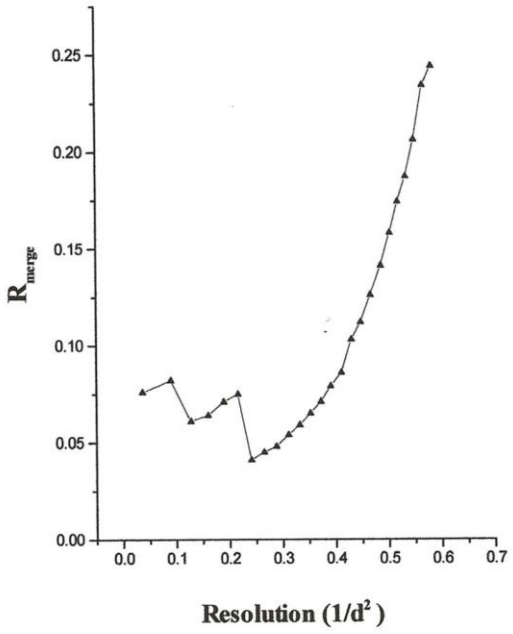
Initially data to a limiting resolution of 1.65 Å were collected at 110 K on beam line BW7A at the EMBL (DESY, Hamburg). The quality of the data was deemed to be only acceptable to 1.78 Å (R_{merge} = 28%; $\langle I / \sigma(I) \rangle$ = 3.2, for the resolution shell 1.81-1.78 Å). The data collected were disappointing, the transport of non-frozen crystals to Hamburg and optimisation of the freezing protocol were probable contributing factors to the lower than expected resolution limit.

Subsequently, fresh crystals were grown at the EMBL outstation and data were collected to 1.3 Å at 100K on beam line BW7B using three resolution sweeps, Table 3.2. The synchrotron data were collected on a MAR research image plate (30 cm). Prior to crystal mounting and freezing the loop was bent to achieve the desired orientation (Figure 2.7 and 3.1b), to ensure that the crystal orientation was such that b* (longest cell dimension) was along the spindle axis to reduce the number of overlaps between lunes in the diffraction image. Data were integrated and scaled using the programs from the *HKL* suite (Otwinowski and Minor, 1997), Table 3.1. The higher R_{merge} in the low resolution shells 20–2.11 Å of both synchrotron datasets is indicative of the weak reflections from the low resolution passes (Figure 3.1) which have larger statistical errors of observation.

	High resolution	Medium resolution	Low resolution
Crystal-to-detector distance (mm)	180	200	320
Resolution range (Å)	4 – 1.3	13 – 2.1	20 – 3.2
Oscillation per image (°)	1	1.5	2.5
No. of Images	127	121	71
Crystal mount along spindle		b*	
Mosaicity (°)		0.2	

Table 3.2 Data collection parameters summary for 1.3 Å PGA data.

(a)



(b)

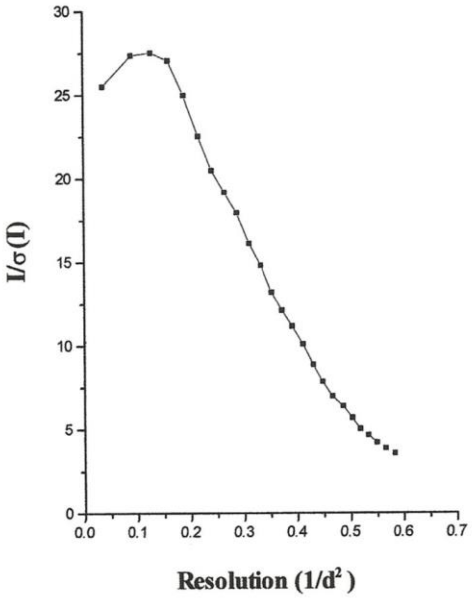


Figure 3.2 ***(a) The merging R(I) factor, defined as*** $\sum|I-\langle I\rangle|/\sum I|$ ***and (b) the average*** $\langle I/\sigma(\mathrm{I})\rangle$ ***of the data as a function of resolution for the 1.3 Å data.***

3.2.3 Molecular replacement

The structure of the monoclinic crystal form using the wild type 1.65 Å data was solved by molecular replacement with *AMoRe* (Navaza and Saludjian, 1997). The search model was the original wild type triclinic '1.9 Å' penicillin acylase model (Duggleby *et al*., 1995); PDB accession code 1PNK, with water molecules removed. The search model was placed in a unit cell with dimensions $a = b = c = 87$ Å, $\alpha = \beta = \gamma = 90°$ and an integration sphere of 37 Å radius and increments between calculations of 2.5° in each Eulerian angle were applied. Data between 10 and 4 Å were included for both the rotation and translation function and between 15 and 3 Å for rigid-body refinement. The results of the molecular replacement gave a single unambiguous solution with a peak height 27 σ, with the next peak only 4.5 σ above the mean. The best solution gave an R-factor of 35.9%, and a correlation coefficient of 62.5%.

The rotation function map shows a clear solution peak (Figure 3.3), and demonstrates the straightforward molecular replacement solution (Table 3.3). The translation function solutions are equally clear with a high correlation factor (Table 3.4). The clear rotation and translation function solutions indicates that the conformation of PGA is the same for both the triclinic and monoclinic crystal forms.

Solution	α	β	γ	TX	TY	TZ	CorrF	RFac
1	182.23	74.71	38.36	–	–	–	35.8	0.0

Table 3.3 The best solution from the rotation function.

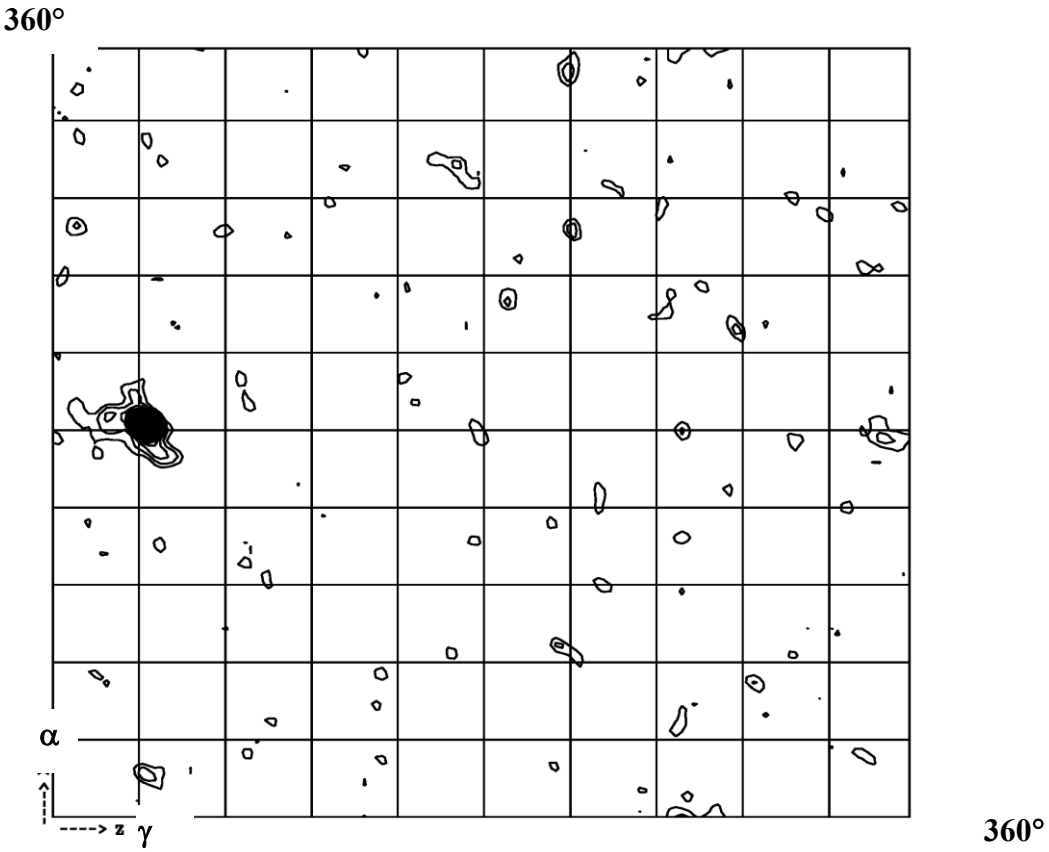


Figure 3.3 A plot of the rotation function from AMoRe showing the solution peak. *The section corresponds to a β angle of 75°. The map is contoured from 2 σ through to the peak height solution of 27 σ in steps of 1 σ.*

Solution	α	β	γ	TX	TY	TZ	CorrF	Rfac
1	182.23	74.71	38.36	0.3036	0.0	0.0588	62.5	35.9

Table 3.4 The solutions from the translation function.

The positional parameters were refined by 10 iterative cycles of rigid body refinement using the *FITING* function of the *AMoRe* program. The R-factor fell to 33% and the correlation coefficient increased to 76%, Table 3.5.

Solution	α	β	γ	TX	TY	TZ	CorrF	Rfac
1	182.33	74.65	38.19	0.3061	0.0	0.0631	76.2	33.0

Table 3.5 The refined solution from rigid-body fitting.

The refined solution was used to reorientate the search model with *LSQKAB* (Collaborative Computational Project, 1994).

3.2.4 Refinement

The reorientated model was refined using maximum likelihood as implemented in *REFMAC* (Murshudov *et al*., 1997) with a bulk solvent correction. 3% of the data were flagged for cross validation. As the monoclinic crystal form exhibited anisotropic mosaicity, overall anisotropic scaling was applied which typically resulted in a 1% drop in free R-factor during refinement. The contribution of hydrogen atoms was included in the model once refinement proceeded below an R-factor of 20% using the program *HGEN* (Collaborative Computational Project, 1994). During

manual rebuilding, difference electron density maps with *SIGMAA* map coefficients $2m|F_o| - D|F_c|$ and $m|F_o| - D|F_c|$ were used to guide the fitting of the model. All model fitting was carried out using the *X-AUTOFIT* module within *QUANTA* (Molecular Simulations Inc).

The refinement of high resolution structures necessitates, especially with large proteins, an automatic procedure for the addition of water molecules. The automated refinement procedure (*ARP*) applies a semi-objective set of criteria for water selection, on the basis of distance, electron density and sphericity (Lamzin and Wilson, 1993). *ARP* was used in combination with *REFMAC* until approximately 80% of water molecules were added. Water molecules were added manually in the later stages using the *X-SOLVATE* module in *QUANTA*. Waters were modelled into difference density based on the criteria that peaks were greater than 3.5 σ and upon inspection of the model made at least one hydrogen bond to protein or to a neighbouring water which itself made hydrogen bonds to protein.

Refinement statistics	***frz*PGA-1.8**	***frz*PGA-1.3**
Resolution (Å)	20 – 1.78	20 – 1.31
†R (‡R_{free}) factor (%)	17.8 (23.0)	14.8 (16.9)
Non-hydrogen atoms		
¶protein	6111	6501
water	752	923
¶ethylene glycol ()	-	108 (24)
R.m.s deviations (Å)		
*bond 1-2 distance (0.02)	0.016	0.014
*angle 1-3 distance (0.04)	0.033	0.027
*planar 1-4 distance (0.05)	0.037	0.032
Average B factor ($Å^2$)		
all atoms	16.6	15.5
protein	15.8	13.7
solvent	25.0	27.9
Ca^{2+} ion	9.8	7.9
ethylene glycol	-	18.6
§Coordinate error (Å)	0.115	0.048

*The numbers in parentheses refer to the target weights used in the restraints.

†R-factor = $\sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl} \left| F_{obs} \right|$

‡Free R-factor = $\sum_{hkl \subset T} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl \subset T} \left| F_{obs} \right|$ where $hkl \subset T$ represents 3% of the diffraction data. §Coordinate error estimated by DPI, the Diffraction-data Precision indicator (Cruickshank, 1996). ¶The number of atoms vary dependent on the number of alternate conformations. #Number of ethylene glycol molecules

Table 3.6 Refinement statistics for wild type PGA.

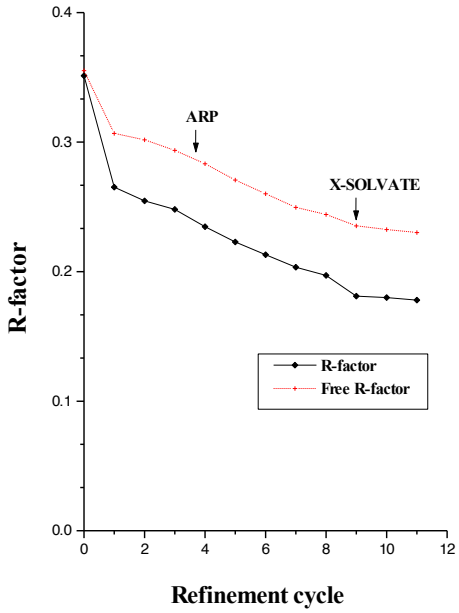


Figure 3.4 Refinement of 1.8 Å data showing the fall in R/Free R factor with number of cycles.

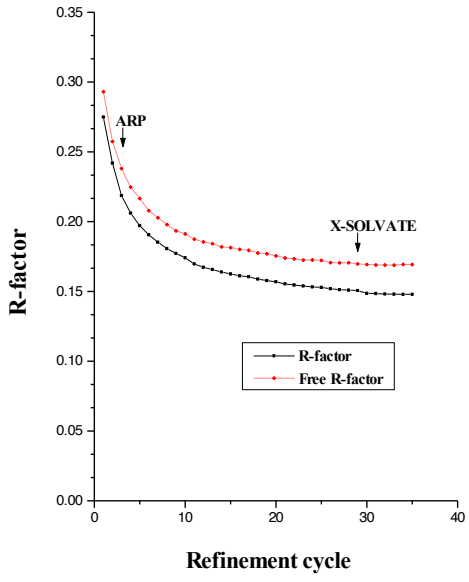


Figure 3.5 Refinement of 1.3 Å data showing the fall in R / Free R factor with number of cycles.

3.3 Results and Discussion

3.3.1 Description and quality of the native 100 K PGA structure

The structure of wild type PGA (*frz*PGA-1.3) was determined at cryogenic temperatures (100 K) to 1.3 Å resolution in space group $P2_1$. The presence of ethylene glycol in the crystallisation buffer permitted flash freezing of the crystal with a minimal addition of cryoprotectant. The monoclinic crystal forms diffracts considerably better than the previously reported triclinic crystal form (Duggleby *et al.*, 1995). This can be attributed, firstly, to the use of synchrotron radiation as an X-ray source, and secondly, the use of cryogenic freezing during data collection. This description of the structure of PGA will focus on the low temperature 1.3 Å structure and will only digress to the triclinic room temperature structure when appropriate.

In the *frz*PGA-1.3 structure, 91.3% of residues are in the most favoured region and 8.7% are in allowed regions of the Ramachandran plot and no residues in the unfavourable region (Figure 3.6).

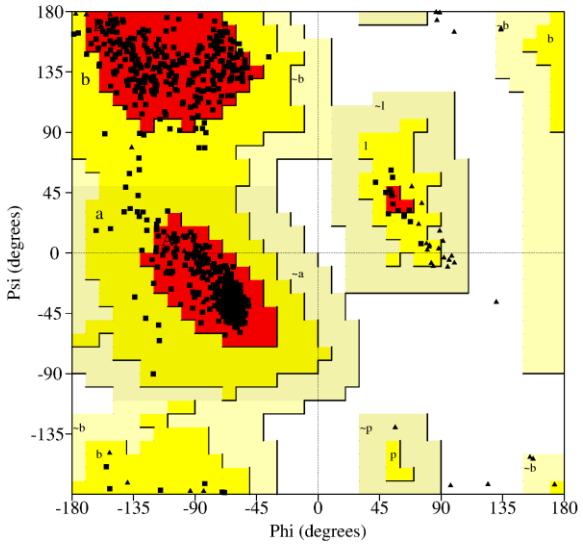


Figure 3.6 Ramachandran plot for* frz*PGA-1.3. *Generated from PROCHECK (Laskowski et al., 1993).*

The secondary structure of the *frz*PGA-1.3 structure is illustrated in Figure 3.7 showing the active site serine, the calcium binding site and the remaining C-terminal residues of the A chain.

Figure 3.7 Secondary structure of PGA. *Showing the A chain (green) and B chain (blue) with the calcium ion (yellow), active site serine (ball and stick), C-terminus of A chain (red) and the active site helix H1.*

3.3.2 Re-refinement of RT-PGA structure with *REFMAC*

Before comparing the *frz*PGA-1.3 structure with the room temperature PGA structure (*RT*PGA-1.9), the *RT*PGA-1.9 structure was re-refined with the same refinement protocol as applied to the *frz*PGA-1.3 structure. The room temperature data for PGA (RT-PGA) which led to the published 1.9 Å structure had been originally refined by least-squares minimisation with *PROLSQ* (Hendrickson and Konnert, 1980) to an R-factor of 19.0% and free R-factor of 24.9 % between a resolution range of 10 – 1.9 Å. The original data extended to 1.8 Å resolution and these additional data were included in the re-refinement of the structure using *REFMAC*. The triclinic model was refined from 30 – 1.8 Å to a final R-factor of 16.2% and free R-factor of 20.8%. Having repeated the same refinement procedures, allowed a more valid comparison between the room temperature and frozen structures.

The low resolution cut-off contributes to the higher free R-factor in the reported structure. Extending the data to include all the high resolution data also contributes to the improved R-factors in the re-refined structure. This has been possible due to the improvements in refinement protocols including maximum likelihood, overall anisotropic scaling and bulk solvent correction.

3.3.3 Comparison of 100 K PGA with room temperature PGA

The re-refined room temperature native triclinic PGA (*RT*PGA-1.8) was compared with *frz*PGA-1.3. The two models were superimposed on their CA atoms using *QUANTA* giving an overall r.m.s deviation of all equivalent atoms of 0.804 Å for all equivalent atoms (Figure 3.8), 0.4 Å for CA atoms and 1.0 Å for side chain and side chain atoms (Figure 3.9). Thus there were no major changes in conformation between the two models, and the active sites were closely similar. There were differences in terms of overall B value, the number of alternate conformations, the number of water molecules and in the accuracy.

The differing characteristics of the *frz*PGA-1.3 structure can be attributed to a number of factors. The change in space group is reflected in increased and different crystal lattice contacts in the monoclinic form. A comparison of the crystal packing in both the monoclinic and triclinic space groups is illustrated in Figure 3.10. The solvent

content of the monoclinic crystal is 49% reduced by approximately 5% compared to the triclinic form. This is expected to increase the order of the monoclinic crystals, and in particular to influence those residues involved in crystal contacts.

The use of ETG as cryoprotectant leads to a set of ordered ETG molecules in the crystals, which may influence both the intermolecular protein contacts and the preferred conformation and flexibility of the protein itself.

The lower temperature and tighter packing of the monoclinic PGA structure leads to lower atomic displacement factors than observed at room temperature (Table 3.8). However this can be attributed to both reduced mobility at low temperature and to more extensive crystal lattice contacts mentioned above (Figure 3.11, 3.12).

There are local differences in conformation between the two structures. For example, two regions of main chain, B: 210-214 an exposed surface loop and B:369-378 close to the active site, show substantial disorder at room temperature. At 100 K, the loop B:210-214 improves only slightly and remains poorly defined. For the second region, B:369-378, two alternate conformation were modelled in the *frz*PGA-1.3 structure. In addition, B:494-495 is involved in symmetry contacts and two side chain conformations with 50% occupancy were modelled for Arg B:495, one involved in an internal salt bridge and the other involved in symmetry contacts.

Re-refinement statistics for *RT*PGA-1.8	
Resolution (Å)	30 – 1.8 Å
Rfactor (%)	16.2
Free Rfactor	20.8
Average B factor ($Å^2$)	
all atoms	28.6
protein	27.9
solvent	38.9
Ca^{2+} ion	18.1
Coordinate error (Å)	0.129

Table 3.7 Summary of refinement statistics for* RT*PGA-1.8.

B values ($Å^2$)	*frz*PGA-1.3	*RT*PGA-1.8
Overall	13.7	27.9
Main chain	12.0	25.4
Side chain	15.3	30.3

Table 3.8 A comparison of temperature factors for low temperature PGA (*frz*PGA-1.3) and room temperature PGA (*RT*PGA-1.8).

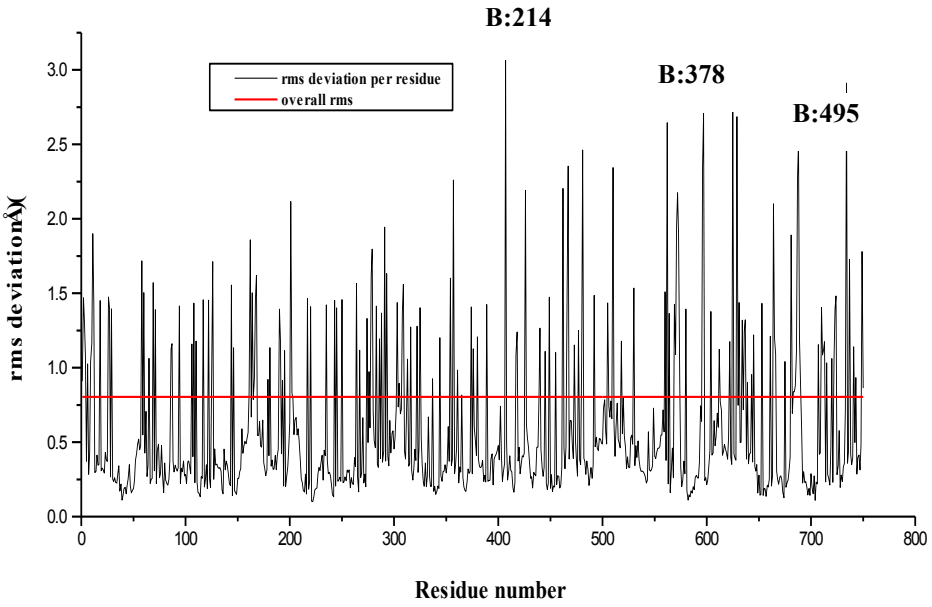


Figure 3.8 A plot of the overall r.m.s deviation per residue between the coordinates of the 1.3 Å frozen structure and the 1.8 Å room temperature structure versus residue number.

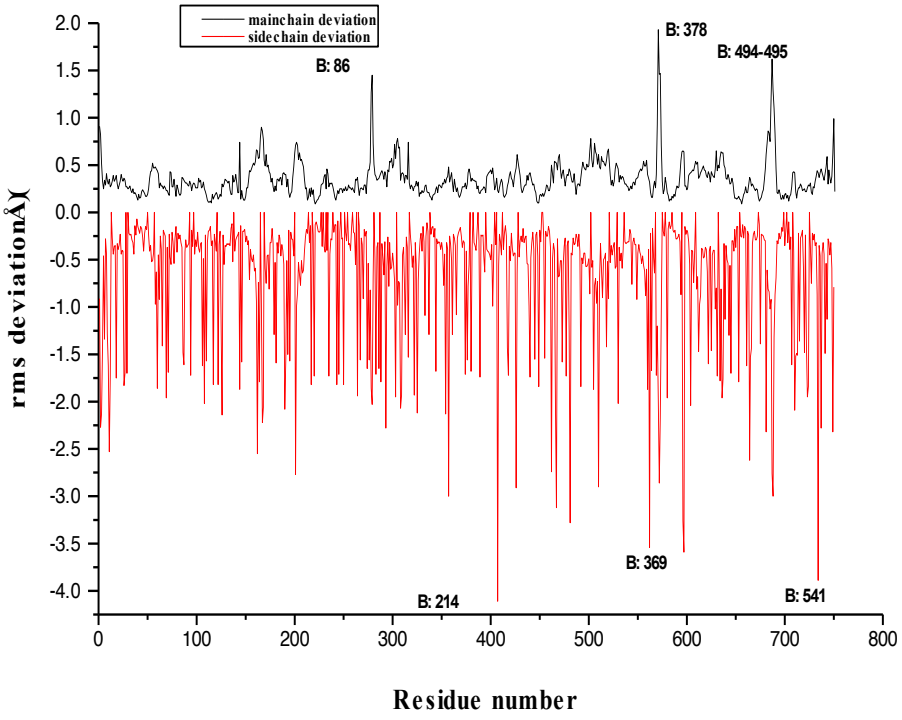


Figure 3.9 A plot of the rms deviation between the overlapped 1.3 Å (100 K PGA) structure and the 1.8 Å room temperature structure for main chain and side chain atoms per residue versus residue number.

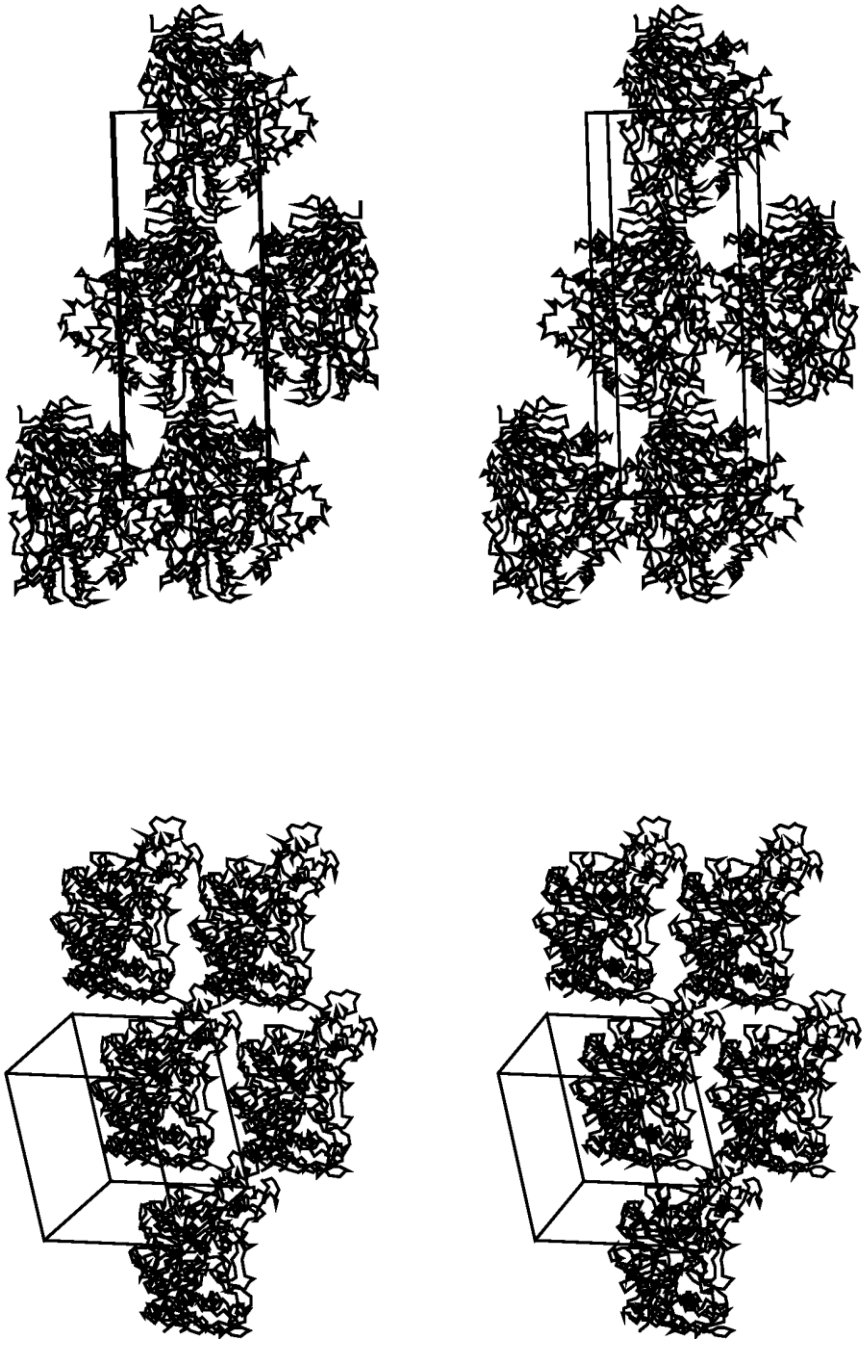


Figure 3.10 Crystal packing within (a) the monoclinic cell of 100 K PGA, (b) the triclinic cell of RT-PGA, where the vertical axis is along b.

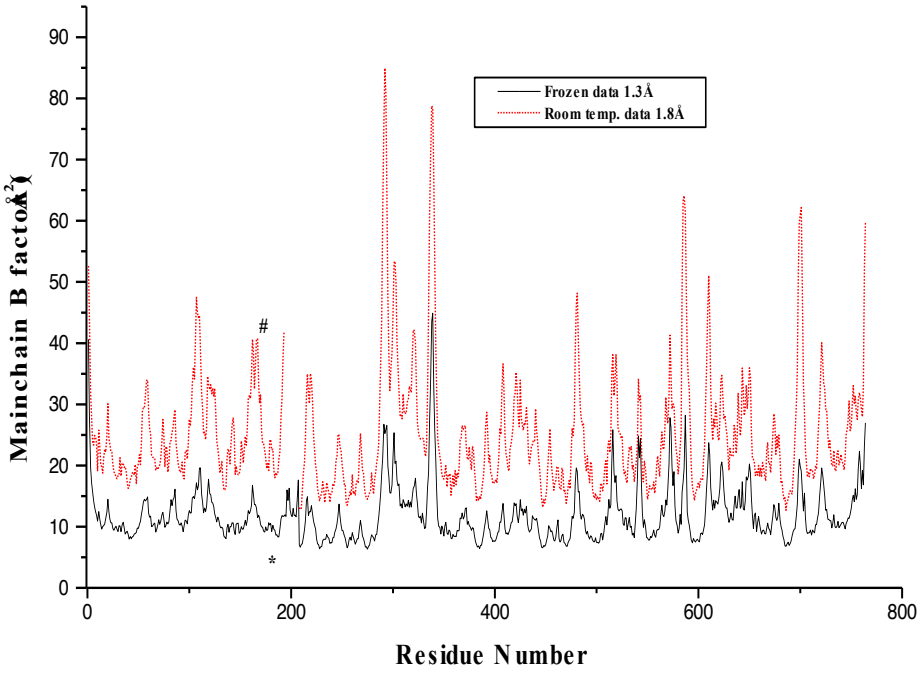


Figure 3.11 Plot of mean B factors for main chain atoms per residue versus residue number. *Comparison of 1.3 Å frozen data (solid line) with 1.8 Å room temperature data (dashed line). # denotes where the A chain ends at Ser A:195 in the room temperature 1.8 Å structure.* denotes Ser B:1.*

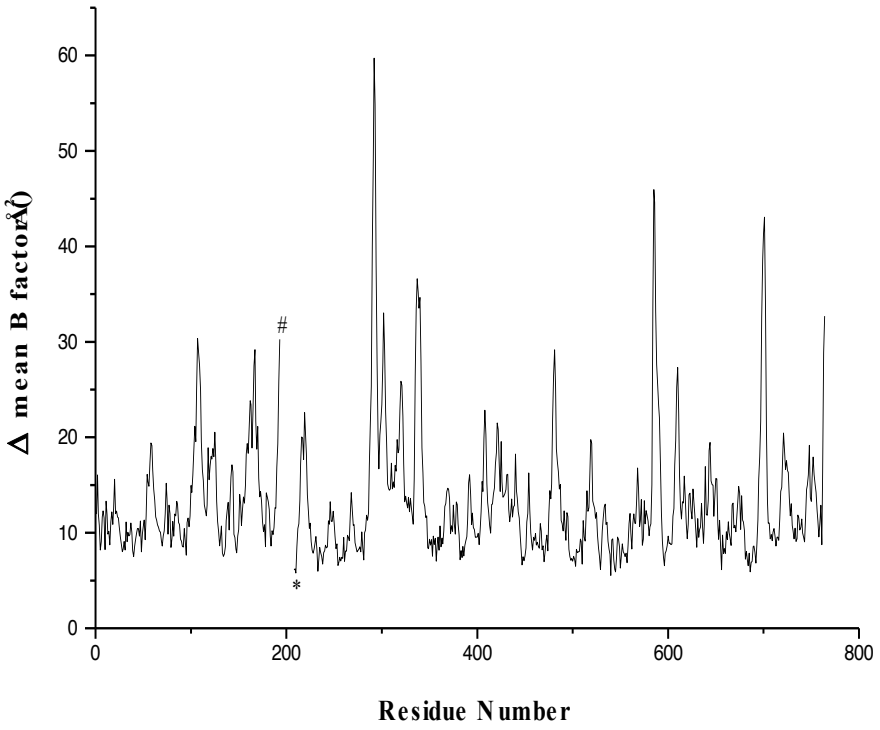


Figure 3.12 Plot of Δ mean B factor versus residue number between the 1.3 Å low temperature structure and the 1.8 Å room temperature structure. # Ser A:195, *Ser B:1

3.3.2.1 Alternate conformations of the active site helix

One advantages of 1.3 Å resolution is the ability to model multiple conformations. The PGA A chain shows two conformations of the main chain for residues A:139 to A:147 which form one face of the active site. In the various structures described in this thesis, the relative occupancy of the two conformations varies. In the room temperature and cryogenic wild type structures of native enzyme, the predominant conformation is that of a long α-helix involving residues A:131 through A:146, with a marked kink at Met A:142, positioned at the back of the specificity pocket which accommodates the phenylacetyl side chain. The ϕ/ψ angles of Met A:142 are unfavourable (-116°, -65°) for a helical conformation, and the long helix is better described as two helices with a kink between them.

The kink in the α-helix probably introduces some instability in the helical conformation, as is evidenced by the second conformation, which involves rearrangement of residues A:142 to A:147 into a coil rather than a helix. The second conformation is barely visible in the room temperature structure, perhaps due to the limited 1.8 Å resolution of the analysis. In the *frz*PGA-1.3 structure, it is visible, albeit at low occupancy (20%), which may partly reflect the presence of ethylene glycol in the specificity pocket. The significance of the two conformations is discussed further in Chapter 5.

3.3.2.2 C-terminus of the A-chain

Processing of the precursor occurs at Ala 235 (A:209)–Ala 236 and Thr 289–Ser 290 (Ser B:1) (O*h et al*., 1987). The presence of a heterogeneous species of A subunit when analysed by SDS-PAGE has been reported (Boc*k et al*., 1983b). In the 1.9 Å triclinic crystal structure, the C-terminal residues were modelled only as far as Ser A:195, and the 1.8 Å re-refined model also showed no electron density beyond this residue.

In contrast, the residues A:196 to A:209 are clearly defined in the electron density maps of *frz*PGA-1.3 crystal structure (Figure 3.13), which can be attributed to a combination of factors. Firstly, the use of recombinant protein and its method of preparation for the monoclinic analysis: the protein used to solve the triclinic

structure was obtained from Boehringer Mannheim as an ammonium sulphate cut and stored in this form before further purification and crystallisation. It is quite probable that residues A:196 to A:209 were cleaved prior to structure determination from the stored material obtained by conventional extraction methods. Secondly, the crystal contacts are different in the monoclinic crystal lattice, some involving residues A:204, A:205, A:208 and A:209 (Table 3.8). Two interpretations are possible here, either the presence of these residues in the recombinant material provides those contacts necessary for growth of monoclinic crystals or they are especially stabilised by the monoclinic crystal lattice. Thirdly, there is 30% ethylene glycol in the cryoprotectant solvent.

Crystal contacts of monoclinic PGA		
Residue : Atom	Symmetry residue : Atom	Distance (Å)
Gln A:204 : O	Arg B:495 : NH2	3.0
Asn A:205 : O	Arg B:495 : NH1	2.9
Thr A:208 : OG1	Asp B:494 : OD1	2.7
Ala A:209 : OCT	Ser B:493 : OG	2.9
Gln B:541 : NE2	Ser A:169 : O	2.9
Gln B:541 : OE1	Ser A:169 : OG	2.7
ETG C:21 : O6	B:494 : OD2	2.5

Table 3.9 The important crystal contacts in monoclinic PGA.

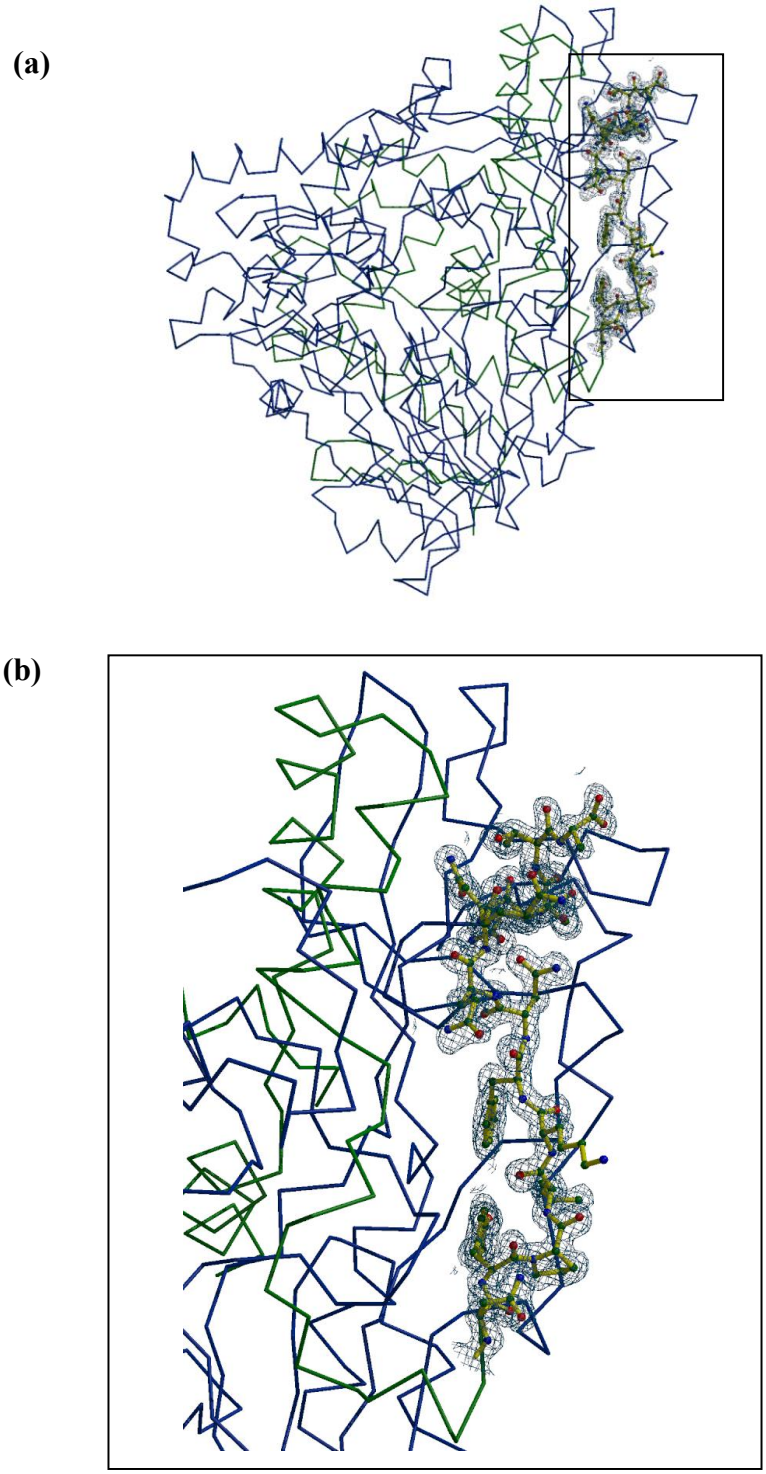


Figure 3.11 The C-alpha trace of PGA showing (a) The electron density for the additional A chain C-terminal residues present in the monoclinic structure and (b) a close up view of the C-terminus of the A chain with the $2F_o - F_c$ electron density map is contoured at 1σ showing the refined atomic positions of residues Ser A:195 to Ala A:209.

3.3.2.3 Calcium binding site

Metal ions serve a variety of functions in proteins. In metalloproteins they generally take on key mechanistic roles, for instance in ligand binding or redox transfer. In many other proteins, they play a structural role, enhancing the structural stability of the conformation of the protein that is required for biological function in the appropriate location in the cell or instead supporting a vital feature, such as the conformation of the active site. Examples of the latter include members of both families of serine proteinases, the trypsin-like and the subtilisin-like, which bind Ca^{2+} for stabilisation and maintenance of catalytic activity (McPhale*n et al*., 1991). Ca^{2+} binding often enhances protein stability primarily by promoting energetically favourable interactions within the protein structure, as the use of peptide carbonyl oxygen atoms as Ca^{2+} ion ligands reduces the flexibility of the polypeptide chain in specific regions.

There is a single calcium ion (Ca^{2+}) in both the triclinic (Duggleb*y et al*., 1995) and monoclinic crystal structures of PGA. Since no details of the calcium binding site were given in the published triclinic structure a brief description will be given in relation to the structure and function of PGA.

The Ca^{2+} ligand distances in room temperature and 100 K PGA are shown in Table 3.9. The calcium ion is at the base of the active site and has typical but distorted octahedral geometry (Kaufmann Katz *et al*., 1996). The calcium is coordinated by seven oxygen atoms at the vertices of an octahedral arrangement, each at a distance of ~2.4 Å, comprised of four side-chain residues, one a bidentate aspartate, (A:152, B:73, B:76 and B:252) and the main-chain carbonyls of Val B:75 and Pro B:205 (Figure 3.14). The binding site can be termed a semicontinuous site since four of the seven ligands are derived from a continuous local segment of the polypeptide chain, residues B:73-B:76. .

The calcium does indeed appear to play an essentially structural role. Firstly, it pins together the A and B chains and may serve to stabilise the protein in the periplasmic space. In the cytoplasm of the cell, the level of Ca^{2+} is normally several orders of magnitude lower than outside the cell with concentrations of calcium ions in the extracellular space, ~1.2 mM, and in the cytosol, ~0.1 μM (Kaufmann Katz *et al*.,

1996). PGA is a periplasmic enzyme and so Ca^{2+} ions may serve to initiate folding of the enzyme only when it has reached its correct cellular location.

Secondly, it is placed strategically at the lip of the active site interacting with Glu A:152, Asp B:73 and Val B:75 so that it helps configure the pocket with regard to substrate binding in which the nearby residues Phe A:146 and Phe B:71 both have a key role (see Chapter 5). Residue B:71 derived from one of the β-strands that comprises the Ntn-fold forms part of an extended loop which contributes to the calcium coordination through B:73, B:75 and B:76. Similarly A:146 is situated at the end of a helix which then forms an extended loop to A:152 making up part of the calcium coordination. Residue Glu A:152 is at the amino terminus of an α-helix in which the side chain projects outwards from the helix axes to interact with the calcium. The side chains of negatively charged residues at the ends of helices are frequently Ca^{2+} ligands (McPhale*n et al*., 1991).

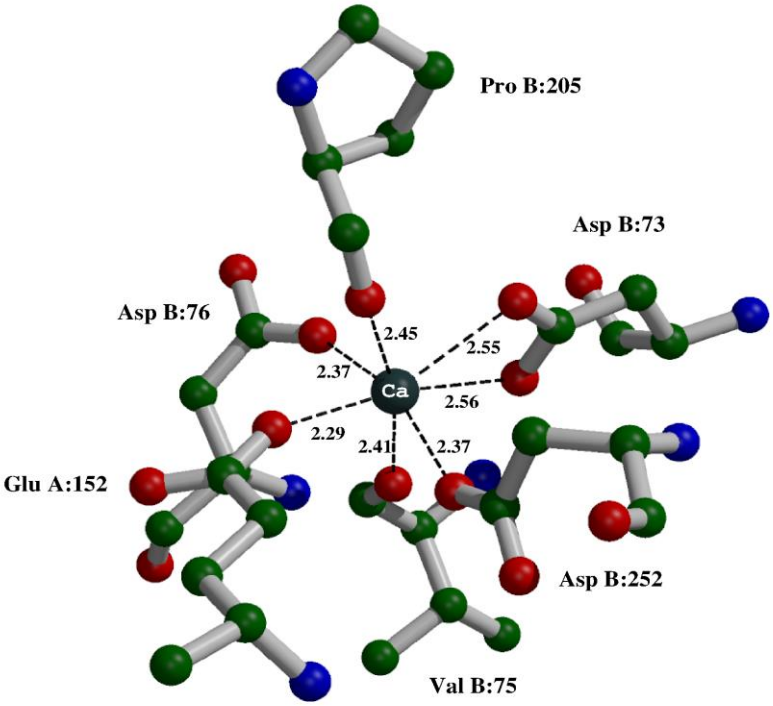


Figure 3.14 Calcium binding site of PGA. *The seven coordinate calcium exhibits distorted octahedral geometry.*

Residue	Ligand atom	Monoclinic	Triclinic
Glu A:152	OE1	2.29	2.49
Asp B:73	OD1	2.55	2.68
Asp B:73	OD2	2.56	2.68
Val B:75	O	2.41	2.58
Asp B:76	OD1	2.37	2.61
Pro B:205	O	2.45	2.62
Asp B:252	OD1	2.37	2.41

Table 3.10 Comparison of calcium ligand distances between monoclinic and triclinic PGA binding sites.

3.3.4 Ethylene glycol binding: Implications for structure interpretation and as a probe for mapping potential substrate interactions

A recent review of cryocrystallography (Garman and Schneider, 1997) suggests some caution should be taken when considering structural results arising from the use of cryoprotectants and describes some potential difficulties in model building and map interpretation when cryosolvents are used especially at low resolution. In particular, they warn of possible conformational changes in proteins induced by binding of cryoprotectants. For example, the cryoprotectant glycerol was shown to be a competitive substrate in studies with glycogen phosporylase *b* (Gregorio*u et al*., 1998; O'Reill*y et al*., 1997; Zographos *et al*., 1997). O'Reilly showed that glycerol mimics

the hydrogen bonding interaction of the O6, O4 and O3 hydroxyls of glucopyranose with only a short soak (< 3 sec) in 30% glycerol required to displace the glucopyranose spirohydantoin inhibitor (O'Reill*y et al*., 1997). In this case they resorted to a change in cryoprotectant using instead 2-methyl-2,4-pentanediol (MPD) which did not exhibit any such deleterious competition (Gregorio*u et al*., 1998).

Proteins interact with their natural ligands in binding pockets that have evolved to optimise the functional viability of the complex. This viability for an enzyme involves factors such as high substrate specificity, a binding constant that reflects the optimum time of duration of the complex and a geometry of binding that is conducive to a chemical reaction. Indeed, a novel approach has been described to locate and characterise ligand binding sites, both functional and non-functional, on protein surfaces (Mattos and Ringe, 1996): multiple solvent crystal structures (MSCS). This entails the X-ray determination of a target protein solved in a variety of organic solvents. Each type of solvent molecule serves as a probe for complementary binding sites on the protein. The probe distribution on the protein surface allows the location of binding sites and the characterisation of the potential ligand interactions within these sites.

Reflecting the 30% ethylene glycol solution used in the cryocooling for PGA, twenty-four ethylene glycol molecules were detected and fitted to the electron density maps of the *frz*PGA-1.3 structure. In general the carbon backbones of the ETG molecules pack against non-polar residues with the hydroxyl groups adopting various conformations in order to form hydrogen bonds. Some ETGs have alternate conformations. The majority lie on the surface of the protein although a few are partly buried in the protein interior. Several are located in regions close to the protein crystal contacts and some directly mediate these contacts, Figure 3.15.

There are four ethylene glycol (ETG) molecules in the active site of PGA, Figure 3.16, and provide information about potential binding sites beyond those identified for the S1 subsite prior to this thesis. ETG1 is positioned in the S1 specificity pocket which holds the phenylacetyl moiety of the substrate. ETG2 spans the active site nucleophile by H-bonding with both the α-amino group and the serine hydroxyl of Ser B:1, displacing the catalytic water. Two more ETG molecules lie close by, ETG3

involved in a hydrophobic interaction with the aromatic ring of Phe B:71 and ETG4 with Phe B:256.

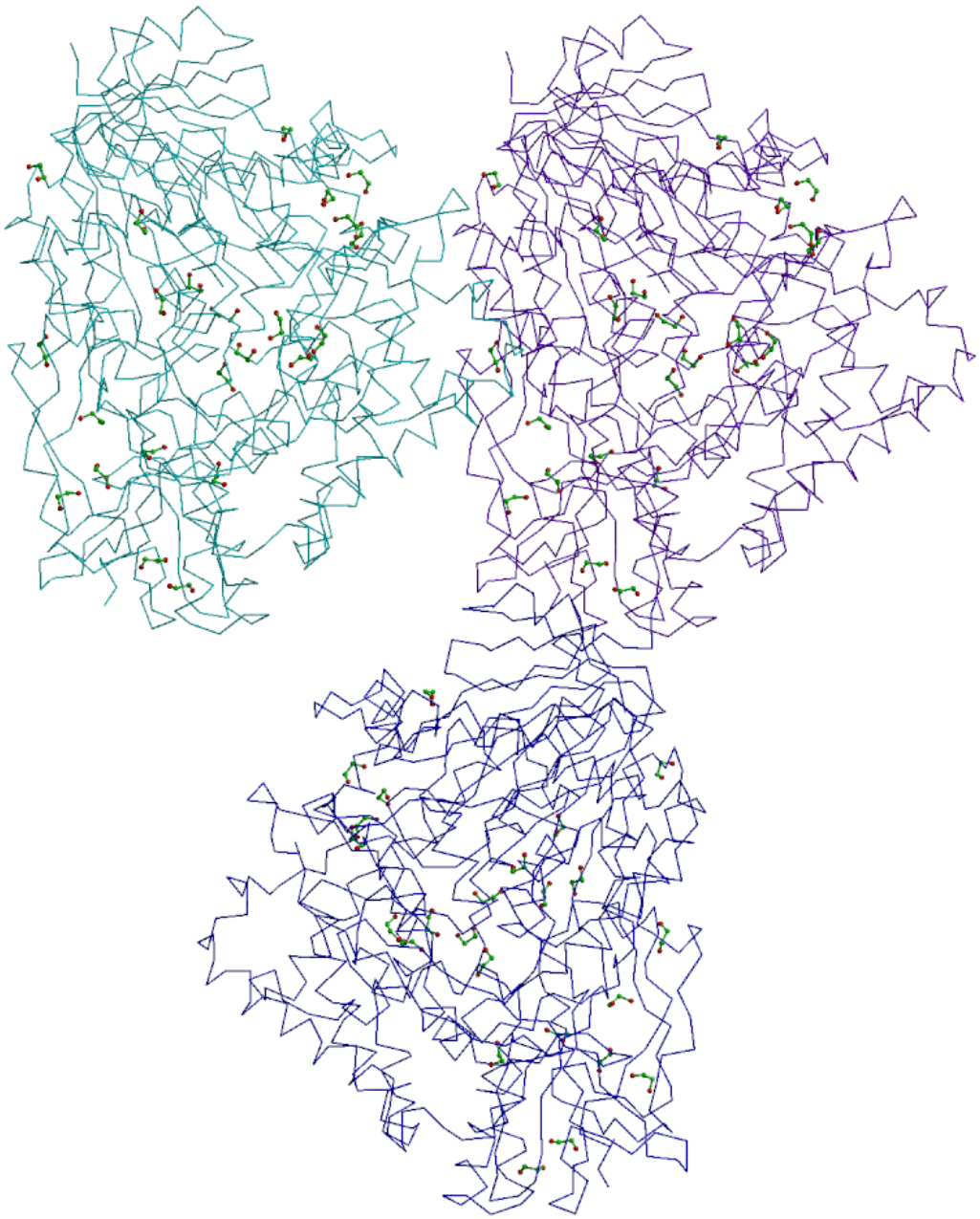


Figure 3.15 Ethylene glycol interactions and their involvement in mediating crystal contacts.

Thus some ETG molecules in the active site parallel the interactions already known to outline the specificity pockets and closely mimic the interactions necessary for binding of phenylacetyl substrate. This may clearly have a bearing on the detailed conformation of PGA, especially at 100 K, which could affect or indeed prevent the binding of substrates in the active site. The results described in later Chapters suggest this is not a problem and that good substrates are more strongly and preferentially bound.

In addition, caution must be taken with respect to ETG binding when interpreting structural changes in the protein with respect to the effect that cryosolvent may have on protein conformation. For example, the observed disorder in the helical region A:139 to A:146 may be influenced by binding of ETG in the S1 specificity pocket. Disorder in this region is observed when some derivatives of PAA are bound in the pocket (Don*e et al*., 1998), so the ETG in the pocket may occupy a similar position to those PAA derivatives which cause this conformational change.

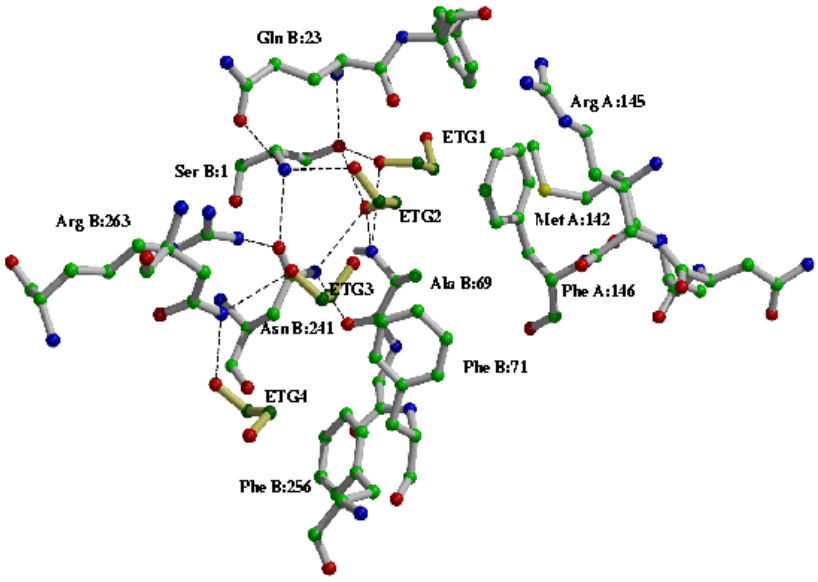


Figure 3.16 The hydrogen bonding interactions of ethylene glycol in the active site of PGA.

3.4 Conclusions

The addition of a suitable cryoprotectant, ethylene glycol, in the crystallisation conditions gave rise to a new crystal form of PGA. The electron density maps at 1.3 Å were vastly superior to those in the 1.8 Å at room temperature structure and substantially easier to interpret. The present results on the monoclinic form provide a more precise description of native PGA with more accurate atomic parameters. Comparison of the *RT*PGA-1.8 and frzPGA-1.3 structures shows a significant drop in overall temperature factor by ~14 $Å^2$. The low temperature high resolution structure has highlighted a number of dual side chain conformations and enabled modelling of these conformations.

In this context, it must be remembered that ethylene glycol binding may have a direct influence on the structure of PGA. The helix-coil transition in the active site might partly result from binding of ETG molecules. The affinity of ETG molecules for the active site may be linked to the its hydrophobicity, since ETG packs against aromatic side chains with the hydroxyl groups forming hydrogen bonds to the main chain. The interaction of ETG2 with the active site serine nucleophile has implications for ETG inhibition of PGA and future work should involve the determination of the inhibition constants for ETG. The potential drawback is displacement of the proposed catalytic water from the active site and prevention of substrate binding. The results in Chapter 5 suggest this is not a problem for good substrates. Nevertheless identifying an alternative cryoprotectant could be useful for weakly binding ligands.

The improved quality of the structure is due to a number of factors. The monoclinic space group results from new interactions between PGA molecules, which are enhanced through additional lattice interactions involving ETG molecules. In combination with cryogenic freezing, this led to data of greatly enhanced resolution and quality to those previously attained. The relative importance of the presence of the cryoprotectant and of freezing for initial crystals quality is likely to be higher for the latter: for reduced radiation damage, the freezing is vital.

The low temperature structure revealed dual main chain conformations for B:369 to B:378. More interestingly it allowed definition of two conformational states for A:145 to A:146, which had previously only been observed in binding of phenylacetic

acid derivatives where two subsets of derivatives with different structural attributes were responsible for each of the observed conformations (Don*e et al*., 1998).

Chapter Four Inactive mutants of PGA and a rationale for their design

4.1 Introduction

Design of an inactive variant of PGA provides scope for studying its function and substrate specificity. The starting point and obvious target for generating inactive mutants is the active site which endows the enzyme with its catalytic capability. The information gleaned from the X-ray structure of PGA provides the means to target specific amino acid residues for structure:function studies, in combination with site-directed mutagenesis (SDM). The prospect of achieving an inactive enzyme is reliant on minimal disruption to the processing mechanism to produce a processed, or at least partially processed inactive enzyme.

The active site of PGA consists predominantly of residues which emanate from the Ntn-fold of the protein which forms the core of the active site and is composed solely of B chain residues (see Figure 1.11). The A chain helix (A:142–A:146) provides the remaining architecture of the active site, and provides the main determinant of substrate specificity. The N-terminal serine forms the focal point of the active site in which its nucleophilic character is proposed to be enhanced by its own α-amino group (see Chapter 1).

Mutation of the N-terminal serine to an alanine disables the nucleophilic action through removal of the side chain hydroxyl and is an obvious choice for the first line of attack. However, mutagenesis experiments demonstrated that Ser B:1 is also critical for processing of active enzyme from the precursor protein (Cho*i et al*., 1992). Mutation to either threonine, arginine or glycine results in non-processing of the precursor and highlights the relationship between the catalytic and processing mechanisms, while mutation to cysteine results in an inactive processed enzyme (Cho*i et al*., 1992). Interestingly, the active site nucleophile of penicillin V acylase is the N-terminal cysteine (Sures*h et al*., in press).

Crystallographic studies (Done, 1996; Duggleby *et al*., 1995) suggested that the oxyanion hole stabilises the transition state through the donation of two hydrogen bonds, one from the main chain amide of B:69 and one from the side chain amide of B:241. The amide side chain is positioned to stabilise the transition state by making a hydrogen bond between its side chain OD1 and the NH2 atom of the neighbouring residue Arg B:263 (Figure 4.1).

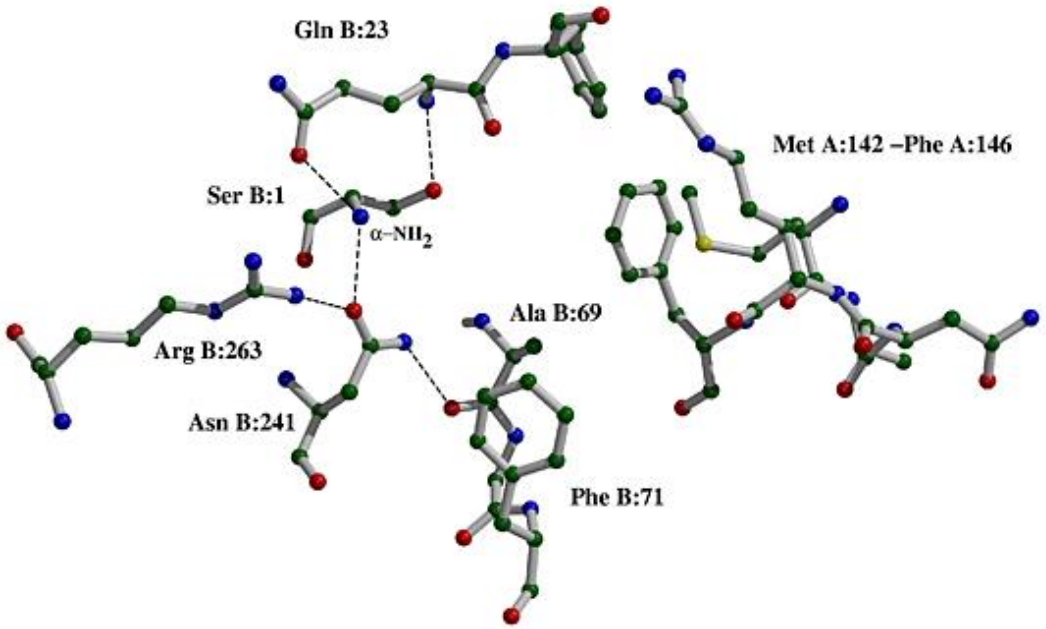


Figure 4.1 The active site of PGA showing the arrangement of catalytic residues and the hydrogen bonding network.

The ultimate goal in obtaining a suitable inactive enzyme was to obtain a 3-D structure of an enzyme : substrate complex, to provide an essential piece in the puzzle of PGA's

function and substrate specificity. The structure of the PAA complex (Duggleby *et al*., 1995) and complexes of tetrahedral intermediate mimics (Done, 1996; Duggleby *et al*., 1995) provide a glimpse of the reaction pathway of PGA. The complex with PAA, a competitive inhibitor, is the end-product of the catalytic pathway of PGA defining the specificity pocket for the acid moiety and provides both a structural rationale for PGA's preference for phenylacetylated compounds and speculation on the possible interactions in the S1' specificity pocket (Figure 4.2). The subsite description of the specificity pockets of PGA is based on the nomenclature as described by Schechter and Berger (Schechter and Berger, 1967), Figure 4.3.

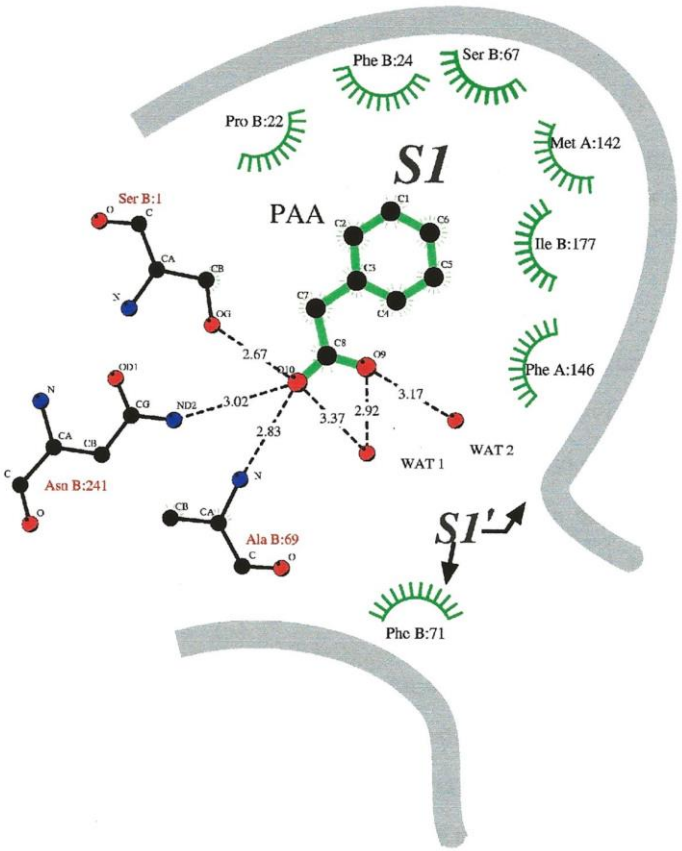


Figure 4.2 A schematic representation outlining the interactions of PAA which define the S1 specificity pocket and show the proposed S1' region.

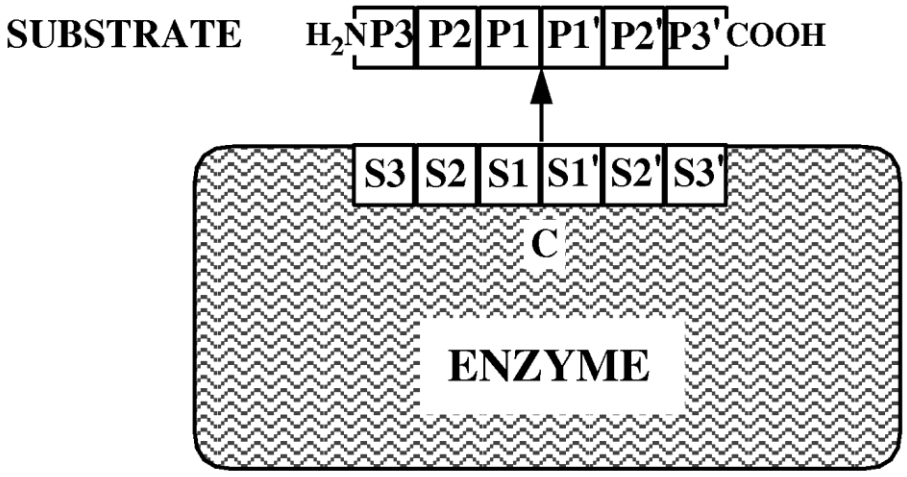


Figure 4.3 A schematic representation defining the S1/S1' nomenclature for the enzyme subsites located on both sides of the catalytic site C where the scissile peptide bond is between P1 and P1'. *Reproduced from Schechter and Berger, 1967.*

Structural determination of two tetrahedral transition state analogues, PMSF (Duggleb*y et al*., 1995) and benzylboronic acid (Done, 1996) provide convincing evidence for the interactions stabilising the transition state (Figure 4.4; see Figure 1.6 for PMSF complex). These results provide snapshots of PGA function as it progresses along the reaction pathway, outlining the primary specificity pocket and highlighting the key catalytic residues involved.

However, obtaining structural information on an enzyme : substrate in the ground state configuration resolves some of the remaining questions of substrate specificity and the catalytic mechanism of the enzyme. While the predominant specificity resides in the S1 pocket, a substrate complex will provide details of the extent of protein interactions in the amide portion of the substrate, i.e. what specificity lies in this region, if any, and what implication this has for the stereospecificity of the enzyme. PGA shows enantioselectivity for L-amino acids on the amide side chain of its substrates (Cole, 1969b; Švedas *et al*., 1996). The crystal structure of a substrate complex would provide insight to enable engineering of an efficient cephalosporin acylase.

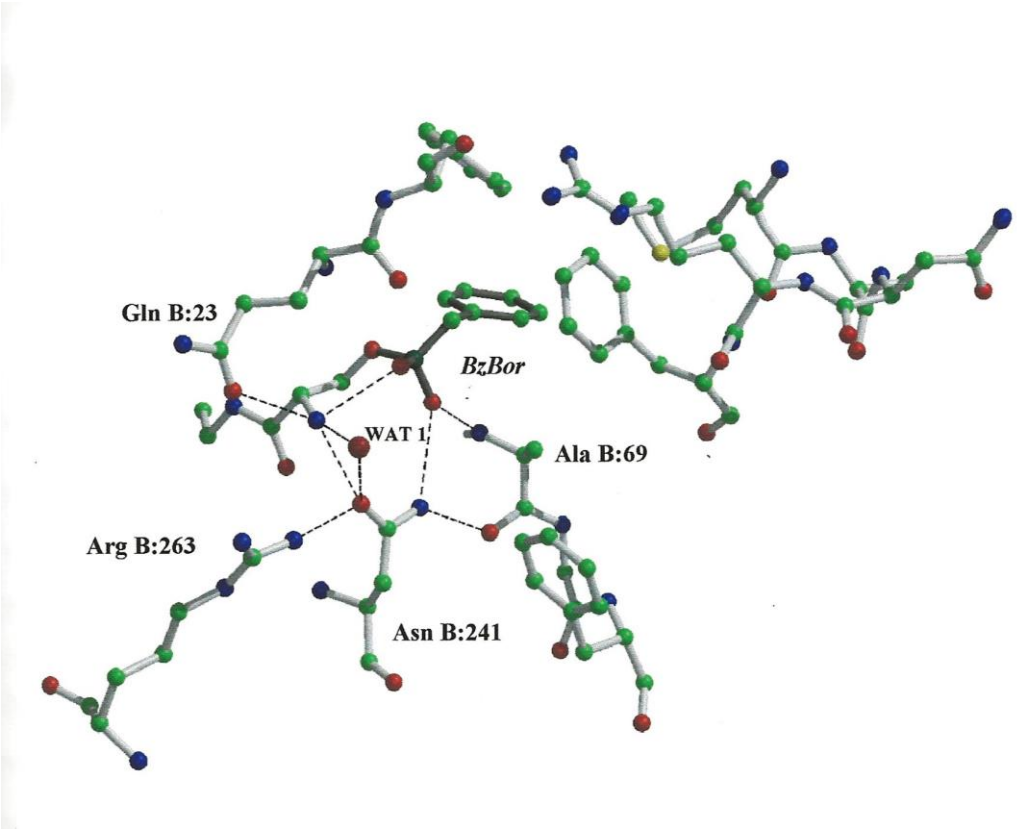


Figure 4.4 Representation of the important hydrogen bonding interactions in the oxyanion hole as defined by the transition state analogue benzylboronic acid (BzBor), WAT1 is the proposed "catalytic" water ((Done, 1996).

A number of potential inactive mutants are described below. The work was a continuation of investigations initiated by Dr. Jim Brannigan in which several active site residues were mutated and screened for enzyme activity. A series of mutations on the catalytic residue Ser B:1 were examined to probe its function, and single mutations were also made at Asn B:241 and Arg B:263.

None of the variant enzymes exhibited any detectable activity in unpurified culture supernatants when assayed with the chromogenic substrate NIPAB. The mutants were investigated to identify suitable candidates for the structure determination of an enzyme:substrate complex.

4.2 Experimental

The mutants were cloned into the wild type expression clone pA1, a derivative of the expression vector pACYC184 containing the 3.24 kb *pac* gene. To minimise spurious mutation sites, a fragment of wild type *pac* in pA1 was replaced with the same region containing the mutation of interest, using specific restriction enzymes unique to *pac* (see Appendix for *E. coli pac* sequence).

4.2.1 Cloning of PGA mutants

The mutants were provided by Dr. Jim Brannigan in the plasmid Bluescript KS (-) in which only low levels of PGA expression are found. To characterise the mutants and provide sufficient quantities of protein both for biochemical and crystallisation experiments, the mutants were replacement subcloned into the expression plasmid pA1. The plasmid is used for constitutive expression of wild type PGA since the cloned *pac* retains its promoter sequence.

Both DNA from the plasmid pA1 and the mutant plasmids were purified from agarose gels following restriction digest with the appropriate enzymes as described in Chapter 2. The mutant Ile B:263 was subcloned from Bluescript KS (-) as a *Bgl*II/*Mlu*I fragment replacing the same DNA fragment from the expression clone pA1. The Asn B:241 Ala (Ala^{B241}) mutant was subcloned as a *Bgl*II/*Spl*I fragment and the Ser B:1 Cys (Cys^{B1}), ΔSer B:1 (ΔSer^{B1}), and Ser B:1 Ala (Ala^{B1}) mutants were subcloned as *Dra*III/*Bgl*II fragments in a similar manner (Figure 4.5). Successful cloning of the mutants was verified by digestion of plasmid DNA with the marker restriction enzyme which gives a restriction pattern unique to each mutant that distinguishes it from wild type *pac*. The Ala^{B241} mutant was verified by restriction digest of the plasmid DNA with *Eco*RI, while the Ile^{B263} mutant was verified with *Eco*RV. Similarly, the mutations at Ser B:1 were each verified by digestion with the restriction enzyme *Kpn*I.

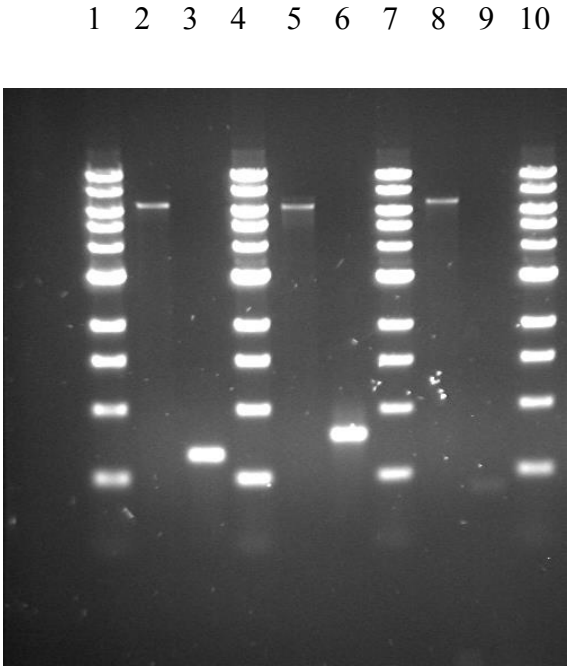


Figure 4.5 Cloning of inactive mutants. *Lane 1,4,7, and 10: 1 kb DNA ladder (NEB). Lane 2: Δ Bgl*II*/Mlu*I *pA1. Lane 3: 655 bp Bgl*II*/Mlu*I *fragment of* Ile^{B263}*. Lane 5: Δ Dra*III*/Bgl*II *pA1. lane 6: 794 bp Dra*III*/Bgl*II *fragments of mutants Δ* Ser^{B1}*,* Cys^{B1}*, and* Ala^{B1}*. Lane 8: Δ Bgl*II*/Spl*I *pA1. Lane 9: 395 bp Bgl*II*/Spl*I *fragment of* Ala^{B241}**.**

4.2.2 Expression of PGA mutants

The five mutant clones in the Bluescript plasmid as provided by Dr. Brannigan (see below), from initial expression and activity screening with NIPAB showed negligible activity. After replacement subcloning into the expression vector pA1, the plasmids containing the mutations were transformed into competent cells of the *E. coli* expression strain BL21 (DE3). Small scale expression trials were carried out with the mutants and analysed by SDS-PAGE, Figure 4.6. Inactivity in the mutants can be subdivided into those which are not processed, or those which are processed and inactive, Table 4.1.

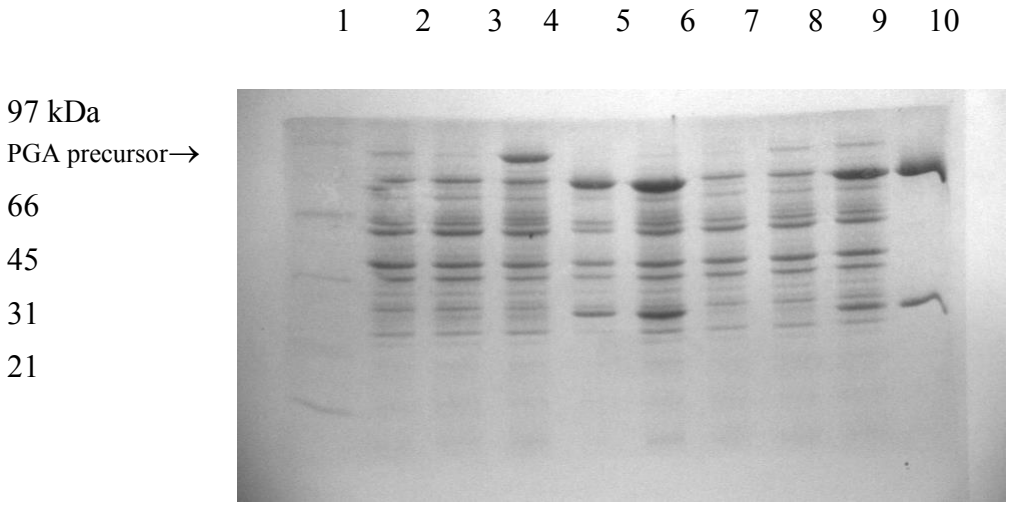


Figure 4.6 Expression of inactive mutants of PGA. *Lane 1: Protein markers. Lane 2: Ala^{B1}. Lane 3: Δ Ser^{B1}. Lane 4: Ile^{B263}. Lane 5: Cys^{B1}. Lane 6: pA1. Lane 7: pACYC184. Lane 8: Ala^{B1} 1/2h RT. Lane 9: Ala^{B1} + PGA 0.5h RT. Lane 10: purified PGA.*

Reference code	Mutation	Marker restriction site	Expressed in periplasm	Processed	Activity with NIPAB assay
Ala^{B241}	B:241 Asn→Ala	*Eco*RI	▤	▤ (slowly)	📻
Ile^{B263}	B:263 Arg→Ile	*Eco*RV	▤	📻	📻
ΔSer^{B1}	B:1 ΔSer	*Kpn*I	**ND**	**ND**	📻
Ala^{B1}	B:1 Ser→Ala	*Kpn*I	▤	📻	📻
Cys^{B1}	B:1 Ser→Cys	*Kpn*I	▤	▤	📻

ND – not detected

Table 4.1 A table showing the characteristics of the inactive mutants.

The mutants Ile^{B263} and Ala^{B1} both disrupt the processing mechanism and lead to expression of precursor protein only, with higher expression levels for Ile^{B263}. The addition of wild type PGA with the Ala^{B1} mutant followed by incubation for 30 min at room temperature showed no change in the level of precursor protein compared to incubation without wild type PGA (Lane 7 and 8; Figure 4.6). This supports the hypothesis that the processing mechanism is intramolecular rather than intermolecular [Kasche, in preparation #976].

The Ala^{B241} mutant was characterised by a mixture of precursor and processed protein in a ratio of ~1:4 (Lane 3; Figure 4.7). Increasing the induction time in the fermentation of Ala^{B241} gave a build up of precursor which suggests that the mutation decreased the processing efficiency but does not have a critical role in the autocatalysis. However, in wild type PGA, Asn B:241 forms a strong hydrogen bond with the NH2 atom of Arg B:263. Since mutation of Arg B:263 completely knocked out processing of the enzyme, this hydrogen bond may have an important role in the processing mechanism. Deletion of the active site serine resulted in no detectable expression of the mutant suggesting this deletion may disrupt the correct folding of the Ntn motif and hence the precursor.

The serine to cysteine mutation at position B:1 was processed and inactive, demonstrating that replacement of one nucleophile by another does not affect the autocatalytic mechanism but specifically disables its catalytic function. Structural determination of the Cys^{B1} mutant was undertaken to elucidate these findings.

1 2 3 4 5 6 7

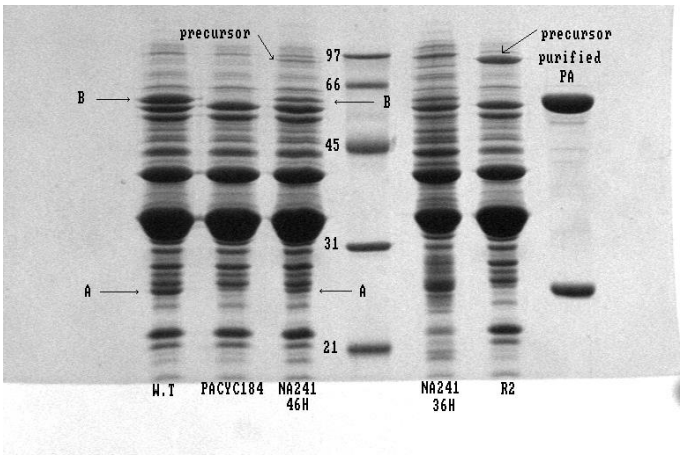


Figure 4.7 Expression of the mutants Ala^{B241} and Ile^{B263} mutants. *Lane 1: pA1. Lane 2: pACYC184. Lane 3: Ala^{B241} (46 h). Lane 4: Protein markers. Lane 5: Ala^{B241} (36 h). Lane 6: Ile^{B263}. Lane 7: Purified PGA.*

4.2.3 Purification of Ala^{B241} and Cys^{B1}

The Ala^{B241} and Cys^{B1} mutants were chosen for structural studies with the aim of determining the structure of an enzyme : substrate complex. Large scale protein preparation and purification were carried out for both, Chapter Two. The protein fractions following phenyl sepharose chromatography also include precursor protein in the acylase fractions (lanes 3-5; Figure 4.8). However the precursor protein band disappears following Q-Sepharose chromatography, probably due to the alkaline conditions (Figure 4.9).

Biochemical characterisation of the two mutants is described below. The results of the structural studies for the Cys^{B1} mutant are described later in this chapter, while structural details of the Ala^{B241} enzyme will be discussed in Chapter five.

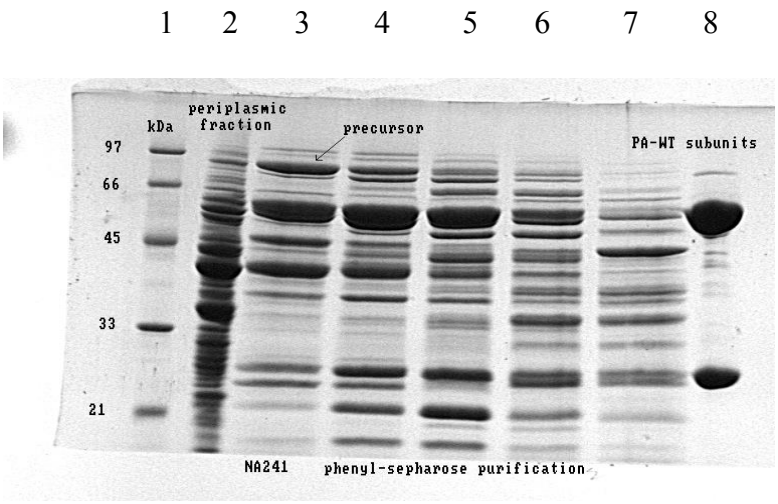


Figure 4.8 Purification of Ala^{B241} by hydrophobic interaction chromatography. *Lane 1: Protein markers. Lane 2: Periplasmic fraction. Lane 3-7: Ala^{B241}. Lane 8: Purified PGA.*

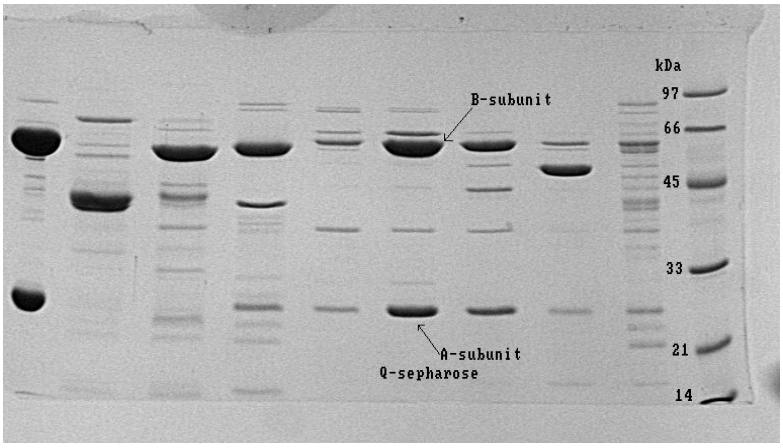


Figure 4.9 Purification of Ala^{B241} by anion exchange chromatography. *Lane 1: Purified PGA. Lane 2-3: non-PGA protein fractions. Lane 4: Ala^{B241} second peak fraction. Lane 5-8: Ala^{B241} protein fractions. Lane 9: Pooled phenyl sepharose fractions. Lane 10: Protein markers.*

4.2.4 Characterisation of Cys^{B1}

Characterisation of the purified Cys^{B1} mutant by non-reducing SDS-PAGE showed the presence of A (23.2 kDa) and B (62.8 kDa) chains, but also an abnormal high molecular weight protein band corresponding to a molecular weight of approximately 120 kDa, larger than the precursor form of PGA (95 kDa), and represented ~50% of the total protein. The protein was analysed under varying conditions including pH and reducing agents to characterise this anomaly (Figure 4.10).

One explanation for the 120 kDa protein band is that it may correspond to a B chain dimer linked by a disulphide bridge between the N-terminal cysteines. Native PGA from *E. coli* contains no cysteine residues. The incorporation of a cysteine residue in the active site could result in the formation of disulphide intermediates during folding of the protein.

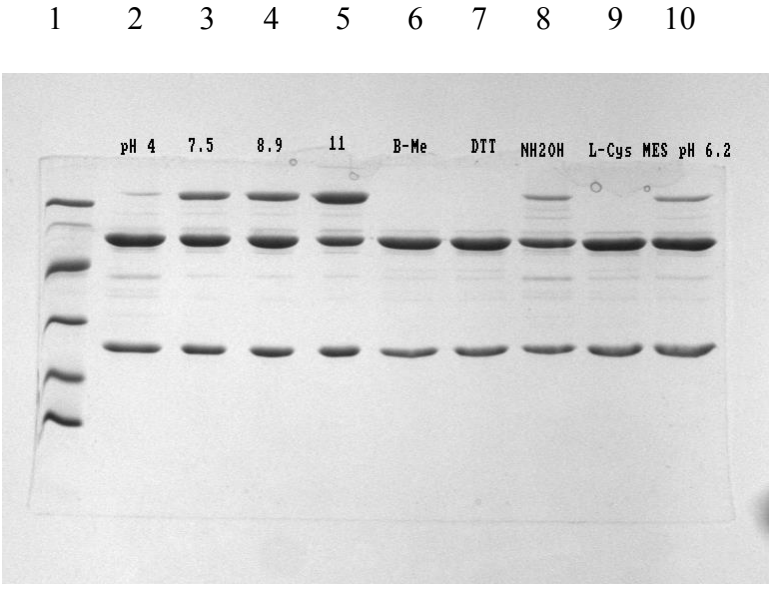


Figure 4.10 A non reducing 12% SDS-PAGE gel showing the effect of pH and various reagents on the Cys^{B1} mutant. *Lane 1: Protein markers. Lane 2: 100 mM sodium acetate pH 4.0. Lane 3: 100 mM Tris-HCl pH 7.5. Lane 4: 100 mM Tris-HCl pH 8.9. Lane 5: 100 mM CAPS pH 11. Lane 6: 1 mM β-ME. Lane 7: 1 mM DTT. Lane 8: 250 mM hydroxylamine. Lane 9: 1 mM L-Cysteine. Lane 10: 100 mM MES pH 6.2.*

The protein band disappears upon addition of a reducing agent, e.g. β-ME, and supports this hypothesis. Refolding experiments of PGA by Lindsay and Pain (1990) suggest a strong tendency of the B chain peptide for rapid self-aggregation to a state in which it will not recombine with the A peptide. This supports the hypothesis that if there is misfolding of PGA protein then self-aggregation of B subunits could occur. Surprisingly, decrease in pH from 9→4 coincided with the disappearance of the protein band which contradicts this hypothesis since one would expect a disulphide to be relatively stable over this pH range. Analysis of the cysteine mutant by native PAGE show the mutant has different mobility dependent on pH and the presence of a reducing agent (Figure 4.11).

Preliminary dynamic light scattering (DLS), however, showed the presence of mainly a single monomeric species with a small percentage of larger aggregates (see Appendix for results). This is inconsistent with the SDS-PAGE analysis and does not support the hypothesis that the anomaly observed is a dimeric species.

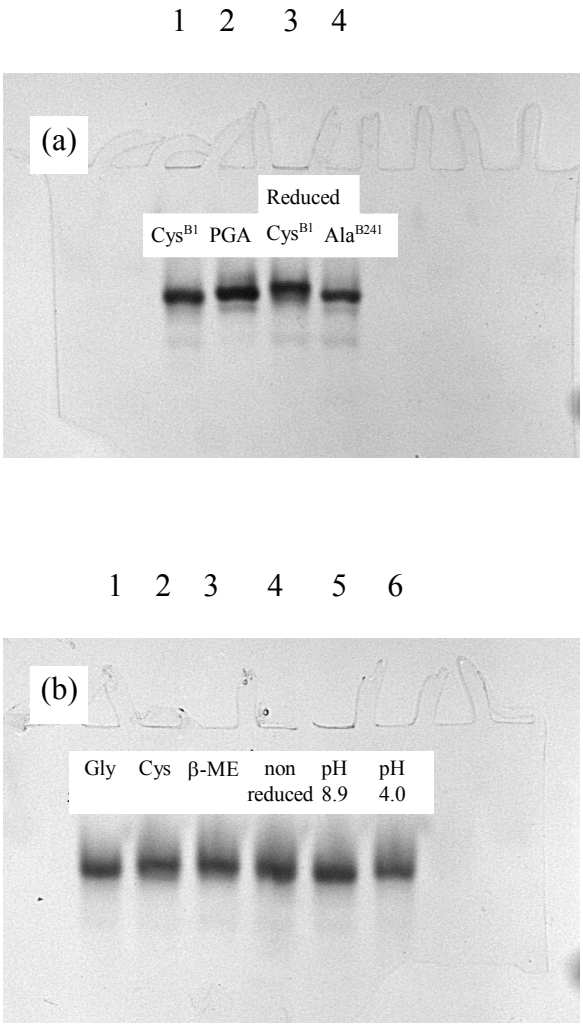


Figure 4.11 A 7.5% native PAGE gel showing the effect of reducing agents and pH on the mutant Cys^{B1}. *(a) Lane 1: Unreduced Cys^{B1}. Lane 2: wild type PGA. Lane 3: Reduced Cys^{B1}. Lane 4: Ala^{B241}. (b) Lane 1: Cys^{B1} in 10 mM L-Glycine. Lane 2: Cys^{B1} in 10 mM L-Cysteine. Lane 3: Cys^{B1} in 1 mM β-ME. Lane 4: non-reduced Cys^{B1} in 100 mM Tris-HCl, pH 7.5. Lane 5: Cys^{B1} in 100 mM Tris-HCl, pH 8.9. Lane 6: Cys^{B1} in 100 mM sodium acetate, pH 4.0.*

4.2.4 Crystallisation, data collection and refinement

Two strategies were used to grow crystals of Cys^{B1} based on previous attempts by Dr Shirley Tolley. Crystallographic data to 3 Å were collected at room temperature by Dr. Marek Brzozowski on the triclinic crystal form of Cys^{B1} grown in the presence of 1 mM β-mercaptoethanol (β-ME). Analysis of the electron density maps confirmed the mutation but also showed extra density around the cysteine side chain, which while difficult to interpret suggested the presence of a β-ME molecule. Work on these data was abandoned because of the limited resolution.

Two different approaches were subsequently taken to crystallise the Cys^{B1} mutant. Firstly, 1 mM β-ME was added to reduce the high molecular weight species of Cys^{B1}. The protein was then buffered exchanged into 50 mM MOPS, pH 7.2 to remove the reducing agent prior to crystallisation. Secondly, the unreduced protein was used immediately for crystallisation in the hope that the percentage of correctly folded protein was sufficient for good crystal quality. Crystallisation was carried out by streak seeding using wild type monoclinic crystals, which were used for a second round of seeding and the crystals obtained were used for data collection.

Crystals were flash frozen at 120 K using an Oxford cryo-system. In-house diffraction data were collected using a MAR image plate, Table 4.2. Three datasets of the Cys^{B1} mutant were collected. The first was obtained from a crystal grown from protein treated with β-ME (Cys^{B1}_R), with a second from protein crystallised without β-ME (Cys^{B1}_NR) and a third from a crystal soaked with Pen G (Cys^{B1}_Ox).

Data statistics	**Cys^{B1}_R**	**Cys^{B1}_NR**	**Cys^{B1}_Ox**
In-house data collection	MAR (30 cm)	MAR (30 cm)	MAR 345(34.5 cm) Mirrors
Space group	$P2_1$	$P2_1$	$P2_1$
Cell parameters (Å)	a = 51.15	a = 51.08	a = 51.18
	b = 131.81	b = 131.43	b = 131.82
	c = 64.32	c = 63.93	c = 64.21
	β =105.81°	β =105.83°	β =105.79°
Resolution (Å)	25 – 2.0	12 – 2.0	15 – 1.9
Mosaicity (°)	0.4	0.4	0.3
Temperature (K)	120	120	120
Reflections measured	359898	195605	377717
Unique reflections	52970	46353	67203
$\langle I \rangle$ / $\langle \sigma(I) \rangle$	18.9 (5.9)[a]	12.7 (3.7)[b]	26.1 (6.5)[c]
Completeness (%)	100(100)	90.8 (89.2)	98.3 (91.4)
*R_{merge} (outer shell)	6.5 (21.9)	5 (19)	3.7 (15.1)

Numbers in parentheses refer to the outermost resolution shell [a](2.06-2.03 Å) [b](2.08-2.04 Å) [c](1.89-1.86 Å).

*R_{merge} = $100\sum_{hkl}|I_{hkl} - \langle I \rangle| / \sum_{hkl} I_{hkl}$

Table 4.2 Data reduction statistics for Cys^{B1}.

The starting model for refinement of the Cys^{B1} mutant was the 1.8 Å wild type structure with water molecules removed from the active site. As discussed later, refinement of β-ME used the parameter information in the *PROTIN* dictionary with additional distance restraints implemented in *REFMAC* between Cys B:1 and β-ME. Water molecules were added to the active site using *X-SOLVATE* with the exception of the 1.9 Å Cys^{B1}_Ox structure where additional waters were added using *ARP* combined with *REFMAC*. 3 % of the data were removed and used as a Free R set for cross validation and the refinement statistics are presented in Table 4.3.

4.3 Results and Discussion

4.3.1 Enzyme activity

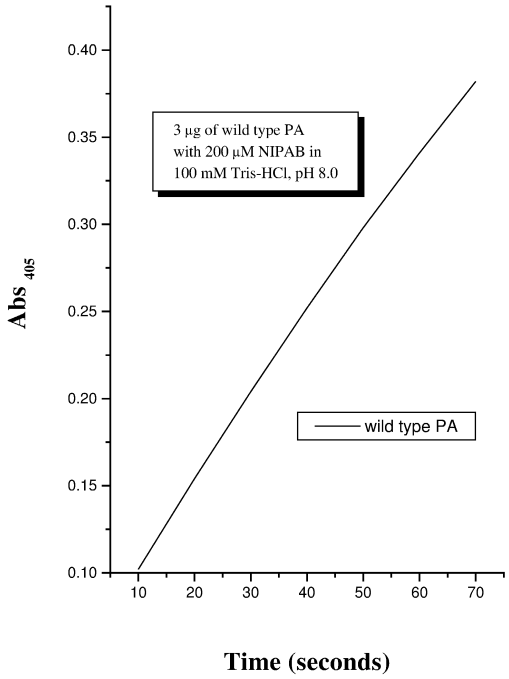
Both purified mutants were assayed with the chromogenic substrate NIPAB. The results confirmed the previous findings obtained from assays carried out with crude periplasmic extracts. As expected the Cys^{B1} mutant showed negligible activity. Surprisingly, the Ala^{B241} mutant showed no activity even after 30 min incubation giving a rate of reaction of 0.006 A.U/hr (A.U; absorbance units) in contrast to wild type enzyme which gave a typical rate of 0.28 A.U/min at λ = 405 nm (Figure 4.12). This corresponded to an approximate 10,000 fold reduction in activity suggesting that the enzyme's catalytic function was dramatically disabled by this mutation. This was unexpected as similar mutations at the equivalent active site asparagine in the serine proteinase subtilisin showed a reduction in activity of between 300 - 2000 fold (Bryan *et al*., 1986).

Refinement	**Cys^{B1}_R**	**Cys^{B1}_NR**	**Cys^{B1}_Ox**
Resolution (Å)	25 - 2.0	12 - 2.0	15 - 1.9
†R (‡R_{free}) factor (%)	16.5 (21.4)	15.7 (21.6)	15.4 (20.0)
Non-hydrogen atoms			
protein	6111	6104	6126
solvent	690	693	833
*β-ME / ethylene glycol	*4	–	68
R.m.s deviations from ideality (Å)			
bond 1-2 distance (0.02)	0.016	0.016	0.016
angle 1-3 distance (0.04)	0.036	0.037	0.032
planar 1-4 distance (0.05)	0.039	0.039	0.037
Average B factor (Å^2)			
all atoms	19.4	22.6	20.2
protein	18.5	21.5	18.6
solvent	27.5	32.2	31.3
Ca^{2+} ion	13	18.8	13.7
*β-ME / ethylene glycol	*27.5	–	26.8
§Coordinate error (Å)	0.179	0.199	0.126

†R-factor = $\sum_{hkl} \| F_{obs} | - k | F_{calc} \| / \sum_{hkl} | F_{obs} |$

‡Free R-factor = $\sum_{hkl \subset T} \| F_{obs} | - k | F_{calc} \| / \sum_{hkl \subset T} | F_{obs} |$ where $hkl \subset T$ represents 3 % of the diffraction data. §Coordinate error estimated by DPI, the Diffraction-data Precision indicator (Cruickshank, 1996).

Table 4.3 Refinement statistics for the Cys^{B1} mutant.



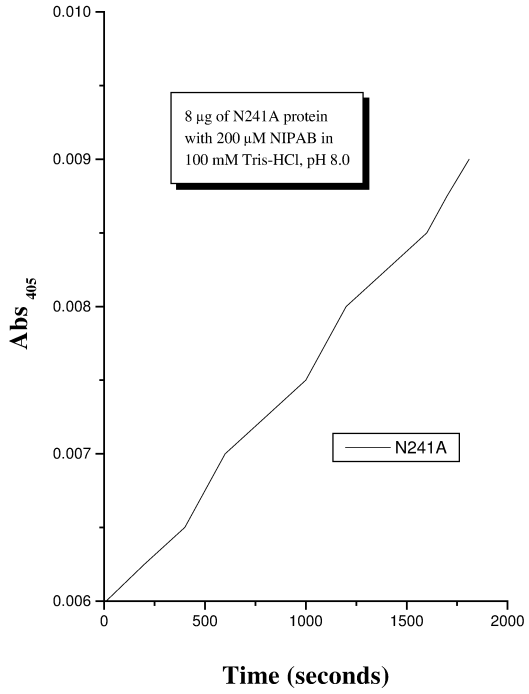


Figure 4.12 The enzyme activity of (a) wild type PGA and (b) the Ala^{B241} mutant, showing the change in absorbance at λ = 405 nm (A_{405}) with time.

4.3.2 Enzyme processing

The expression studies of active site mutants emphasise the intimate relationship between enzyme catalysis and processing of PGA (Figure 4.7). The mutagenesis results have built up a picture of the important residues involved in the processing mechanism and have highlighted residues that are common to both the catalytic and processing mechanisms. Co-purification of the precursor protein with the mature mutant Ala^{B241} in the first column purification step (Figure 4.8) hinted at a similar hydrophobic nature for both mature and precursor protein and suggests that the precursor protein has a similar overall fold. This has been confirmed by the recent structural determination of a precursor mutant (Dr. Lorraine Hewitt, personal communication). Increase of pH in the anion exchange column to pH 8.9 forced the processing of the protein, with a second fraction of Ala^{B241} eluting at a lower salt concentration, Figure 4.9.

4.3.3 Preliminary structural studies of Cys^{B1}

The X-ray structure of three forms of the mutant Ser B:1 Cys (Cys^{B1}) were all determined between 1.9 and 2 Å resolution. The mutation leads to a processed PGA which shows no measurable activity when assayed with NIPAB, in agreement with previous published mutagenesis results (Cho*i et al*., 1992; Slad*e et al*., 1991). The structure of Cys^{B1} was determined in three oxidation states resulting from the presence or absence of β-mercaptoethanol in the crystallisation procedure, and a substrate soak with penicillin G. The overall conformation of the active site remains unchanged in each structure apart from the mutated residue.

4.3.3.1 Reduced Cys^{B1}

The initial electron density maps showed a strong positive peak at the active site confirming the Ser B:1 Cys mutation. After several cycles of refinement of the cysteine, additional electron density around the sulphur (SG) atom was visible in the P1 pocket. This was interpreted as β-mercaptoethanol forming a disulphide bridge with the SG atom (Figure 4.13), although the sulphur of the β-ME had much lower density and a higher B value suggesting partial occupancy or high mobility.

The addition of β-ME to the protein prior to crystallisation to remove the disulphide intermediate species followed by buffer exchange to remove the β-ME, cannot however rule out reduction of the cysteine residue with β-ME. The β-ME molecule refines with higher B values than the cysteine side chain and suggests partial occupancy to be more likely.

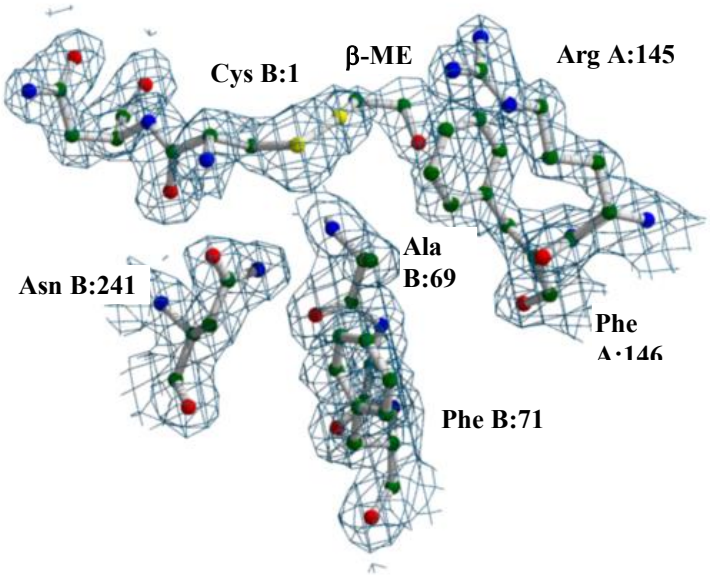


Figure 4.13 The final $2F_o - F_c$ electron density map contoured at 1 σ showing the refined atomic positions of the reduced active site cysteine B:1.

4.3.3.2 The Cys^{B1} structure in the absence of β-ME

The structure of the Cys^{B1} mutant was determined from a crystal grown from protein that had not been exposed to β-ME. The initial electron density maps confirmed the cysteine mutation and there was no additional electron density in the vicinity of the cysteine. After refinement the difference maps showed some positive peaks suggesting possible disorder of the cysteine or oxidation of the sulphur atom (Figure 4.14). In place of the β-ME in the reduced Cys^{B1} structure the electron density was modelled as two alternative water sites in the S1 pocket.

A side chain - main chain hydrogen bond between B:1 Ser OG and B:23 Gln main chain amide NH is lost when the serine is mutated (Figure 4.1). The B:1 Cys CB–SG bond is approximately 1.8 Å long in contrast to the CB–OG bond (~1.5 Å) of serine makes this hydrogen bond less favourable. This is conveyed by B values for the cysteine side chain which ranged from 17.5 $Å^2$ (CA) to 27.5 $Å^2$ (SG) suggesting side chain flexibility

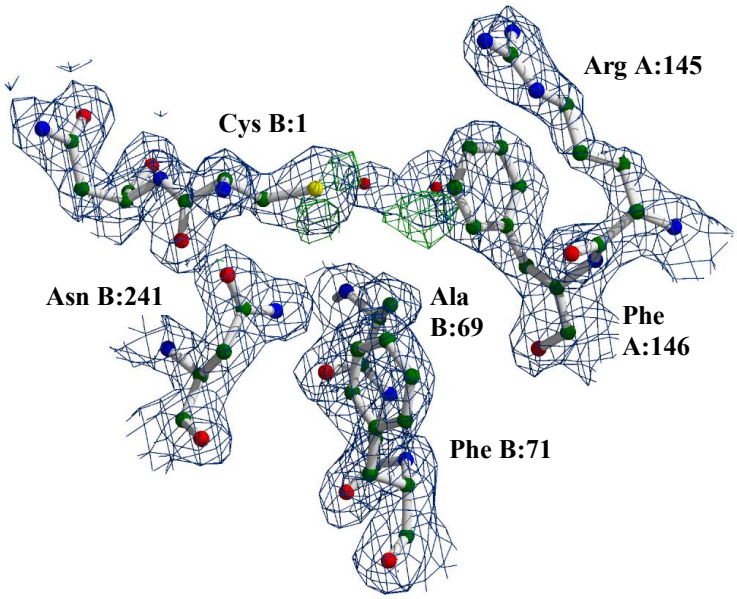


Figure 4.14 The final $2F_o - F_c$ electron density map contoured at 1 σ showing the refined atomic positions of the cysteine B:1 mutant.

4.3.3.3 Cys^{B1} soaked with Pen G

A crystal grown from protein of the Cys^{B1} mutant which had not been reduced with β-ME, was soaked with Penicillin G (10 mM) dissolved in mother liquor for 30 min. The maps showed clearly defined positive $F_o - F_c$ difference density around the sulphur atom of Cys B:1, suggesting the oxidised form of the cysteine side chain, which was modelled as two oxygen atoms to give the cysteic acid derivative (Figure 4.15). Residual positive density around the sulphur atom and the presence of partially occupied water molecules Wat1a, Wat1b, Wat2a and Wat2b, suggested an alternation conformation of the

oxidised sulphur. No further electron density was visible that could be interpreted as bound Pen G substrate.

The active site serine and residue Phe A:146 form either side of the neck of the S1 specificity pocket (Figure 4.2). Mutation to cysteine results in constriction of the neck due to the longer C–S bond and the greater van der Waals radius of the sulphur atom. The cavity volume of the S1 specificity pocket of Cys B1 is 98 $Å^3$ compared to the wild type pocket of 120 $Å^3$, the 22 $Å^3$ difference being attributed to the volume occupied by the cysteine side chain (Figure 4.16). The displayed cavities were calculated in *VOIDOO* (Kleywegt and Jones, 1994) and represents the volume occupied by a probe of radius 1.4 Å. One would expect movement of A:146 to accommodate Pen G in the binding pocket but the inactivity may not correspond to the situation observed in the Ala^{B241}-PG complex described later but may rather be due to the larger sulphur atom preventing access of the phenylacetyl moiety to the S1 subsite. Alternatively, oxidation of the active site cysteine could prevent substrate binding in the S1 pocket.

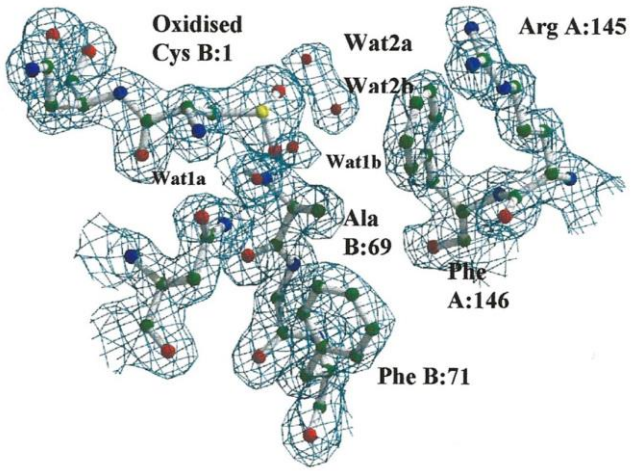


Figure 4.15 The refined atomic positions of the 1.9 Å structure of oxidised Cys^{B1} mutant following an attempted Pen G substrate soak. The $2F_o - F_c$ electron density map (blue) is contoured at 1 σ showing the oxidised cysteine.

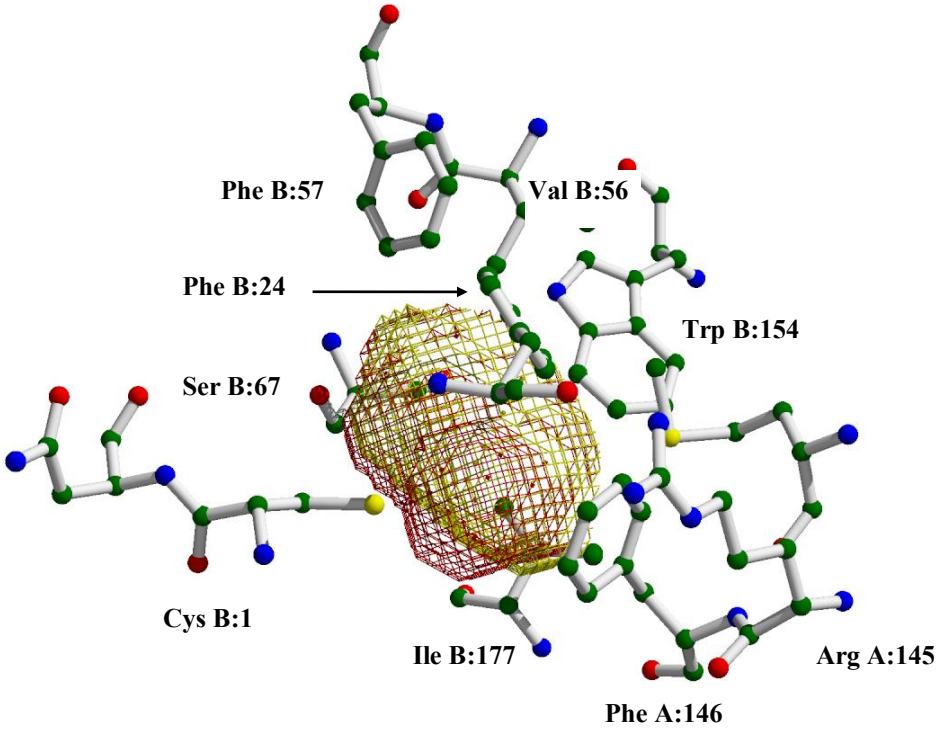


Figure 4.16 The calculated volume of the S1 specificity pocket. The active site volume for wild type PGA is coloured red while the volume of the Cys^{B1} mutant is coloured yellow.

4.3.4 Comparison with penicillin V acylase

The recent structure determination of penicillin V acylase (PVA) from *Bacillus sphaericus* revealed a similar active site structure (Suresh *et al*., in press). PVA is tetrameric with a monomer M_r of 35,000 in contrast to PGA which is a heterodimer of M_r 86,000 and the two enzymes share no obvious sequence homology. However, the X-ray structure established that its β-sheet and α-helix organisation and connectivity are characteristic of the Ntn hydrolase family. The N-terminal nucleophile in PVA is a cysteine which has a similar eclipsed geometry to that observed in the Cys^{B1} mutant of PGA.

Structural alignment reveals that the five key residues in the catalytic mechanism of PGA and PVA overlap (Figure 4.17) with a rmsd of 0.98 Å for CA atoms and 1.21 Å for all equivalent atoms.

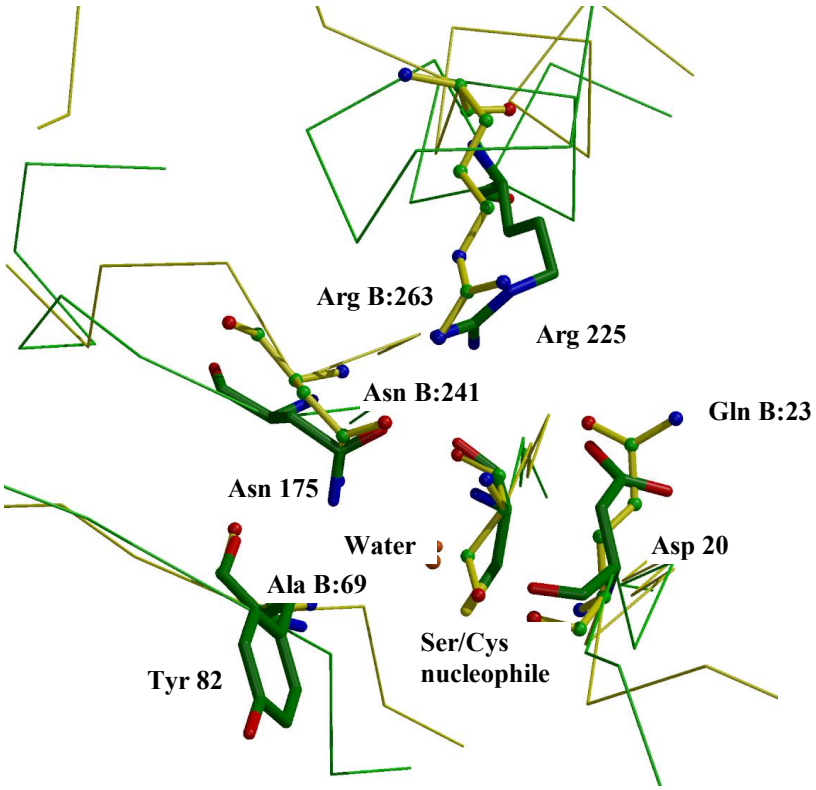


Figure 4.17 The overlap of the catalytic environment in PGA and PVA. *The PGA structure is represented with a yellow Cα trace and ball-and-stick representation for amino acids in the catalytic site and the oxyanion hole. The PVA structure is represented in liquorice but coloured green.*

4.4 Conclusions

Mutations were introduced at positions Ser B:1, Asn B:241 and Arg B:263 of PGA and the resulting variants overexpressed to varying levels of success. The experiments were designed to produce inactive mutants which would shed light on the active site of the

enzyme with regard to activity and specificity. Ser B:1 is the catalytic nucleophile, Asn B:241 is in the oxyanion hole and its position is stabilised by a hydrogen bond to Arg B:263.

Substitutions at Ser B:1 clearly demonstrate that this residue is not only involved in catalysis, but also in the nucleophilic attack needed for enzyme processing. This is supported by the fact that Ser B:1 Ala knocks out enzyme processing whereas mutation to a cysteine, also a nucleophile, still allows processing of the precursor, albeit at a reduced level. Experiments are currently underway in York to determine the structure of the Ala^{B1} unprocessed precursor mutant (Hewit*t et al*., in press).

Mutation of the neighbouring active site residue Asn B:241 to alanine resulted in a precursor which was only slowly processed to mature enzyme. Assay of the processed enzyme showed negligible activity. The slow processing by the mutant suggests a minor role for this residue in precursor processing. Inactivity of the mature enzyme demonstrates a more significant role in catalysis.

Overexpression of Arg B:263 Ile results in high levels of precursor protein and gives no processed enzyme, suggesting a key role in the processing mechanism. Thus mutation of Asn B:241 Ala and Arg B:263 Ile both have an additional influence on enzyme processing, not involving nucleophilic attack by the serine or other nucleophile, with the latter mutation having the most significant effect.

Thus in practice the mutations have revealed an intimate relationship between the autocatalytic processing of the precursor and the mechanism of catalysis by the mature enzyme. This to some extent confirms previous findings (Cho*i et al*., 1992) but also reveal some surprising results.

The lower mutant enzyme expression levels are probably related to the mechanism of processing of the pro-form of PGA during secretion into the periplasm. A similar finding has been reported for subtilisin (Brya*n et al*., 1986). The mechanism of enzyme processing in PGA is proposed to involve an autocatalytic step [Kasche, In preparation]. Characterisation of the mutants has provided insight on both the autocatalytic processing mechanism, which is a function of the pro-enzyme, and secondly, the catalytic function of the processed enzyme. These results highlight key residues in both the processing and catalytic functions and reveal the role of the active site residues in

both mechanisms. One explanation for the higher expression of the precursor variant Ile^{B263} compared to the Ala^{B1}, is that this mutation may have a greater destabilising effect on folding of the precursor.

In the transition state, the carbonyl group of the amide bond to be hydrolysed is proposed to adopt a tetrahedral configuration rather than the ground-state planar configuration. The specific binding forces involved in stabilising the transition state have been investigated crystallographically using inhibitors that mimic the tetrahedral intermediate (Done, 1996; Duggleb*y et al*., 1995). An important hypothesis emerging from these and similar investigations is that hydrogen bonding between protein groups and the developing negative charge on the substrate carbonyl oxygen of the transition state is a major contributing factor in lowering the free energy of the activated complex (Brya*n et al*., 1986).

Bryan showed that mutation of the oxyanion residue Asn 155 to Leu in subtilisin BPN' did not alter the K_m but greatly reduced the catalytic rate constant k_{cat} by a factor of 200-300 fold. The surprising 10,000 fold decrease in activity when the equivalent residue Asn B:241 in PGA was mutated to alanine suggests that loss of this hydrogen bond has other consequences in addition to stabilising the transition state. This hypothesis is supported further since the oxyanion hole could still form a single hydrogen bond with the Ala B:69 peptide NH group, which one would expect to provide sufficient stabilisation of the oxyanion to generate low levels of enzyme activity.

Two possible candidates were examined as suitable inactive enzymes for studying substrate specificity by X-ray crystallography. The most appropriate was the active site serine where the mutation Ser B:1 Cys would enable a study of both specificity and function. However, structural investigations of the cysteine mutant in this chapter suggested that it is not amenable to binding of substrate. The mutation may preclude substrate binding by restricting access to the S1 pocket (see Figure 4.16) which coupled with the problem of oxidation of the sulphur or β-ME binding discounts it as a viable option for structural studies of substrate binding.

The anomaly observed when Cys^{B1} is analysed by non-reducing SDS-PAGE remains unresolved and further work is needed. That the cysteine mutation in PGA is inactive

whilst the PVA enzyme's nucleophile is indeed a cysteine demonstrates the subtleties of enzyme active sites.

The second residue selected for SDM was Asn B:241, a key oxyanion hole residue, which was substituted by alanine. In contrast, Ala^{B241} was successfully processed and expressed at a much higher level, and a detailed account of substrate binding to this inactive mutant is given in the next Chapter. This exemplifies how mutation studies may often highlight the precise definition of catalytic residues, where mutations both drastic and similar can have dramatic effects on enzyme catalysis, but that other factors often come into play, such as activation of precursor in the case of PGA.

Chapter Five Substrate specificity of PGA: Studies on inactive substrate analogues and an inactive enzyme : substrate complex

5.1 Introduction

The substrate specificity of PGA has been extensively studied (Claridg*e et al*., 1963; Cole, 1969a; Plaski*e et al*., 1978; Ro*a et al*., 1995; Švedas *et al*., 1996). PGA is highly specific for phenylacetyl compounds but the structure of the leaving group of the substrates has minimal effect on the hydrolysis rate (Cole, 1964). Penicillin G (benzylpenicillin) is a preferred substrate for PGA (Margoli*n et al*., 1980). In contrast, penicillin G sulphoxide (PGSO) has been described as a non-hydrolysable substrate analogue for penicillin acylase (Plaski*e et al*., 1978), whilst Martín and co-workers (Martí*n et al*., 1993) have reported it to be a poor substrate for PGA from *Kluyvera citrophila* with activity less than 1% of that shown for penicillin G (Pen G), but present no kinetic data to support this. The only structural difference between Pen G and PGSO is the oxidation of the sulphur atom of the thiazolidine ring (Figure 5.1) and PGA has comparable affinities for both molecules with K_M values reported to be in the range 12–20 μM (Alvar*o et al*., 1991). Thus the oxidation of the sulphur of the thiazolidine ring is responsible for a dramatic decrease in enzyme activity without loss of binding affinity and the penam moiety of the substrate must play an important role in the catalytic efficiency of the enzyme so that it is not just the acid moiety of the substrate that determines specificity.

PGSO behaves as a general stabilising agent of immobilised PGA's from both *E. coli* and *K. citrophila* and this role is related to its strong inhibitory effect on the enzymes (Alvar*o et al*., 1991). The stabilising effect has been observed during thermal inactivation of soluble enzymes at alkaline pH (pH 10) and inactivation of immobilised enzymes as a consequence of covalent multiinteraction with activated agarose aldehyde gels. Nonetheless the structure of the PAA complex at 2.5 Å resolution was the result of a PGSO crystal soak at room temperature (Duggleby *et al*., 1995). Hydrolysis of

PGSO presumably resulted from the long data collection time required, several days, and the fact that the data were recorded at room temperature.

In chapter 3 the native structure of PGA in the monoclinic space group $P2_1$ was described in which binding of several ethylene glycol molecules traces the subsites where part of the substrate interactions are known to occur. Chapter 4 gives a detailed description of the design and expression of an inactive enzyme with the purpose of solving the structure of a substrate complex.

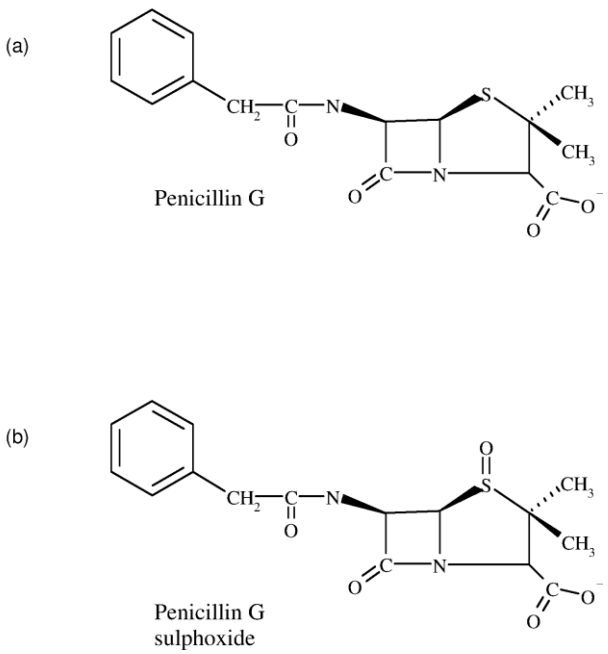


Figure 5.1 A schematic drawing of the chemical structures of (a) Penicillin G and (b) Penicillin G sulphoxide.

In this chapter substrate binding by PGA is probed by determining the crystal structures of the unliganded mutant Ala^{B241} along with its complexes with the substrate Pen G and PGSO. The structure of PGSO bound to the wild type enzyme is also presented. Collectively, this information has been used to infer the likely mode of substrate binding in the active enzyme. The structural data confirm and rationalise the essential role of Asn B:241 in catalysis, especially in light of the decrease in enzyme activity upon mutation to alanine.

5.2 Experimental

5.2.1 Crystallisation and substrate soaks

Inactive mutant PGA was crystallised under identical conditions as the wild type enzyme (McVey *et al*., 1997; Chapter Three, this thesis) to give isomorphous crystals in space group $P2_1$. Crystals of the Ala^{B241} mutant were soaked in Pen G (10 mM) or PGSO (7 mM), dissolved in a solution identical to the mother liquor from the crystallisation well. The final concentration of PGSO was essentially saturated at 7 mM owing to its poor solubility in the mother liquor solution. Crystals of wild type PGA were soaked in PGSO for a shorter time to minimise hydrolysis. The crystal soak times are summarised in Table 5.1. Following the substrate soaks, the crystals were briefly transferred to mother liquor containing 30% ethylene glycol before flash cooling but without the inclusion of substrates as solubility problems were experienced at higher ethylene glycol concentrations. This could lead to loss of Pen G or PGSO from the complex, but the results suggest this was not a serious problem.

Pen G (K^+ salt) was obtained from Sigma and PGSO (Na^+ salt) was a gift from Dr. José Guisán (Instituto de Catálisis, C.S.I.C. Serrano, Madrid, Spain).

Substrate Soaks	Penicillin G 10 mM	Penicillin G sulphoxide 7 mM
wild type PGA	–	3 min
Ala^{B241}	5 min	4 min

Table 5.1 Substrate soak times with crystals of PGA.

5.2.2 Data collection

For data collection the crystals were flash frozen in a gaseous nitrogen stream at 100 K using an Oxford cryo-system. All data were collected on beam line BW7B with a MAR 300 research imaging plate at the EMBL Hamburg Outstation. Data were indexed and integrated using the HKL suite (Otwinowski and Minor, 1997). The collection strategy for the high resolution data for the unliganded mutant was limited by crystal orientation and mosaicity (Table 5.2). The crystal diffracted beyond 1.2 Å but the resolution was restricted to 1.4 Å because of the orientation of the crystal along the spindle axis and time constraints of data collection. Data were collected in two resolution sweeps to ensure complete low resolution data by recollecting reflections overloaded in the high-resolution and statistics are presented in Table 5.3. Low resolution sweeps were collected for all the substrate complexes with the exception of the 2 Å Ala^{B241}–PGSO complex.

	High resolution data	Low resolution data
crystal-to-plate (mm)	200	260
Resolution range (Å)	10-1.41	25-2.7
Oscillation range per image (°)	0.5	1.0
Number of Images	216	100
Crystal mount along spindle axis	a*	
mosaicity (°)	0.3	

Table 5.2 Summary of Data collection Parameters for $Ala^{B:241}$. Crystal size 0.15 x 0.1 x 0.07 mm^3. *Crystallisation conditions: 11% (v/v) MME-PEG 2K, 20% (v/v) ethylene glycol, 56mM MOPS pH 7.2. Cryo-buffer: gradually transferred from 20-25-30% (v/v) ethylene glycol over a 5-8 minute time period.*

Data statistics	**Ala^{B241}**	**Ala^{B241}–PG**	**Ala^{B241}–PGSO**	**WT–PGSO**
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell parameters (Å)	a = 51.2	a = 51.6	a = 51.2	a = 51.2
	b = 131.7	b = 131.8	b = 131.2	b = 131.7
	c = 63.9	c = 63.9	c = 63.9	c = 64.0
	β = 105.7°	β = 105.6°	β = 105.8°	β = 105.9°
Resolution (Å)	25-1.4	30-1.45	15-2.0	30-1.8
Temperature (K)	100	100	100	100
Reflections measured	710945	1012905	248768	710896
Unique reflections	152163	144035	50974	77861
Mosaicity (°)	0.3	0.5	0.6	0.6
$\langle I \rangle / \langle \sigma(I) \rangle$	19.8 [a](4.4)	18.7 [b](3.3)	7.2 [c](1.7)	24.3 [d](4.6)
Completeness (%)	97.4 (82.5)	99.6 (99.9)	93.4 (77.8)	96.1 (71.9)
R_{merge}	5.2 (17.6)	6.1 (33.9)	8.6 (32.1)	4.8 (20.3)

Numbers in parentheses refer to the outermost resolution shell [a](1.43-1.41) [b](1.47-1.45) [c](2.03-2.0) [d](1.81-1.78)

$R_{merge} = 100 \sum_{hkl} |I_{hkl} - \langle I \rangle| / \sum_{hkl} I_{hkl}$

Table 5.3 Data collection statistics for the Ala^{B241} mutant and complexes with mutant or wild type PGA.

5.2.3 Refinement and Model building

The starting model for refinement of the mutant and the complexes was the *frz*PGA-1.8 structure with water molecules removed. The crystals for the mutant and the substrate complexes were all isomorphous with wild type. The models were refined using restrained maximum likelihood method as implemented in *REFMAC* with a bulk solvent correction and overall anisotropic scaling. 'Riding' hydrogen atoms were included using the program *HGEN*, however hydrogens were not included for substrates or ligands. During manual rebuilding, difference electron density maps with coefficients $2m|F_o| - D|F_c|$ and $m|F_o| - D|F_c|$ (output weights generated from *REFMAC*) were used to guide the fitting of the model. Initially, water molecules were added using the automated refinement program *ARP* (Lamzin and Wilson, 1993) in concert with several rounds of refinement with *REFMAC*. In the later stages, additional water molecules were added using the *X-SOLVATE* module in *QUANTA*. During model building and refinement, 3% of the data were flagged for cross-validation using the free R to monitor the progress of refinement. Refinement statistics for each structure are presented in Table 5.4.

5.2.3.1 Modelling of bound substrates and modified amino acids

REFMAC parameter files for the substrates were generated with the *CCP4* program *MAKEDICT* (Collaborative Computational Project, 1994). The parameter files for Pen G and PGSO were based on related small molecule X-ray structures deposited in the Cambridge Structural Database (CSD) (Allen *et al*., 1979), benzylpenicillin 1'-diethylcarbonate ester (Ref. Code BENPEN10) and cloxacillin sulphoxide (Ref. Code IPENSX) respectively. Planar restraints were imposed upon the phenylacetyl moiety, the amide bond, the carbonyl of the lactam ring and the carboxyl side chain of the thiazolidine ring. The parameter file for the oxidised Met A:16 was based on the related small molecule X-ray structure, carboxymethyl-L-cysteine sulphoxide (Ref. Code CMXLCS).

Refinement statistics	**Ala^{B241}**	**Ala^{B241}–PG**	**Ala^{B241}–PGSO**	**WT–PGSO**
Resolution (Å)	25 - 1.41	30 - 1.45	15 - 2.0	30 - 1.78
†R factor (%)	14.9	14.5	16.5	15.4
‡Free R factor (%)	16.9	17.1	23.2	19.0
Non-hydrogen atoms				
protein*	6252	6293	6085	6103
water	832	877	626	702
ethylene glycol (#)	52 (13)	88 (22)	–	44 (11)
ligand	–	23	24	24
R.m.s. deviations from ideality (Å)				
bond 1-2 distance (0.02)	0.011	0.019	0.017	0.018
angle 1-3 distance (0.04)	0.026	0.034	0.039	0.035
planar 1-4 distance (0.05)	0.032	0.041	0.044	0.041
Average B factor (Å^2)				
all atoms	14.4	14.8	18.2	16.9
protein	13.1	13.2	17.4	15.7
water	23.9	26.1	25.4	26.4
ligand	–	15.0	28.5	30.1
Ca^{2+} ion	7.2	7.4	14.5	10.8
ethylene glycol	17.5	19.3	–	23.6
§Coordinate error (Å)	0.054	0.056	0.185	0.110

†R-factor = $\sum_{hkl} ||F_{obs}|-k|F_{calc}|| / \sum_{hkl} |F_{obs}|$

‡Free R-factor = $\sum_{hkl \grave{I} T} ||F_{obs}|-k|F_{calc}|| / \sum_{hkl \grave{I} T} |F_{obs}|$ where $hkl \subset T$ represents 3% of the diffraction data. §Coordinate error estimated by DPI, the Diffraction-data Precision indicator (Cruickshank, 1996). *The number of protein atoms vary dependent on the number of alternate conformations. # Number of ethylene glycol molecules

Table 5.4 Refinement statistics for the Ala B:241 mutant and its substrate complexes and PGSO with wild type PGA.

5.3 Results & Discussion

5.3.1 Unliganded mutant structure

The inactive PGA mutant was produced by introducing a single point mutation at the oxyanion hole, Asn B:241 to Ala, henceforth Ala^{B241}. The oxyanion hole can be envisaged as a positively polarised binding pocket that accommodates the carbonyl oxygen of a scissile peptide bond (Henderson, 1970), placed so as to bind preferentially the transition state complex for the tetrahedral intermediate. Disruption of the oxyanion pocket the mutation Ala^{B241} results in an approximately 10,000-fold reduction in enzyme activity when assayed with NIPAB. The X-ray structure at 1.4 resolution of Ala^{B241} was determined to elucidate the primary structural consequences of the mutation and to discriminate conformational changes as a result of the mutation from those induced by substrate binding. The protein structure away from the active site closely resembles that of the wild type enzyme with an overall rmsd on CA atoms of 0.39 Å.

The initial difference map calculated with coefficients $F_o - F_c$, showed regions of both positive and negative density in the active site. The density clearly confirmed the alanine mutation at position B:241 (Figure 5.2) and also indicated conformational changes in the vicinity. The side chains of Arg A:145 and Phe A:146 occupy two alternate conformations, one of which corresponds to that in the wild type (helix conformation) and the other, which is the predominant conformation (70% occupied), is that seen in the penicillin complexes described below (loop conformation). The side chain density is weak and ill-defined for A:145 and A:146 in the helix conformation but main chain and side chain disorder of A142-A:144 confirm the presence of both conformations. The residues Arg A:145 and Phe A:146 lie at the end of a 16 residue α-helix (A:131 to A:146; helix H1) which is interrupted by a kink at position A:142 (ϕ = 116°, ψ = 65°) to A:143 which is positioned at the back of the S1 specificity pocket. The kink at position Met A:142 destabilises the end of the helix which results in two energetically favourable conformations being possible.

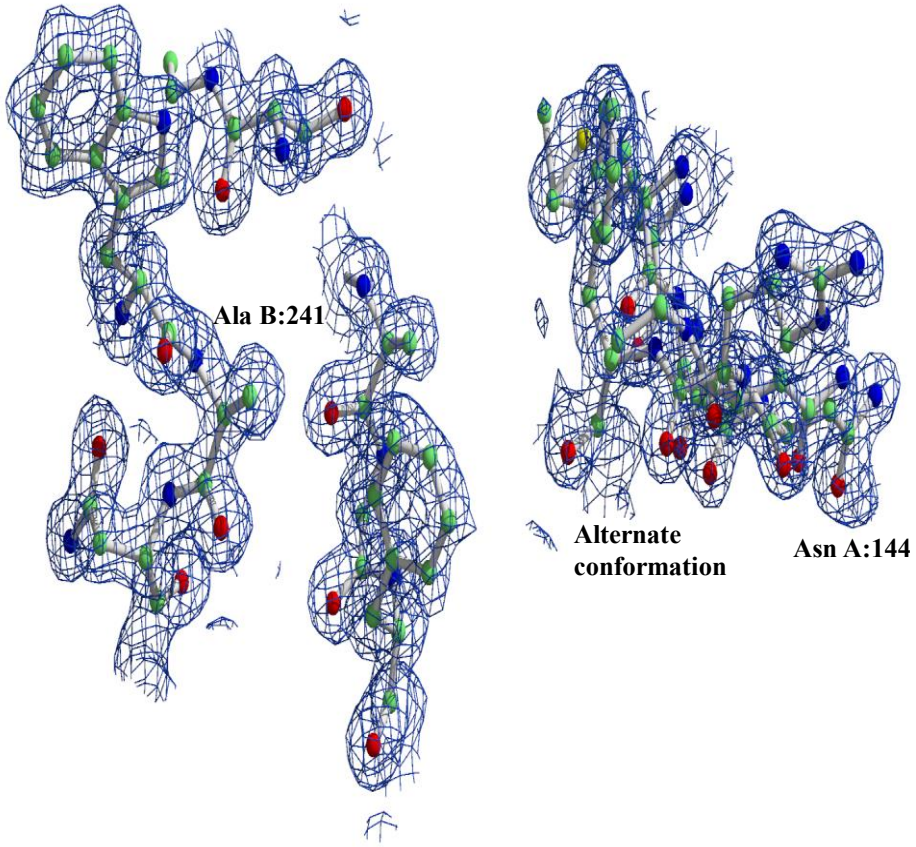


Figure 5.2 The final $2F_o - F_c$ electron density map contoured at 1.5 σ with the refined atomic positions of the active site mutant Ala^{B241}.

The density corresponding to Met A:16 showed additional positive difference density which was modelled as the first oxidation state, the sulphoxide form. This tends to be the fast step in the oxidation, with further oxidation to give the sulphone being much slower requiring stronger oxidising conditions (Buz*y et al*., 1995; Kachur*in et al*., 1995). The charged environment around the methionine, with a lysine side chain whose NZ is 3.1 Å away from the sulphur and a water at 2.7 Å may be a contributing factor to its susceptibility to oxidation, Figure 5.3.

The oxidised methionine is remote from the active site of the enzyme and is neither relevant to ligand binding or structure interpretation in the area around the active site. Met A:16 is accessible to solvent via a channel and its oxidation is presumably a consequence of the chemical environment of the protein.

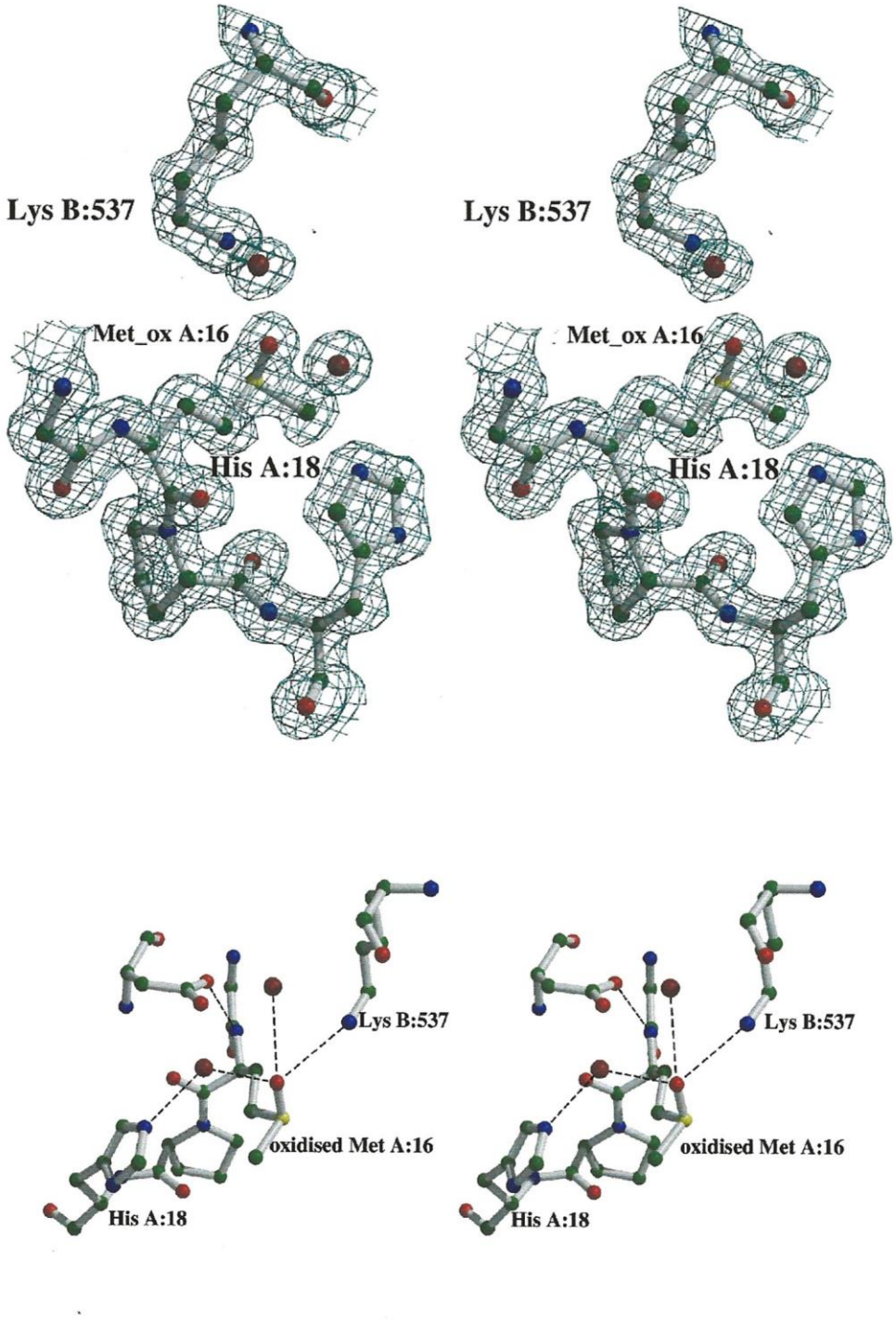


Figure 5.3 (a) The final $2F_o - F_c$ electron density map contoured at 1 σ showing the refined atomic positions of the oxidised Met A:16 and neighbouring residues, (b) an orthogonal view to (a) of the environment around Met A:16 showing the hydrogen bonding network.

5.3.2 Ala^{B241}–Pen G complex

The structure of Ala^{B241} complexed with Pen G (Ala^{B241}-PG) was determined to 1.45 Å resolution. The overall structure is similar to the wild type and mutant PGA structures, with substrate induced conformational changes restricted to the region around the active site. The overall rmsd on CA atoms between wild type and Ala^{B241}-PG is 0.21 Å, and 0.44 Å on active site residues within a 12 Å sphere around the C15 atom of Pen G, (Refer to Figure 5.6 for atom labelling), Table 5.5.

The initial electron density maps showed positive difference density in the active site pocket and the *X-LIGAND* module in *QUANTA* was used to fit the Pen G substrate (Figure 5.4). The refined structure of Pen G bound in the active site and the residues which interact with it are shown in Figure 5.5.

R.M.S deviations for the Ala^{B241}-PG complex from wild type PGA (Å)							
CA atoms			12 Å sphere of active site CA atoms			12 Å sphere of active site all equivalent atoms	
0.21			0.44			0.98	
selected R.M.S deviations per residue (Å)							
Thr A:141	Met A:142	Ala A:143	Asn A:144	Arg A:145	Phe A:146	Ser A:147	Phe B:71
1.43	0.84	0.86	1.98	5.46	2.33	0.35	1.01

Table 5.5 R.M.S deviations of the structure Ala^{B241}-PG from the structure of frzPGA-1.3.

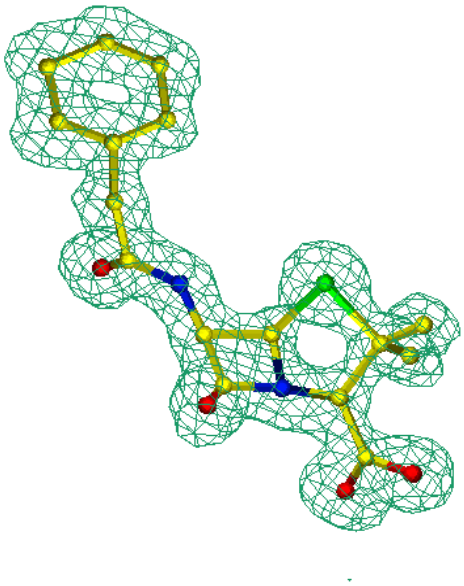


Figure 5.4 The initial $F_o - F_c$ difference density contoured at 3 σ showing the Penicillin G molecule (not in phasing).

A summary of the contacts between Pen G and Ala^{B241} is given in Table 5.5 with a schematic representation of the key interactions shown in Figure 5.6. The interactions between Ala^{B241} and Pen G are dominated by aromatic-aromatic and sulphur-aromatic contacts. The phenylacetyl side chain of the substrate sits in the well defined hydrophobic hole at the back of the active site interacting with the aromatic rings of Phe A:146, B:24, and non polar residues Ile B:177 and Met A:142. The only direct hydrogen bond is between the OG atom of Ser B:1 and the O16 atom of Pen G. The aromatic-aromatic contacts made to the phenylacetyl moiety are similar to those described previously for the phenylacetic acid complex (Duggleby *et al*., 1995). The polar groups of the penam moiety all make favourable hydrogen bonds, however none of these are direct to the protein but rather to water or ethylene glycol.

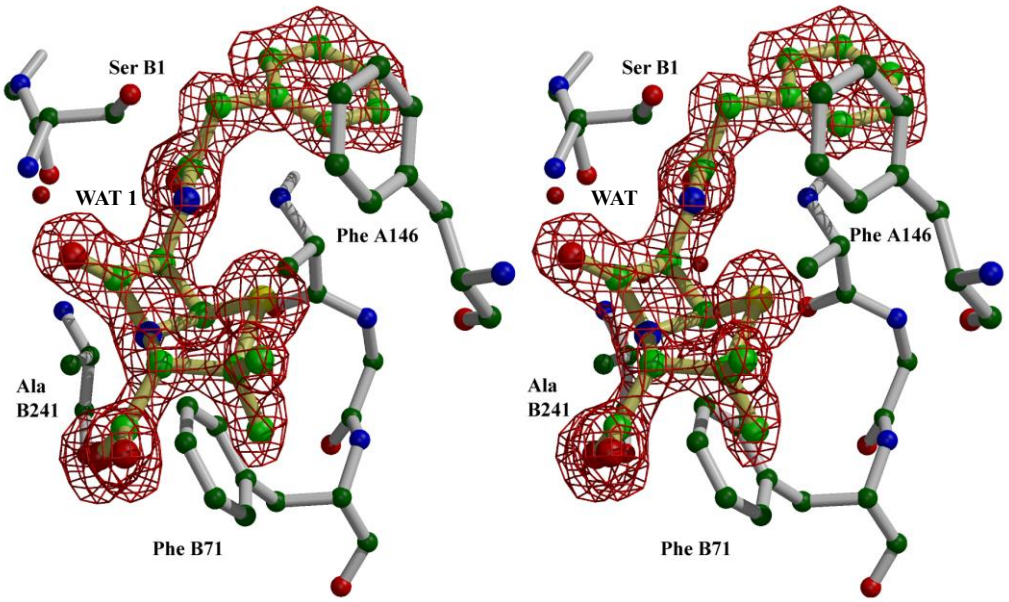


Figure 5.5 Stereoview showing the final $2F_o - F_c$ electron density map contoured at 1 σ for the refined atomic positions of the Pen G complex and the key active site residues, without contoured density, interacting with the substrate.

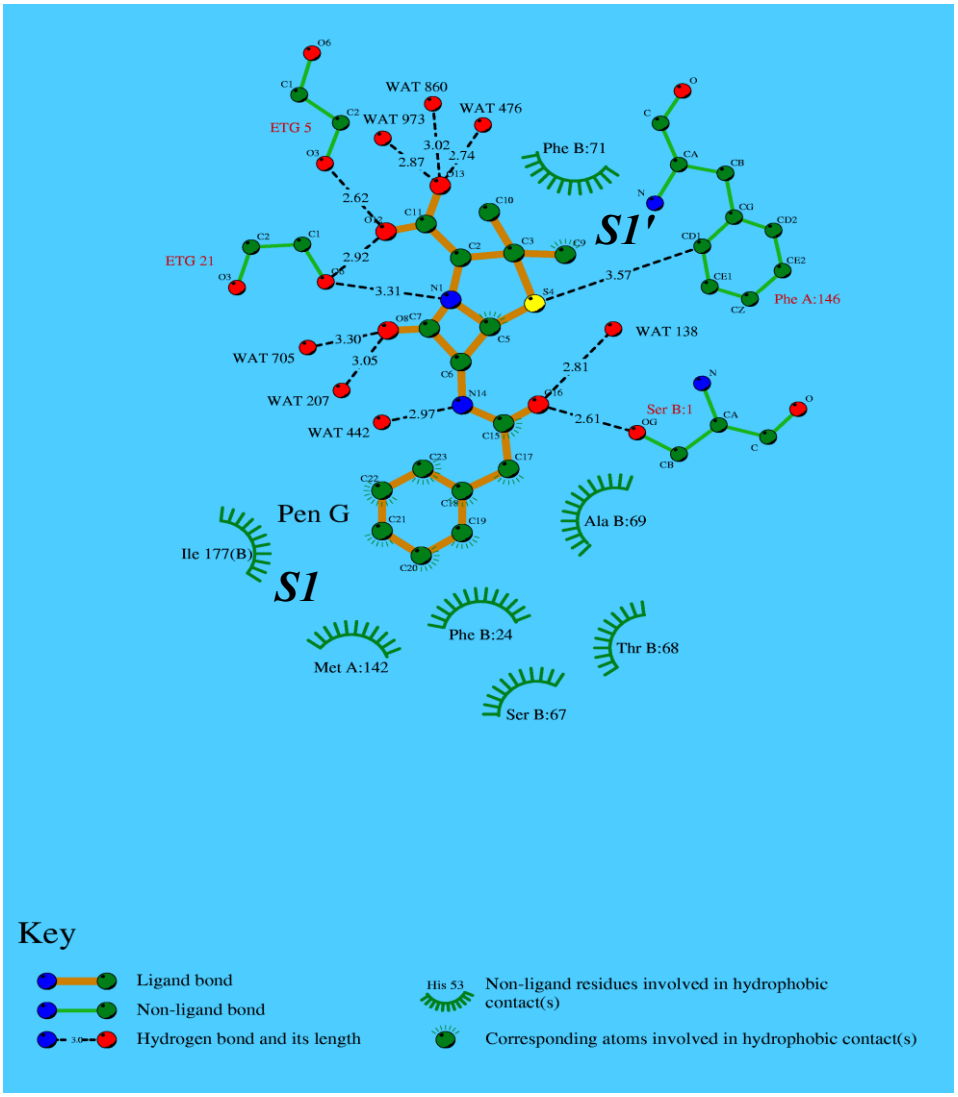


Figure 5.6 A schematic representation showing the interactions of Pen G in the active site of the inactive mutant Ala^{B241}. *The figure was prepared with the programs HBPLUS (McDonald and Thornton, 1994) and LIGPLOT (Wallace et al., 1995).*

(a) Pen G polar interactions		
Donor	Acceptor	Distance (Å)
WAT 138	Pen G O16	2.81
Pen G N14	WAT 442	2.97
Pen G O13	WAT 973	2.87
Pen G O13	WAT 860	3.02
Pen G O13	WAT 476	2.74
WAT 705	Pen G O8	3.30
WAT 207	Pen G O8	3.05
Pen G N1	Pen G O12	2.79
ETG C21 O6	Pen G O12	2.92
Pen G N1	ETG C 21 O6	3.31
ETG C5 O3	Pen G O12	2.62
SER B1 OG	Pen G O16	2.61
PHE A146 CD1	Pen G S4	3.57

(b) Pen G non polar interactions		
Atom 1	Atom 2	Distance (Å)
Pen G C15	ALA B69 CB	3.74
Pen G C17	PHE B24 CD1	3.49
Pen G C17	PHE B24 CE1	3.43
Pen G C18	PHE B24 CE1	3.34
Pen G C18	PHE B24 CZ	3.77
Pen G C19	PHE B24 CE1	3.64
Pen G C19	PHE B24 CZ	3.54
Pen G C20	SER B67 CB	3.86
Pen G C20	THR B68 C	3.82
Pen G C20	THR B68 CA	3.78
Pen G C21	MET A142 CE	3.84
Pen G C22	ILE B177 CD1	3.87
Pen G C22	MET A142 SD	3.67
Pen G C23	ALA B69 CB	3.77
Pen G C23	PHE B24 CE1	3.79
Pen G C5	PHE B71 CE2	3.64
Pen G C5	PHE B71 CZ	3.39
Pen G C9	PHE B71 CE1	3.82
Pen G C9	PHE B71 CZ	3.78

Table 5.6 A summary of the protein : Pen G interactions. *(a) The polar interactions of Pen G with the mutant Ala^{B241}. (b) The non polar interactions of Pen G with the mutant Ala^{B241}.*

Direct protein interactions with the penam moiety are limited to hydrophobic contacts. The penam nucleus points out of the active site and is sandwiched between the aromatic side chains of Phe A:146 and Phe B:71. There are contacts between the thiazolidine ring and the edge (CD1 and CE1) of Phe A:146 with the sulphur atom of the thiazolidine ring 3.6 Å from the CD1 atom of Phe A:146. These contacts may be weakly electrostatic interactions between the electronegative δ^- sulphur atom and the δ^+ hydrogens atoms on the aromatic ring edge (Burley and Petsko, 1988). The side chain of Phe B:71 is positioned under the base of the thiazolidine ring of the penam moiety. The C5 atom and C9 atom of the penam side chain are 3.4 Å and 3.8 Å from the CZ atom of Phe B:71 respectively.

Close inspection of additional positive and negative difference density around the ligand, helped by the high resolution, showed that binding of Pen G induced a conformational change starting at phenylalanine A:138 and culminating in a large side chain movement of Arg A:145 and Phe A:146. The residues are part of helix H1 in which Met A:142 is positioned at a kink (ϕ -116°, ψ -65°) in the helix of the wild type structure. On binding of Pen G the C-terminus of the helix unwinds into a coil with torsion angles for Met A:142 of (ϕ -54°, ψ -47°). Residues A:142 - A:146 are totally in the loop conformation in this complex, quite different to the situation in the wild type.

Analysis of the individual residues (A:138 to A:146) involved in the conformational change in the helical region shows that it is almost entirely dominated by the side chain movements of Arg A:145 and Phe A:146 (Table 5.5). The rmsd of the Arg A:145 and Phe A:146 residues from their equivalent in wild type is 5.5 Å and 2.3 Å respectively. The side chain of A:146 compared with its position in the wild type structure moves 3.3 Å with the main chain CA moving ~1 Å when compared to wild type PGA (Figure 5.7). In the native structure the guanidinium group of Arg A:145 is sandwiched between the δ^- π electron cloud of both Phe A:146 and Phe B:459. The CA atom of Arg A:145 moves 2 Å in tandem with A:146 and A:145 and adopts a position where the side chain is solvent exposed into the active site.

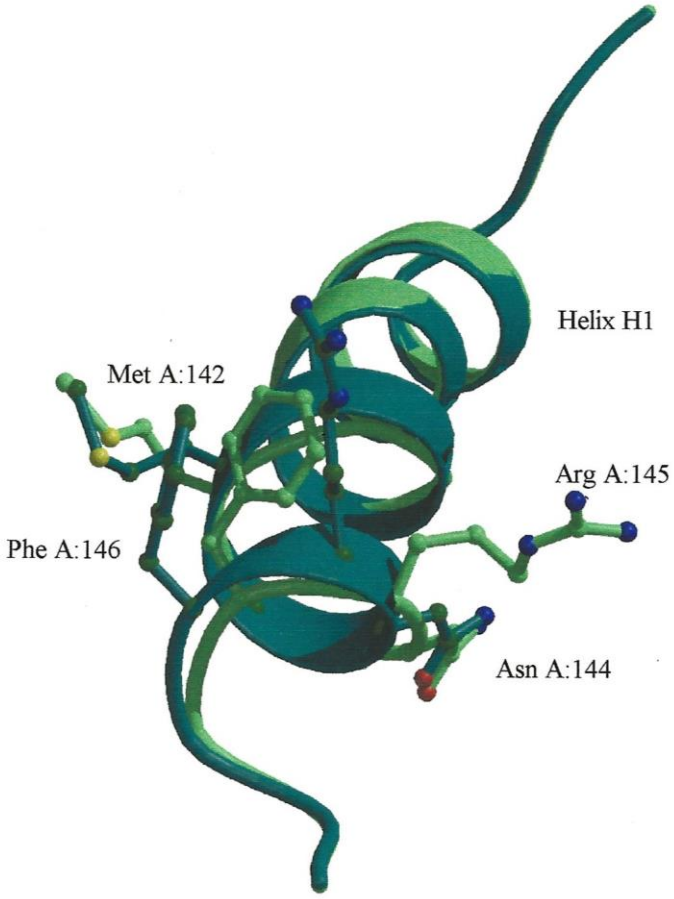


Figure 5.7 Overlay of the two conformations of helix 1, H1. *The helix conformation of wild type PGA is shown in slate blue while the Ala^{B241}-PG helix-loop conformation is shown in green.*

The position of the phenylacetyl side chain is close to, but not identical to the position of the phenyl acetic acid product in the triclinic structure (PDB accession code 1PNL, Duggleby *et al*., 1995). The change in conformation of helix H1 leads to a loss in van der Waals contacts between the phenylacetyl group of Pen G and the side chain of Phe A:146. In comparison with the PAA complex, the phenylacetyl moiety of the Pen G substrate with the Ala^{B241}-PGA mutant does not sit as tightly in the pocket and is slightly shifted out of the hydrophobic pocket, Figure 5.8. Phe A:146 forms the neck of the hydrophobic pocket that accommodates the phenylacetyl moiety. The substrate shift

can be correlated with widening of the neck of the pocket through the movement of A:146 and the unwinding of the C-terminus of the helix (Figure 5.7). The phenylacetyl group moves laterally which coincides with the side chain movement of A:146. In comparison with the PAA ligand this lateral movement results in a 1 Å shift of the carbonyl C15 atom. The greater side chain shift of Phe A:146 destroys the close aromatic-aromatic interaction between the phenyl groups.

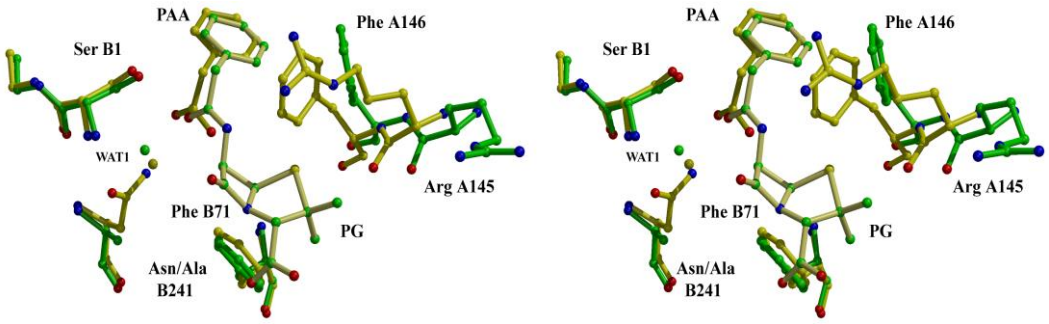


Figure 5.8 Superimposition of the active site of the PAA ligand structure (yellow) with the structure of Pen G complex (green).

The active site nucleophilic serine OG lies 3.3 Å from the carbonyl carbon (atom C15) of Pen G. This is an unfavourable distance for catalysis, about 0.5 Å greater than expected from the phenylacetic acid complex. This results from the loss of the side chain interaction of the AsnB:241Ala mutation, with the much smaller side chain at this position losing its hydrogen bonding capability. The result is that the Pen G substrate shifts out of the active site pocket. Interestingly, the proposed catalytic water (WAT1) makes a hydrogen bond (2.9 Å) to the α-NH_2 of Ser B:1 as seen in the PAA complex.

The electron density for the penam side chain of the penicillin substrate is well defined in the S1' subsite pocket. The interactions are perhaps unexpected in the S1' subsite considering how open this part of the binding pocket appears to be (Figure 5.5) and that the substrate specificity is not governed by the penam moiety, implying that the interactions are non-specific with little substrate preference. Part of the explanation may be that the crystal was frozen at 100 K. A more likely explanation is that the limited conformational flexibility of Pen G restricts the position of the penam moiety

once the phenylacetyl group is buried in the S1 pocket. The clear definition of the electron density for the penam nucleus is also a reflection of the contacts made by the alternate conformation adopted by A:146 and the synergy of binding through the aromatic interactions of A:146 and B:71.

Superimposition of the crystal structure of free penicillin G on the molecule bound to the Ala^{B241} mutant shows a distinct conformational change in the complex (Figure 5.9). The position of the phenylacetyl moiety is bent to one side when bound in the S1 specificity pocket. Overall the penam moiety remains rigid with the exception of the thiazolidine ring which adopts a different pucker conformation when bound in the complex in contrast to a much flatter ring in the small molecule structure. This can be explained partly by the clasping of the penam moiety by the aromatic residues A:146 and B:71 and also by the sulphur-aromatic interactions which are formed in preference with Phe A:146 rather than with Phe B:71.

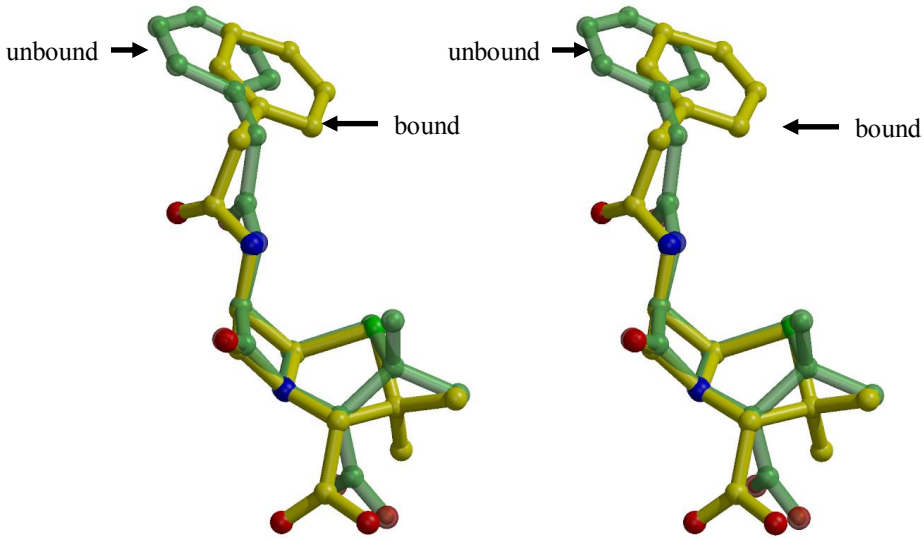


Figure 5.9 Superimposition on all atoms of the substrate Penicillin G in its bound conformation (yellow) with that of the small molecule crystal structure of unbound Penicillin G (green).

5.3.3 Ala^{B241}–PGSO complex

The structure of the Ala^{B241}–PGSO complex was determined to 2 Å resolution and is similar to the Pen G complex in its overall structure and active site conformation. The initial difference maps showed positive density representative of bound PGSO. The initial starting model for PGSO was the small molecule crystal structure IPENSX, edited using the Molecular Editor option in *QUANTA* to remove the additional side chain modifications from the carboxylic acid group. Because of the ambiguity in assigning the relative positions of the sulphur and the oxygen on the thiazolidine ring of the PGSO and to ensure that the substrate was the oxidised form of Pen G, the composition of the compound was verified by mass spectrometry and NMR (see Appendix). The mass spectrometry results confirmed the molecular mass corresponding to the sodium salt of PGSO. The ^{13}C NMR analysis confirmed that the substrate was intact and not degraded. The mass spectrometry analysis was carried out by Dr. Trevor Dransfield and the NMR analysis by Dr. Paul Murphy.

The PGSO is in a similar position to Pen G in the active site differing only by subtle changes in conformation and in protein contacts. The protein-substrate interactions and distances are illustrated in Figure 5.10, 5.12 and listed in Table 5.7. The position of the phenylacetyl side chain is identical to that in the Ala^{B241}–PG complex, and the B:1 Ser OG lies 3.3 Å from the carbonyl carbon (C16) of the substrate. The carbonyl oxygen (O17) makes one direct hydrogen bond of 2.5 Å to protein with B:1 Ser OG and a second hydrogen bond to a water molecule (WAT149).

The penam side chain of PGSO sits between A:146 and B:71 (Figure 5.10) and again the interactions are confined mainly to hydrophobic contacts. Similarly, the oxidised form of the penam side chain has each of its polar groups satisfied by hydrogen bonds to water molecules. In contrast to the Ala^{B241}–PG complex, the side chain of Phe A:146 now forms an electrostatic interaction with the O5 atom of PGSO where the δ^+ charge of the edge of the aromatic ring is positioned to interact with the electronegative oxygen. The distance between the CD1 atom of A:146 and the O5 of PGSO is 3.2 Å, closer than the equivalent sulphur interaction in the Pen G complex. The base of the penam moiety is in van der Waals contact with the side chain of Phe B:71. The

aromatic ring now makes closer contact with the β-lactam of PGSO with the C6 and C7 atoms a distance of 3.2 Å and 3.9 Å respectively from the CZ atom of B:71.

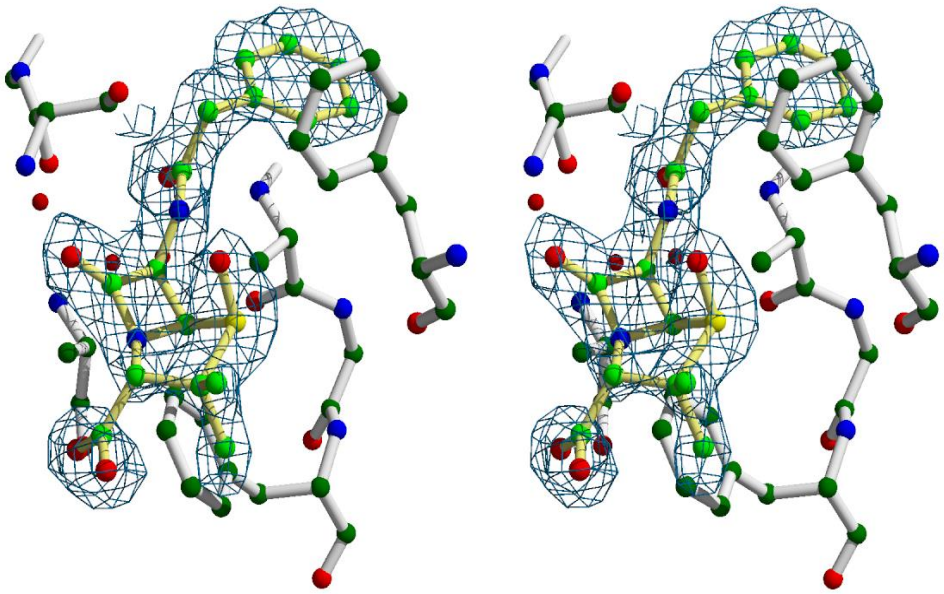


Figure 5.10 Stereoview showing the final $2F_o - F_c$ electron density map contoured at 1 σ and the refined atomic positions of the PGSO complex in the active site of Ala^{B241} and surrounds the PGSO substrate only with the active site residues not contoured for clarity.

(a) Electrostatic interactions for PGSO		
Donor	Acceptor	Distance (Å)
WAT W149	PGSO O17	2.62
WAT W601	PGSO O14	3.26
PGSO O13	WAT W297	2.98
WAT W225	PGSO O9	2.70
WAT W154	PGSO O9	2.96
PHE A146 CD1	PGSO O5	3.17
SER B1 OG	PGSO O17	2.53

(b) Hydrophobic interactions for PGSO		
Atom 1	Atom 2	Distance (Å)
PGSO C23	ILE B177 CD1	3.72
PGSO C7	PHE B71 CZ	3.89
PGSO C6	PHE B71 CZ	3.15
PGSO C7	PHE B71 CE2	3.87
PGSO C6	PHE B71 CE2	3.55
PGSO C10	PHE B71 CE1	3.87
PGSO C24	ALA B69 CB	3.88
PGSO C16	ALA B69 CB	3.58
PGSO C22	THR B68 C	3.88
PGSO C21	THR B68 C	3.83
PGSO C21	THR B68 CA	3.78
PGSO C21	PHE B24 CZ	3.76
PGSO C20	PHE B24 CZ	3.43
PGSO C19	PHE B24 CZ	3.71
PGSO C24	PHE B24 CE1	3.69
PGSO C20	PHE B24 CE1	3.55
PGSO C19	PHE B24 CE1	3.28
PGSO C18	PHE B24 CE1	3.42
PGSO C19	PHE B24 CD1	3.88
PGSO C18	PHE B24 CD1	3.43
PGSO C23	MET A142 SD	3.67
PGSO O5	PHE A 146 CE1	3.58

Table 5.7 A summary of the protein interactions with PGSO. *(a) Polar interactions of PGSO in the active site of* Ala^{B241}*. (b) The non polar interactions of PGSO with the* Ala^{B241} *mutant.*

The conformational change observed in the Pen G complex from A:142 to A:146 culminating in the downward movement of the side chains of Arg A:145 and Phe A:146 is also seen in the PGSO complex. However, the presence of an oxygen atom on the sulphur of the thiazolidine ring causes a movement of the aromatic rings of A:146 and B:71 compared to their position in the Pen G complex. The side chain of A:146 is rotated by -15° due to its contact with the O5 atom of PGSO (Figure 5.11).

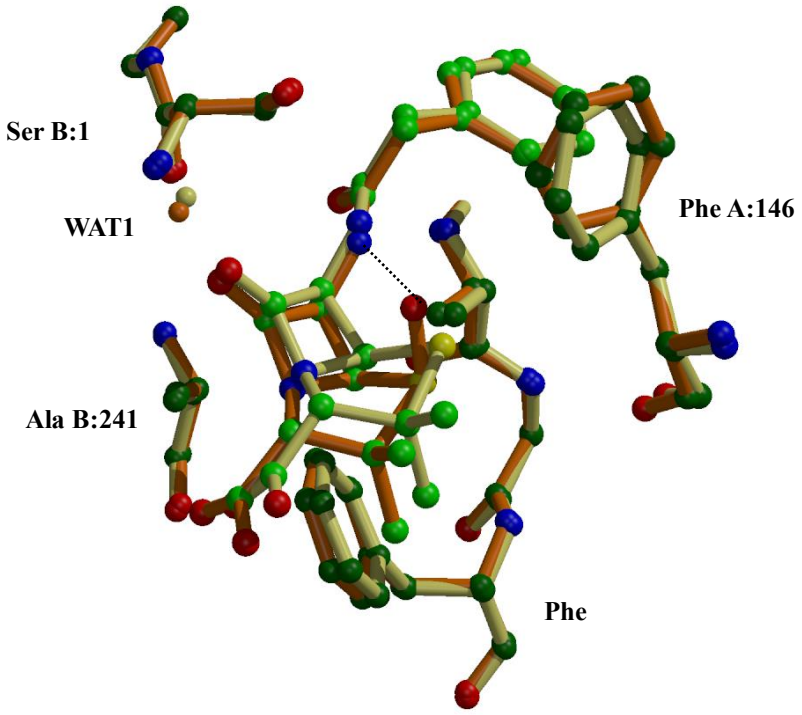


Figure 5.11 Superimposition of Ala^{B241}–PGSO (orange) with Ala^{B241}–PG (cream) showing the differences in binding of both substrates. *The dotted line represents the intramolecular hydrogen bond of PGSO.*

The side chain of B:71 also moves further back from the active site resulting in closer contact with the penam moiety which may explain the improved affinity of PGSO over Pen G (Figure 5.11, LIGPLOT schematic). The K_M of Pen G determined from semi-purified extracts of PGA from *E. coli* is 20 μM and the K_i of PGSO is 12 μM (Alvaro *et al*., 1991).

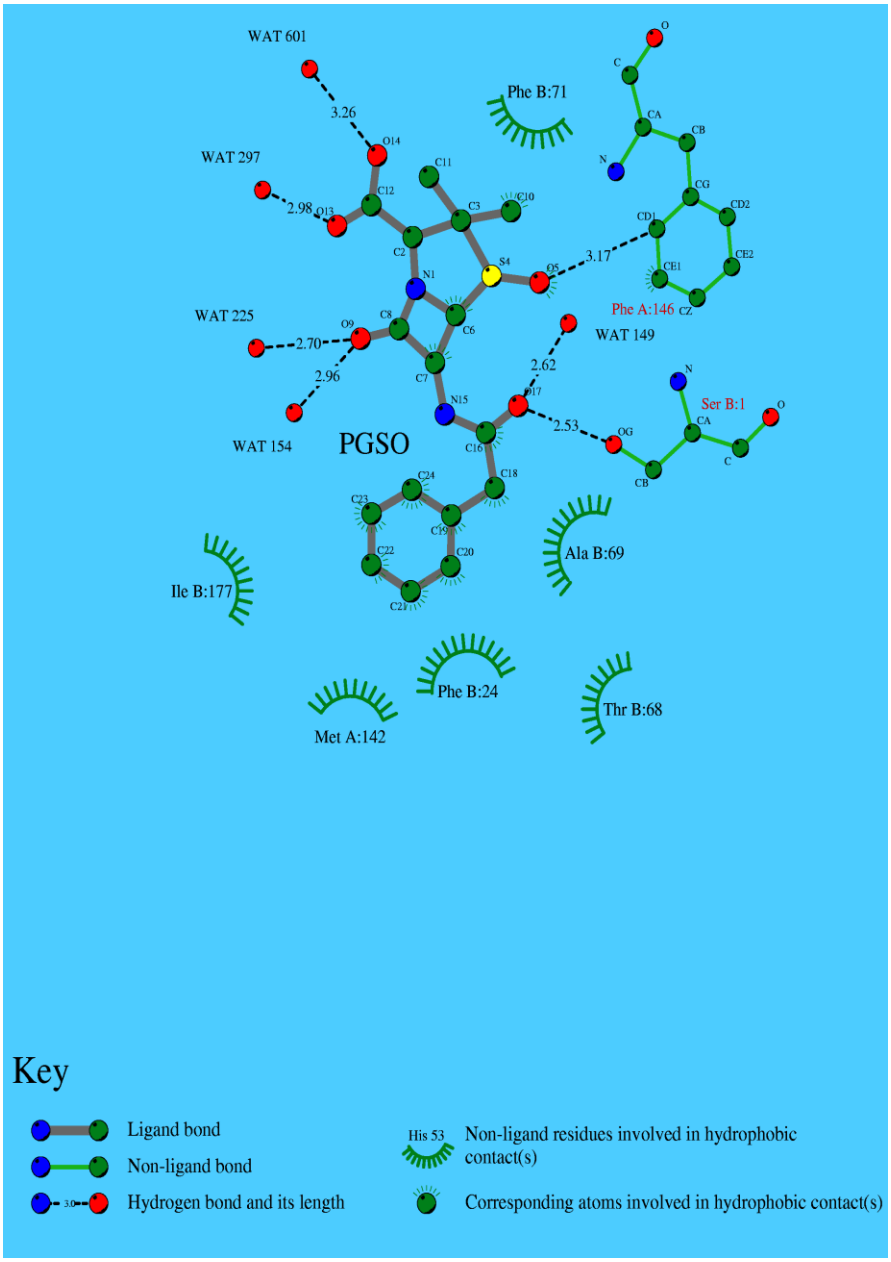


Figure 5.12 Schematic representation showing the interactions of PGSO with the Ala^{B241} mutant.

5.3.4 Wild type PGSO complex

The structure of wild type PGA complexed with PGSO (WT–PGSO) has been determined to 1.8 Å resolution. The initial difference map around the active site was difficult to interpret and assignment of the position of PGSO was less straightforward than for the previous complexes. There was clearly defined difference density for the phenylacetyl moiety but the desnity was much less clear for the remainder of the substrate. The final density map is shown in Figure 5.13.

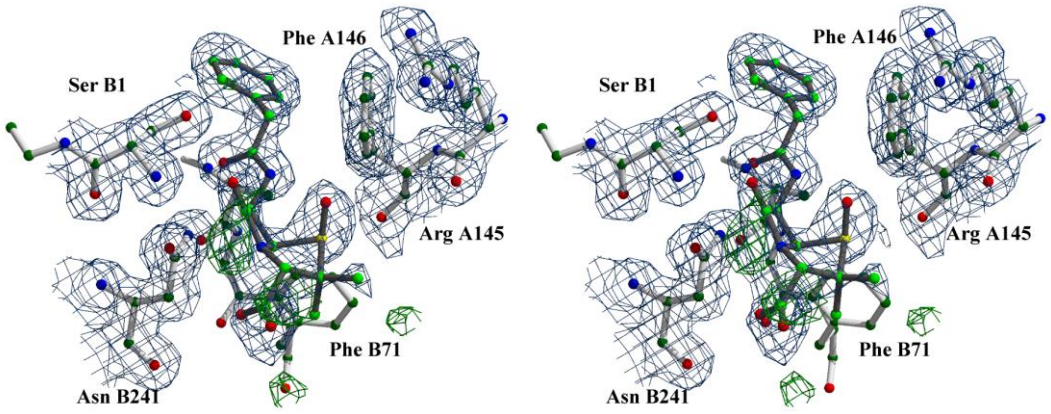


Figure 5.13 The electron density map of PGSO bound in the active site of wt PGA. *The $2F_o - F_c$ map coloured blue is contoured at 1 σ and the $F_o - F_c$ map coloured green is contoured at 3 σ and surround the active site residues.*

The average B values ~ 15 Å^2 for the phenylacetyl supports its tight binding and essentially full occupancy. The phenylacetyl side chain is shifted 0.5 Å back into the pocket in comparison to the Ala^{B241}–PGSO structure and is in a similar position to that of the phenylacetic acid (PAA) ligand in the triclinic complex (1PNL.pdb). This means that the phenylacetyl side chain it the present complex makes closer contacts with the hydrophobic pocket than in the two Ala^{B241} complex, more similar to those seen with

PAA. However, the -CH_2 group (atom C17) is displaced from its position in the PAA complex (Figure 5.14). In contrast to the Ala^{B241}–PGSO complex the Ser OG no longer makes a hydrogen bond to the O17 atom of PGSO, instead as expected the ND2 atom of Asn B:241 makes a direct hydrogen bond to the carbonyl oxygen.

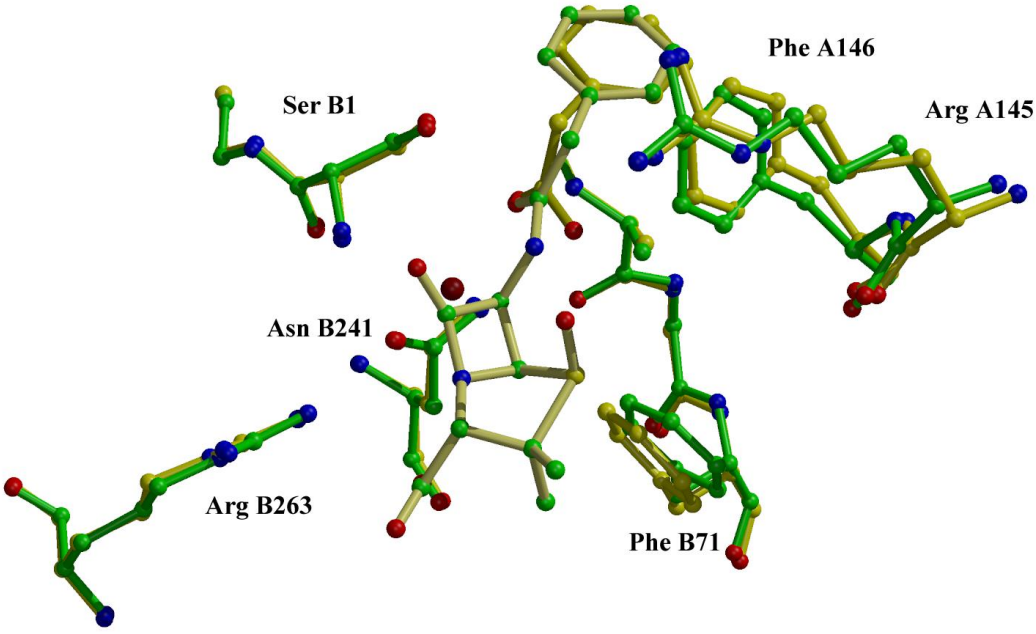


Figure 5.14 Superimposition of WT-PGSO with WT-PAA (1PNL.pdb).

The quality of the density corresponding to the penam nucleus is much poorer than in the mutant PGSO complex, (Figure 5.13), and modelling of the substrate in this region is more speculative. The B values for the penam moiety (average of 40 $Å^2$) are much higher than for its phenylacetyl side chain and for the surrounding protein atoms (16 $Å^2$). There is continuous residual difference density from the well-defined phenylacetyl electron density to the weak and confusing density of the penam moiety.

The relatively weak density of the penam moiety suggests either the presence of more than one conformation for the penam, or low occupancy due to partial hydrolysis (Figure 5.13). PGSO is a slowly hydrolysed substrate and this could give rise to a mixed

species of PGSO and the product 6-amino-4-oxypenicillanic acid during preparation of the complex. Alternatively there could have been some modification of the substrate at the lactam ring, but this seems much less likely.

The density for the side chain of Phe B:71 is significant but rather broadened from that expected for similar side chains in the rest of the structure, which could reflect the presence of two overlapping alternate conformations, but these were not modelled at the present resolution. Two conformations for B:71 would be in keeping with both the two conformations and the partial hydrolysis hypotheses. The conformation and location of the penam ring therefore remains imperfectly defined and the following discussion is somewhat speculative. This needs further work to resolve the present ambiguity.

The lateral movement of Phe B:71 in the active site compared to the two mutant complexes suggests that the specific interactions have changed for the WT–PGSO. B:71 now makes direct contacts with the face of the thiazolidine ring in which the CZ and CE2 atoms of B:71 are 3.4 Å from the sulphur atom of PGSO, much closer than observed in the two inactive mutant complexes where Phe B:71 makes non polar interactions with the base of the thiazolidine ring.

In the WT-PGSO complex the Ser OG is 2.9 Å from the carbonyl carbon of the substrate analogue and ideally placed for nucleophilic attack and the carbonyl oxygen (O17) makes one direct hydrogen bond of 2.5 Å to B:1 Ser OG. Unlike the inactive mutant complexes, the position of Arg A:145 and Phe A:146 in the wild type PGSO complex are the same as in the uncomplexed wild type enzyme, suggesting a dual role for Asn B:241. The first is to position the substrate correctly in the hydrophobic pocket via hydrogen bonding of ND2 of B:241 and the carbonyl oxygen of the substrate, and the second is to stabilise the tetrahedral intermediate during catalysis. The difference in binding mode of PGSO to the wild-type enzyme compared to the inactive mutant, reinforces the important role of Asn B:241 during catalysis. The binding mode to the wild type should more closely resemble true substrate binding. A firm explanation as to why PGSO is a poor substrate is less clear.

The absence of water molecules and in particular the proposed catalytic water in the active site of the WT-PGSO complex raises two separate issues concerning the catalytic mechanism and the poor hydrolysis of this substrate by PGA. It is possible that poor

hydrolysis of PGSO by PGA is attributed in part to displacement of the 'catalytic' water by the lactam oxygen. This could result from a shift in conformation of the penam moiety arising from the interaction of the sulphonyl oxygen with the side chain of Phe A:146. It is also possible that when Pen G is bound, the proposed 'catalytic' water is still displaced by the lactam moiety and the serine's nucleophilic character is enhanced by other means, possibly by direct interaction with the α-amino group.

Another significant observation is the presence of an intramolecular hydrogen bond between the amide nitrogen and sulphonyl oxygen of PGSO (Figure 5.15). The obvious consequence of the hydrogen bond would be that the substrate is held more rigidly with the penam and amide tied together, restricting movement of the amide bond and hindering the conformational change necessary. This could have a bearing on the poor hydrolysis of this substrate and the mechanism of catalysis, possibly by reducing the ability of the amide to be a good leaving group.

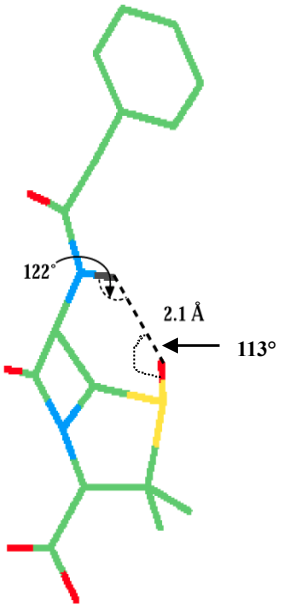


Figure 5.15 The structure of PGSO showing the intramolecular hydrogen bond between the amide NH (hydrogen drawn in grey) and the sulphonyl oxygen of the thiazolidine ring.

5.3.5 Comparison of the three PGA complexes

By comparing the above structures, a better understanding of the catalytic mechanism and the enzyme specificity is obtained. Penicillin G is one of the preferred substrates for PGA, whilst Penicillin G sulphoxide is a much poorer substrate. Comparison of Pen G and PGSO binding with the inactive Ala^{B241} mutant illustrates the differences in their conformation and contacts. Unfortunately the comparison is in the context of an inactive enzyme in which the substrate is displaced 1 Å out the pocket and any conclusions drawn from substrate binding in the mutant must remain tentative. The interaction of Phe A:146 with both substrates affects the position of the substrates in the active site and the extent of their interactions with other active site residues. The only discriminating feature which distinguishes the good substrate (Pen G) from the substrate analogue (PGSO) is the oxidation of the sulphur of the thiazolidine to give the sulphoxide form. This shows that Phe A:146 is the key to substrate specificity. The complex of PGSO with the mutant results in a shift of the penam moiety to accommodate a closer electrostatic interaction to that observed in the Pen G complex (Figure 5.10, overlap of Pen G/PGSO complexes).

In Figure 5.16, an overlay of all three complexes is shown and the importance of the oxyanion hole residue Asn B:241 is immediately apparent. This residue helps stabilise the tetrahedral intermediate and initially fixes the substrate in the S1 specificity pocket, providing the optimal stereochemistry for catalysis. Loss of the Asn side chain results in ~1 Å movement of the phenylacetyl moiety out of the S1 specificity pocket as evident from the inactive mutant complexes. In both mutant complexes A:145 and A:146 move to accommodate the phenylacetyl moiety resulting in the α helix unwinding, explaining the inactivity of the mutant. Addition of an oxygen atom on the sulphur of the thiazolidine ring to give the PGSO analogue influences the primary interactions between the aromatic rings of A:146 and B:71 and the substrate analogue. The important interactions and their respective distances are shown in Table 5.8 using the nomenclature outlined in the substrate interaction schematics, Figure 5.6 & Figure 5.12.

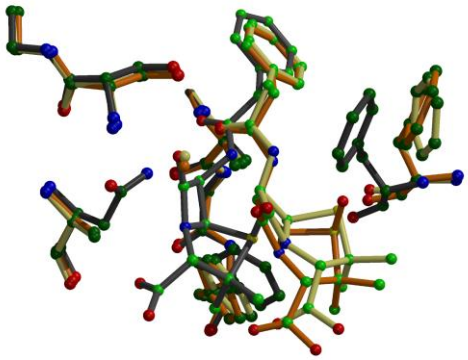


Figure 5.16 Superimposition of all three substrate complexes bound to penicillin acylase. *WT–PGSO (grey), Ala^{B241}–PGSO (orange), and Ala^{B241}–PG (cream).*

The proposed 'catalytic' water is present in the two mutant complexes, but rather disturbingly absent from the wild type PGO structure. Overlap of the three complexes reveals that the catalytic water may be displaced by the carbonyl oxygen of the lactam ring. This raises several important questions about the role of this water molecule during catalysis. If the displacement does indeed occur in the native enzyme with Pen G then it follows that the water does not have a role in the nucleophilic attack and the deprotonation of the Ser B:1 hydroxyl. In contrast, if the Pen G complex with the native enzyme has different contacts to those with PGSO and the water remains next to the B:1 Ser OH, then the origin of PGSO's resistance to hydrolysis could be associated with the displacement of water from the active site by PGSO.

What is important is that PGSO is a poor substrate for PGA. Apart from the presence of the oxygen, the only obvious difference between Pen G and PGSO is an intramolecular hydrogen-bond between the amide NH and the sulphur-oxygen of PGSO which might influence the chemistry and electrostatics of the reaction. An overlay of the complexes illustrating three stages along the reaction pathway of PGA is shown in Figure 5.17.

The complexes represent snapshots of the three fundamental stages of the reaction pathway, the Michaelis-Menten complex is represented by the PGSO complex, the tetrahedral intermediate mimicked by the benzylboronate covalently linked to Ser B:1 and the reaction product represented by the PAA complex.

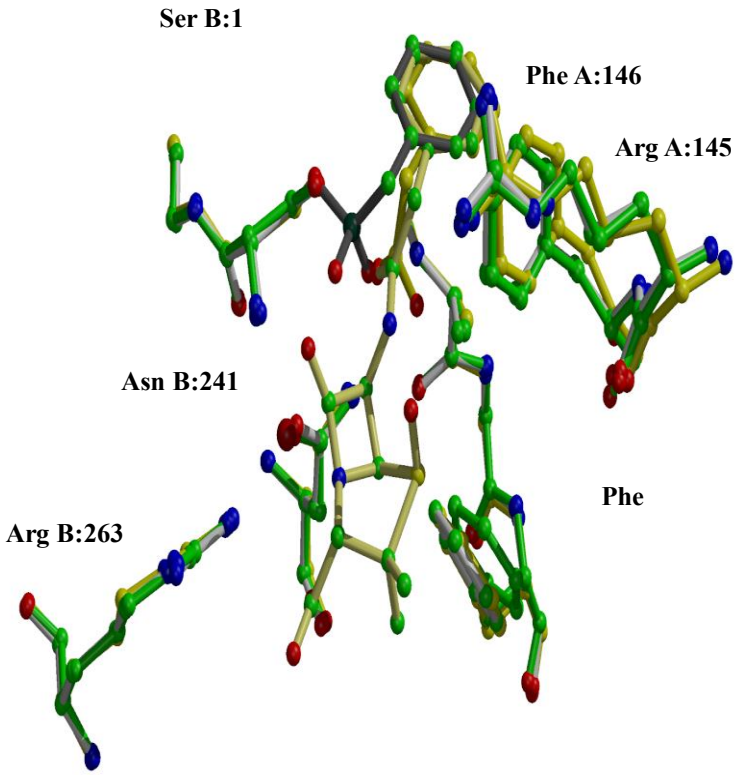


Figure 5.17 Superimposition of the WT-PGSO complex with the PAA and benzylboronate complexes. *The WT-PGSO complex ball-and-stick is coloured green with the PGSO molecule coloured cream. The benzylboronate complex is coloured grey with white ball-and-stick and the PAA complex is coloured yellow.*

Selected distances (Å) between ligand and active site residues				
Ligand atom	**Protein atom**	**Ala^{B241}–PG**	**Ala^{B241}–PGSO**	**WT–PGSO**
O16/O17*	OG Ser B:1	2.6	2.5	2.6
S4	CD1 Phe A:146	3.6	–	–
S4	CE2 Phe B:71	3.4	3.2	3.4
C5	CZ Phe B:71	3.4	–	3.4
C6*	CZ Phe B:71	–	3.2	–
C6*	CE2 Phe B:71	–	–	3.1
C15/C16*	OG Ser B:1	3.3	3.3	2.9
O5*	CD1/CE1 A:146	–	3.2/3.6	3.7

*PGSO complexes only

Table 5.8 A summary of the important interactions and their distances in the substrate complexes.

5.3.6 Active site configuration by calcium binding

The interactions in the specificity pocket of PGA are governed by A:146 and B:71 from which the A and B chains extend to form the calcium binding site (A:152, B:73, B:75 and B:76). The close proximity of the calcium binding site may have both a stabilising function and structural role in relation to catalysis. The calcium site ties down the active site and in doing so, helps configure it. The environment of the calcium binding site remains constant irrespective of conformational change in the close vicinity of the active site on binding of substrate (Table 5.8). The rigidity of the calcium site during for example, the conformational change at A:142 to A:146, is probably because the movement is dissipated by the time it reaches the calcium ligand Glu A:152.

Rmsd of the calcium binding sites from that of wild type PGA (Å)					
	PGA	WT-PGSO	Ala^{B241}-PGSO	Ala^{B241}-PG	Ala^{B241}
Main chain	–	0.1	0.1	0.1	0.1
side chain	–	0.1	0.2	0.1	0.1

Table 5.9 A summary of the superimposition of the calcium binding sites in PGA showing the rmsd of main chain and side chain atoms of the residues involved in binding.

5.4 Conclusions

Ligand binding in wild-type and the inactive mutant Ala^{B241} provide insight into substrate recognition and the role of active site residues in catalysis. The aromatic residues within the active site are the main contributors to substrate recognition and binding. Phe A:146 forms part of both the S1 and S1' substrate subsites and is the hinge of the substrate recognition machinery.

In the binding of Pen G and PGSO to Ala^{B241}, the expected H-bond to Asn B241 in the oxyanion hole is lost, and the ligands take a position further out of the S1 pocket which involves movement of A:145 and A:146 to make new contacts. Asn B:241 as well as forming part of the oxyanion hole has a role in holding the phenylacetyl moiety in the hydrophobic pocket presenting the amide bond for catalysis. The inactivity of Ala^{B241} can be attributed to both these properties of Asn B:241. Mutation of Asn B:241 to leucine should be tried in the future to probe further the importance of hydrogen bonding at this position since it is more isosteric than alanine but still has lost the capability to form side chain hydrogen bonds.

The alternate conformation of the active site helix (H1) observed upon binding of Pen G and PGSO to the Ala^{B241} mutant is identical to that observed by Done (Done *et al*., 1998), using a subset of substituted PAA derivatives bound to the wild type enzyme. Bulkier substituents on the phenyl ring of PAA caused displacement of the ligand from

the S1 pocket which resulted in an alternate conformation of the C-terminus of helix H1 and consequent loss of hydrogen bonds with either Ala B:69 or Asn B:241 and the acid group of PAA. However the opposite situation is observed in the Ala^{B241} complexes where loss of hydrogen bonding to the carbonyl of the amide bond through mutation of the oxyanion residue Asn B:241 Ala results in substrate displacement and the concomitant helix-loop transition.

One might also suggest that Phe A:146 has a dual role in determining specificity interacting with both acyl and amine moieties. Binding of PGSO to the wild type enzyme highlights the influence of A:146 in substrate recognition where the addition of an oxygen modifies the key interactions and disrupts the precise geometry around the amide bond resulting in a poor substrate for the enzyme. What is most notable in the WT-PGSO complex is the water structure on the serine B:1 side of the substrate analogue and in particular the absence of the proposed catalytic water. Overlap of the substrate complexes demonstrates that the position of the catalytic water in the inactive substrate complexes superimposes perfectly with the modelled position of the carbonyl oxygen of the lactam ring. Although speculative, one might suggest that PGSO behaves as a poor substrate because it disrupts the water environment surrounding the nucleophile. Taking this hypothesis a step further one could also suggest that disruption of the water structure by the lactam carbonyl oxygen might likewise occur when binding the preferred substrate penicillin G.

The research described above raises three hypotheses for the poor hydrolysis of PGSO

- steric clashes with the sulphonyl oxygen
- the presence of an intramolecular hydrogen bond
- displacement of the catalytic water

In the inactive enzyme complexes the thiazolidine ring is positioned between the side chains of A:146 and B:71. In the WT–PGSO complex the aromatic ring probably interacts with the sulphur (S4) from a frontal position (~3.2 Å) rather than being positioned under the penam moiety. The role of Phe B:71 as evident from the three complexes is to guide the penam side-chain into position for catalysis. Its somewhat flexible nature probably contributes to the broad substrate specificity shown by the enzyme. The opening of the active site cleft is another contributing factor to the broad

specificity. The determinants of substrate specificity, however, are known to correlate not only with the rigid structural features of the active site, but also with an inherent capacity for structural flexibility, both of the helical region from A:142 to A:146 and also the observed synergism between the aromatic residues A:146 and B:71.

Chapter Six Altered specificity of PGA

6.1 Introduction

PGA has a marked substrate specificity for hydrophobic acyl groups, in particular phenylacetyl, at the S1 subsite in which the pocket is lined with non polar residues (see Figure 4.2). It can tolerate, however, a variety of substituents on P1', including amino acids, without loss of catalytic activity. Modifications of the substrate specificity of PGA might expand its use not only in the semi-synthesis of new β-lactam antibiotics, e.g. cephalosporins, but in other synthetic and hydrolytic reactions. There are nine naturally occurring enzymes showing weak cephalosporin acylase activity from *Pseudomonas* (Lee and Park, 1998), but the activity is too low for industrial application. To date, enzyme screening has not revealed an enzyme capable of efficiently hydrolysing the α-aminoadipyl side chain of cephalosporin C which would provide a one-step process for production of a range of semi-synthetic cephalosporins which remains an ongoing objective in the pharmaceutical industry. Hence currently, the production of semi-synthetic cephalosporins involves a two enzyme process, utilising D-aminoacid oxidase and glutaryl 7-ACA acylase (Alfan*i et al*., 1997; Bruggin*k et al*., 1998).

There are several approaches to selection of an enzyme with modifed properties, such as substrate specificity or improved stability, which can be broadly classified as follows. (1) The identification of naturally occurring variants whose properties can be exploited directly. (2) Use of selective pressure, for example variation of substrate, often combined with the presence of mutagens, to induce non-natural variants. (3) Use of selective mutagenesis without prior knowledge of 3-D structure based on information from the first two approaches. This can be supplemented by knowledge of sequence homology to provide information on conserved residues. (4) SDM based on knowledge of the 3-D structure supplemented by information from the previous approaches. (5) Use of random mutagenesis via phage display to screen a greater number of variants. This has become tractable only in recent years and is often used in combination with information from the 3-D structure.

Site-specific mutagenesis without the aid of 3-D information often produces spurious results where modifications are often silent or deleterious (Knowles, 1987). Changing substrate specificity through selective pressure can circumvent these problems, as use of selective pressure to alter enzyme characteristics offers the advantage that only catalysts with greater effectiveness are obtained without prior knowledge of the enzyme structure. It has the advantage that positive selection from a random pool of mutants can pick up protein variants that would otherwise be non-intuitive from structural observations. This approach has been used previously to direct the evolution of a variety of enzymes, e.g. β-lactamase (Hall and Knowles, 1976), aliphatic amidase (Bet*z et al*., 1974) and ribitol dehydrogenase (Mortlock, 1982).

Selected modification of the substrate specificity of PGA has potential benefits in the semi-synthesis of β-lactam antibiotics and in the expansion of the range of substrates for other synthetic reactions. Approaches 1-3 have been previously applied to PGA and the results briefly summarised here. They lead directly into the work described in this chapter on the use of SDM as in approach 4.

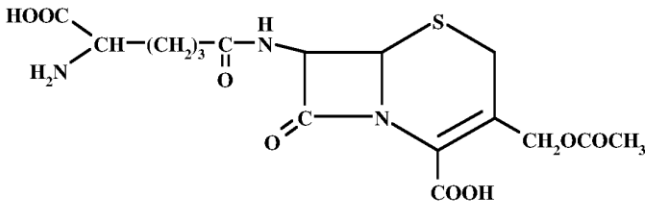
A mutant strain from *E. coli* (*broA*) with modified substrate specificity for PGA was characterised (Williams and Zuzel, 1985), with Met A:142 substituted by leucine, which permitted the hydrolysis of a novel substrate N-(3-carboxy-4-nitrophenyl)-6-bromoadipamide. Subsequent mutants to either valine, alanine, threonine, glycine or histidine at this site were also active on this substrate.

At almost the same time, a limited sequence homology was proposed between *E. coli* penicillin binding proteins (PBP's) and a region in the A subunit of *E. coli* PGA between Met A:142 and Lys A:165, suggesting this region could be involved in the binding of penicillin (Oliv*er et al*., 1985). Mutations of Met A:142 (to Ala A:142 and Val A:142) in *K. citrophila* PGA were also generated based on this proposed homology and were shown to decrease the specificity constant of the enzyme for Pen G, Pen V and phenylacetyl-4-aminobenzoic acid (Prieto *et al*., 1990). The kinetic data suggested that the region interacted directly with substrate, with Met A:142 interacting with the acid moiety of the substrate. Met A:142 was proposed as a potential target for altering the substrate specificity of PGA to hydrolyse substrates of industrial significance other than penicillins.

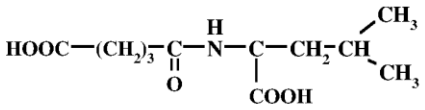
Prieto and colleagues also proposed a second putative penicillin binding region in the B subunit, Ser B:83 to Lys B:98. However the crystal structure of PGA revealed that the latter residues are part of a exposed loop 40 Å away from the active site pocket, showing how caution should be taken in the absence of a 3-D model.

Following the work of Williams and Zuzel, selective pressure has been used successfully with a variety of novel substrates to direct PGA evolution towards new substrate specificity (Daum*y et al*., 1985a; Forney and Wong, 1989b; Forne*y et al*., 1989a; Niersbac*h et al*., 1995a; Niersbac*h et al*., 1995b; Ro*a et al*., 1994). Daumy *et al*. showed that variant acylases could be derived from the penicillin acylases of *E. coli* ATCC 9637 and *Proteus rettgeri* ATCC 31052 by serial passage of the strains in medium containing primary amides as a sole nitrogen source (Daum*y et al*., 1985a)

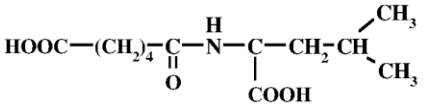
Directed evolution of PGA through selective pressure is best demonstrated by the work of Forney and later Roa who used selective pressure to alter substrate specificity towards two cephalosporin mimics, glutaryl-leucine, and adipyl-leucine (Forne*y et al*., 1989a; Ro*a et al*., 1994) (Figure 6.1) using amino acid auxotrophs. For example, if the amine portion of the substrate is replaced by an amino acid, a positive genetic screen can be used to select for activity from a randomly generated mutant pool in *E. coli* strains auxotrophic for that amino acid.



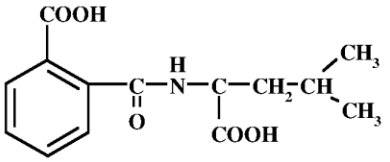
Cephalosporin C



Glutaryl–(L)–leucine



Adipyl–(L)–leucine



Phthalyl–(L)–leucine

Figure 6.1 Chemical structure of cephalosporin C and some novel substrates used to direct the evolution of PGA substrate specificity.

pac (pA1) was transformed into a leucine auxotroph and mutagenised with either UV light or chemical mutagenesis (*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine). Mutants were selected for their ability to hydrolyse glutaryl-(L)-leucine. Since PGA is a periplasmic protein, growth of the mutant strain was restricted by modulation of medium pH and substrate concentration. Subsequently, successive generations of mutants that hydrolysed glutaryl-(L)-leucine more efficiently than the parental mutants were generated using a *mutD5* strain. This strain has a defective DNA polymerase 3'-5' exonucleolytic editing activity (Echol*s et al*., 1983) that results in a increased mutation frequency when the strain is grown in rich medium (Fowle*r et al*., 1974). First and second generation mutants (PA251 and PA251-47 respectively) were thus selected for their ability to hydrolyse glutaryl-(L)-leucine most rapidly between pH 5 and 6 and showed increased catalytic efficiency with a second order rate constant (k_{cat}/K_M) for PA251-47 2.3 times greater than the first generation mutant PA251, while the wild type enzyme showed no activity.

Prior to the structural determination of PGA, work was undertaken at York to characterise these altered specificity mutants provided by Dr. Larry Forney. They were sequenced in our laboratory by Dr. Jim Brannigan and showed a mutation at residue Phe B:71 to either leucine, TTC→CTC (Leu^{B71}; 1st generation) or cysteine, TTC→TGC, (Cys^{B71}; 2nd generation). To confirm that a single mutation was the unique cause of the change in substrate specificity, the first generation and second generation mutations were reproduced by site-directed mutagenesis by Dr. J. Brannigan and their properties shown to be identical to those of Forney.

Similar experiments using adipyl-(L)-leucine and phthalyl-(L)-leucine as substrates resulted in mutations at Phe B:71 to valine (Ro*a et al*., 1994) and Gly B:70 to aspartic acid (Niersbac*h et al*., 1995b) respectively. Thus, selective pressure by quite different substrates led to amino acid alterations in the same region of the active site pocket, especially B:71 and this region must play an important role in determining substrate specificity.

Morillas *et al*. has carried out a detailed kinetic analysis of the Leu^{B71} and Cys^{B71} mutants (Morilla*s et al*., in preparation) using enzyme prepared in York by myself. Using a series of phenylacetate and non phenylacetate substrates both steady state and pre-steady state kinetic parameters were determined, including the effect of pH. The

conformational stability of the mutants was also compared with wild type as a function of pH and various denaturants. The results of the steady state kinetic parameters are shown in Table 6.1. Both mutants catalysed the hydrolysis of glutaryl-(L)-leucine with a value of k_{cat}/K_m of ~50 s^{-1} M^{-1} and gave similar results to those described previously for mutants of *E. coli* PGA (Forney *et al*., 1989a) and for the homologous enzyme from *K. citrophila* (Roa *et al*., 1994). Although the specific activity for glutaryl-(L)-leucine hydrolysis is modest for the mutants, displaying both a high K_M (~16 mM) and low k_{cat} (~1s^{-1}), this was an advance since hydrolysis by wild type enzyme is not detectable.

The specific activity k_{cat}/K_M of both mutants for phenylacetate derivative substrates were 5-10 fold higher with NIPAB and 2-3 fold higher with *p*-nitrophenyl phenylacetate, compared to wild type enzyme. The K_M of both mutants for NIPAB and *p*-nitrophenyl phenylacetate showed 4-5 fold increase in affinity. Both wild type and mutants show a decrease in specific activity by 10^2 to 10^4 fold for substrates which do not have a phenylacetate moiety including glutaryl-(L)-leucine, but while the activity is essentially zero for the wild type, it is improved for the mutant. If the rate determining step is the acylation rate, e.g. with NIPAB, then the mutants show an increase in specific activity. Surprisingly, this seems to be also the case with penicillin G and cephalosporin C hydrolysis (Morillas *et al*., in preparation).

	Wild type	Cys^{B71}	Leu^{B71}
NIPAB			
k_{cat}/K_m (s^{-1} μM^{-1})	1.11 □ 0.25	6.13 □ 0.74	10.7 □ 3.1
k_{cat} (s^{-1})	20.1 □ 3	19 □ 2.1	37 □ 3.5
K_m (μM)	18 □ 2.2	3.1 □ 0.35	3.4 □ 0.43
p-nitrophenyl phenylacetate			
k_{cat}/K_m (s^{-1} μM^{-1})	11.3 □ 0.41	20.0 □ 5.9	34.4 □ 6.9
k_{cat} (s^{-1})	140 □ 20	72 □ 15	135 □ 25
K_m (μM)	12 □ 2.2	3.6 □ 0.51	3.92 □ 0.61
p-nitrophenyl acetate			
k_{cat}/K_m (s^{-1} μM^{-1})	7.8 □ 12 x 10^4	2.33 □ 0.58 x 10^4	5.40 □ 0.52 x 10^4
k_{cat} (s^{-1})	0.71 □ 0.06	0.091 □ 0.0082	0.19 □ 0.09
K_m (μM)	9.10 □ 11	3.90 □ 0.53	3.52 □ 0.46
phenylacetate			
k_{cat}/K_m (s^{-1} M^{-1})	1.97 □ 0.19 x 10^5	1.22 □ 0.25 x 10^5	2.40 □ 0.33 x 10^5
k_{cat} (s^{-1})	0.84 □ 0.09	0.089 □ 0.012	0.18 □ 0.021
K_m (M)	4.27 □ 0.71 x 10^{-6}	7.31 □ 0.81 x 10^{-7}	7.51 □ 0.92 x 10^{-7}
o-nitrophenyl acetate			
k_{cat}/K_m (s^{-1} M^{-1})	349 □ 156	438 □ 135	498 □ 280
k_{cat} (s^{-1})	0.15 □ 0.039	0.15 □ 0.021	0.21 □ 0.031
K_m (M)	4.37 □ 2.49 x 10^{-4}	3.42 □ 1.35 x 10^{-4}	4.21 □ 1.02 x 10^{-4}
glutaryl-(L)-leucine			
k_{cat}/K_m (s^{-1} M^{-1})	NA	45.50 □ 15.21	55.97 □ 17.27
k_{cat} (s^{-1})		0.81 □ 0.11	0.89 □ 0.10
K_m (mM)		17.8 □ 4.0	15.9 □ 5.8

NA: No activity detected

Table 6.1 Steady state constants of wild type and the mutants Leu^{B71} and Cys^{B71} of Penicillin G acylase (Morillas, 1996).

The mutants exhibit a broader pH profile for hydrolysis of NIPAB. The mutants were more stable at pH 10-11 in the presence of substrate compared to wild type, but Morillas

observed no significant differences in the unfolding transitions and proposed that local interactions between the enzyme and substrate determined the enzyme stability at alkali pH. The mutants have greater stability in the presence of urea and guanidinium chloride.

The X-ray structure of one of these altered specificity mutants was determined in order to rationalise on a structural basis, the effect of the mutations, produced by different substrate selected pressures. The structure is related to the kinetic data on the altered specificity mutants (Leu^{B71} and Cys^{B71}) obtained by Dr. Richard Virden and Dr. Manuel Morillas, at the University of Newcastle-upon-Tyne (Morillas *et al*., in preparation).

6.2 Experimental

6.2.1 Cloning and expression of the mutants Leu^{B71} and Cys^{B71}

Clones of the altered specificity mutants Phe B:71 Leu (Leu^{B71}) and Phe B:71 Cys (Cys^{B71}) were provided in the plasmid Bluescript KS (-), pBS-L71 and pBS-C71 respectively by Dr. Jim Brannigan. These were subcloned into the *Dra*III/*Bgl*II site of the wild type expression clone pA1 replacing the wild type *Dra*III/*Bgl*II fragment with the respective mutant fragment. Expression and purification of these mutants were as described in Chapter Two .

The 750 bp *Dra*III/*Bgl*II fragments of pBS-L71 and pBS-C71 were gel purified and ligated to the Δ*Dra*III/*Bgl*II pA1 linearised plasmid, Figure 6.1. Following transformation into the *E. coli* strain XL-1 Blue MRF' and plasmid DNA purification, the mutations were confirmed by restriction digest with *Bam*HI, Figure 6.2. *pac* is cloned into the *Bam*HI restriction site of pACYC184 and so restriction with *Bam*H1 releases a 3.25 Kb fragment. The Leu^{B71} and Cys^{B71} mutants have incorporated a unique *Bam*HI site within *pac* close to the mutation site and so DNA digestion with *Bam*HI releases the mutated *pac* gene as two DNA fragments of approximately 2.1 Kb and 1.1 Kb.

The mutants were expressed in BL21 (DE3) cells and purified as described previously. Purified protein was analysed by SDS-PAGE (Figure 6.3) and the fractions containing

pure protein pooled. The protein yield for the mutants were approximately 4 mg/l of culture, approximately 50 % of the yield obtained for wild type PGA.

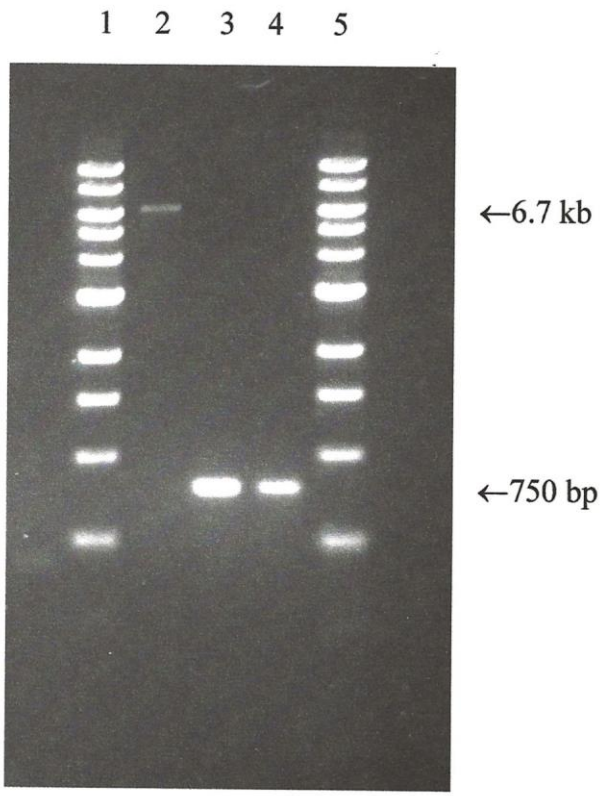


Figure 6.2 The agarose gel purified DraIII/BglII DNA fragments of pBS-L71 and pBS-C71 and DraIII/BglII linearised pA1. *Lanes 1 and 5: 1Kb ladder (NEB). Lane 2: ΔDraIII/BglII pA1. Lane 3: DraIII/BglII fragment of mutant LeuB71. Lane 4: DraIII/BglII fragment of mutant CysB71.*

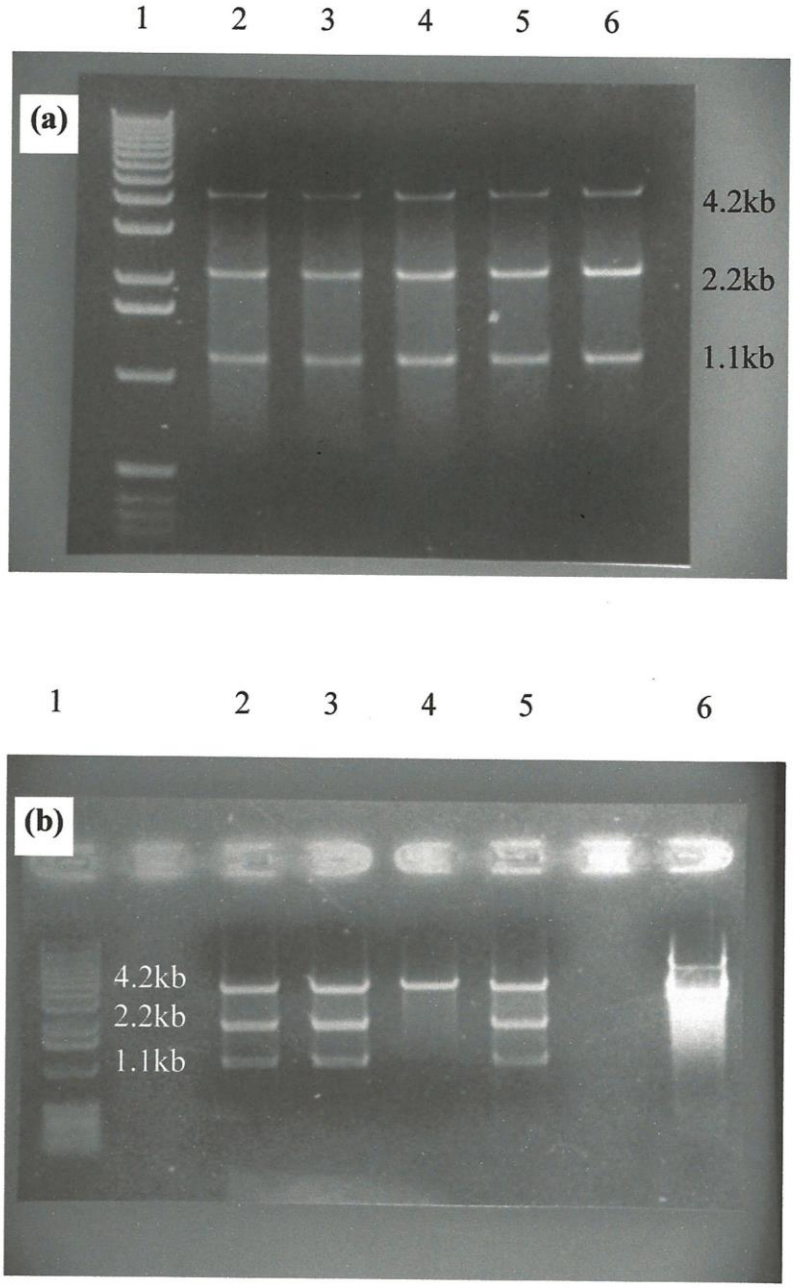


Figure 6.3 Restriction analysis of (a) Leu^{B71} DNA and (b) Cys^{B71} DNA with Bam*HI. *(a) Lane 1: 1Kb ladder (Gibco BRL). Lanes 2 to 6: Bam*HI *digest of Leu^{B71} clones. (b) Lane 1: 1Kb ladder. Lanes 2 to 5 Bam*HI *digest of Cys^{B71} clones and Lane 6: Bam*HI *digest of pA1.*

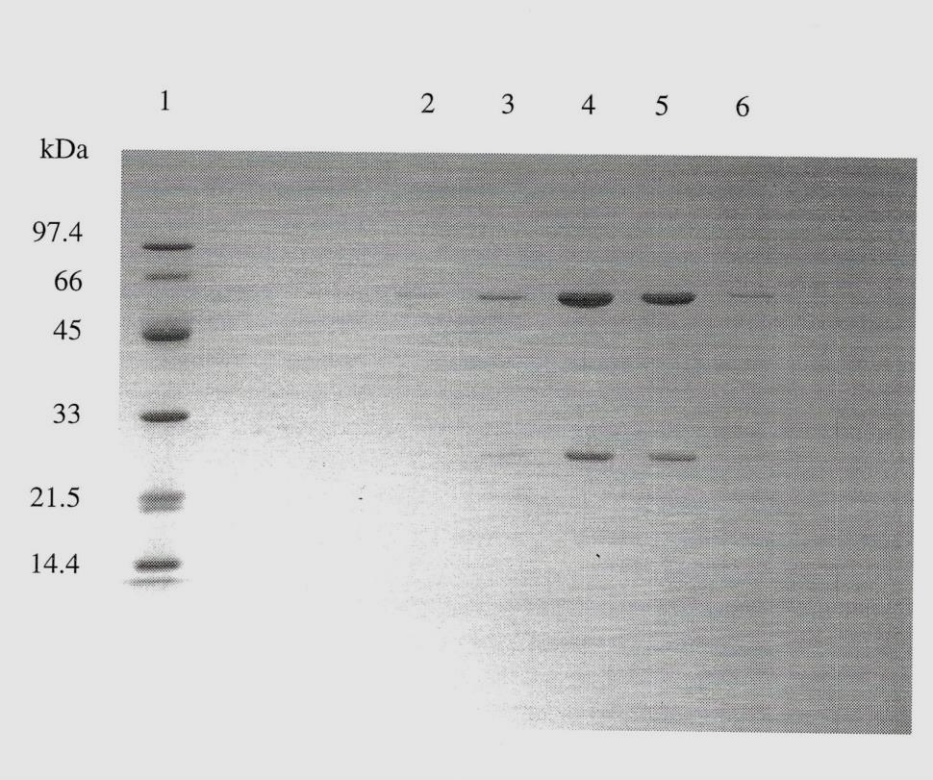


Figure 6.4 Final purification step of Leu^{B71}. *Lane 1: Protein markers (Biorad). Lanes 2 to 6: FPLC fractions of Leu^{B71} after Q-Sepharose chromatography.*

6.2.2 Crystal growth and data collection

Crystals of Leu^{B71} (Phe B:71 Leu) were obtained using a combination of the hanging drop method and streak-seeding techniques with a protein concentration of 11 mg/ml in 11% (w/v) PEG MME 2K, 50 mM MOPS (pH 7.2). After serial transfer (5% steps) to 30% (v/v) ethylene glycol in mother liquor for cryoprotection, data were collected at 120 K for Leu^{B71} to 2 Å resolution on a RAXIS II image plate detector with a Rigaku RU200 rotating anode X-ray generator. The space group was *P*1 with unit cell dimensions a = 52.0 Å, b = 64.2 Å, c = 70.7 Å, $\alpha = 70.6°$, $\beta = 72.8°$ $\gamma = 73.8°$, with one molecule in the asymmetric unit. Data were indexed and integrated using *HKL* suite (Otwinowski and Minor, 1997) and scaled using *ROTAVATA* and *AGROVATA* from the *CCP4* suite (Collaborative Computational Project, 1994). Data reduction statistics are presented in Table 6.2.

Data statistics for Leu^{B71}	
Space group	*P*1
Cell (Å)	a = 52.0 b = 64.2 c = 70.7
Cell (°)	α = 70.6 β = 72.8 γ = 73.8
Resolution (Å)	20 - 2.0
Wavelength (Å)	1.5418
Temperature (K)	120
Measured reflections	102704
Unique reflections	50677
Mosaicity (°)	0.3
Completeness (%)	93 (80)
$\langle I \rangle$ / $\langle \sigma(I) \rangle$	14.8 (7.1)
R_{merge} * (%)	4 (10.1)

Numbers in parentheses refer to the outermost resolution shell (2.1-2.0 Å)

*$R_{merge} = 100 \sum_{hkl} |I_{hkl} - \langle I \rangle| / \sum_{hkl} I_{hkl}$

Table 6.2 Data collection statistics for the altered specificity mutant Leu^{B71}.

6.2.3 Molecular replacement

The unit cell dimensions of the room temperature triclinic PGA were *a* = 52.1 Å, *b* = 65.1 Å, *c* = 76.3 Å, α = 100.2°, β = 111.4°, γ = 105.8°. Due to the shrinkage of the cell dimensions caused by freezing of the crystals (difference of 5.6 Å in *c* dimension), the structure of the mutant Phe B:71 Leu (Leu^{B71}) was determined by MR with *AMoRe* (Navaza and Saludjian, 1997). The search model was the wild-type 1.9 Å penicillin acylase structure (Duggleby *et al*., 1995; PDB accession code 1PNK) with the water molecules removed. An integration sphere of 32 Å radius and increments of 2.5° in each Eulerian angle were applied between calculations. Data between 10 and 4 Å were included for the rotation function and between 15 and 3 Å for rigid-body refinement. The solution was unambiguous with a peak height 33σ above the mean. As the molecular replacement solution was for a P1 unit cell, a translational search was not needed as the origin is arbitrary. The rotation parameters were refined by rigid body refinement using the *FITING* function in *AMoRe*. The refined solution gave an *R*-factor of 31.9%, and a correlation coefficient of 77.8% (Table 6.3). The search model was positioned according to the refined solution and orientated using *LSQKAB* (Collaborative Computational Project, 1994).

Solution	α	β	γ	CorrF	RFac
ROTATION Function	195.67	92.40	308.75	56.8	0.00
FITING Function	195.64	92.74	308.63	77.8	31.9

Table 6.3 The solution from the rotation function and after rigid body refinement.

6.2.4 Refinement

The reoriented model was refined using maximum likelihood as implemented in *REFMAC* (Murshudov *et al*., 1997) with bulk solvent correction and anisotropic scaling, as described in Chapters 2, 3 and 5. Towards the end of refinement the contribution of hydrogen atoms was included in the model using *HGEN*. Initially the addition of water molecules used the automated refinement programme *ARP* (Lamzin & Wilson, 1992) in concert with *REFMAC*. Water molecules were added manually in the later stages using X-SOLVATE, resulting in a final model of 762 amino acids, 557 waters, one calcium ion and six ethylene glycol molecules. The final R-factor was 15.8% with free R-factor of 21.5%, based on 5% of the diffraction data. The refinement statistics are given in Table 6.4.

6.3 Results and discussion

6.3.1 Structure of the altered specificity mutant Leu^{B71}

The overall structure of the mutant is the same as the triclinic wild type despite the unit cell shrinkage. The initial difference maps showed regions of both positive and negative density in the active site. The electron density at position B:71 clearly reconfirmed the leucine mutation (Figure 6.5) and additional density in the vicinity suggested conformational changes in the neighbouring residues, Leu B:253 and Phe B:256: Leu B:253 adopted a different rotamer while Phe B:256 displayed a more dramatic reaction to the mutation with a pronounced side chain movement.

Mutation of phenylalanine B:71 to leucine clearly alters the hydrophobic environment of the active site. Loss of aromatic stacking interactions leads to a flip of the hydrophobic side chain of Phe B:256 out into the periphery of the active site (Figure 6.6) which may then alter the substrate environment, in particular the S1' subsite. The CZ atom of Phe B:256 moves 4.9 Å and as a result narrows the entrance to the active site. Phe B:256 is now pinned between Leu B:71 and Leu B:253 whose side chain has flipped in the opposite orientation (Figure 6.6) to the wild type structure.

Refinement statistics for Leu^{B71}	
Resolution range (Å)	20 - 2
R-factor † (%)	15.8
R_{free} ‡ (%)	21.5
Water molecules	557
Ethylene glycol (ETG) molecules	6
rms deviations from ideality (Å)	
bond 1-2 distance (0.02)	0.013
angle 1-3 distance (0.04)	0.030
planar 1-4 distance(0.05)	0.036
Average *B*-factor (Å^2)	
Protein	21.7
Water	29.6
ETG	24
Ca^{2+}	18.37
Coordinate error § (Å)	0.18

†R-factor = $\sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl} \left| F_{obs} \right|$

‡Free R-factor = $\sum_{hkl \subset T} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl \subset T} \left| F_{obs} \right|$ where $hkl \subset T$ represents 5% of the diffraction data. §Coordinate error estimated by DPI, the Diffraction-data Precision indicator (Cruickshank, 1996).

Table 6.4 Refinement statistics for the altered specificity mutant Leu^{B71}.

Leu^{B71} not only leads to a drastic alteration of the binding surface accessible to the substrate, but also to more subtle changes in the active site. The electron density for residues A:142 – A:146 suggests the presence of both the kinked helix and alternative loop conformations (Chapters 3 and 5) for these residues. Thus the Leu^{B71} mutation appears to affect the equilibrium between the two conformations, increasing the proportion of the loop form to about 50%, from a much lower level in the wild type. This illustrates once more the sensitivity to disruption of the hydrophobic network.

The slow stepwise transfer of the crystal into cryosolvent provides an explanation for the six ETG molecules in the structure. The serial transfer took place over a period of 30 min and shows PGA has a reasonable affinity for ethylene glycol. There are two ETG molecules in the active site, one of which is situated in the S1 pocket (ETG1), and the second is equivalent to ETG3 in the monoclinic wild type structure (Figure 3.16).

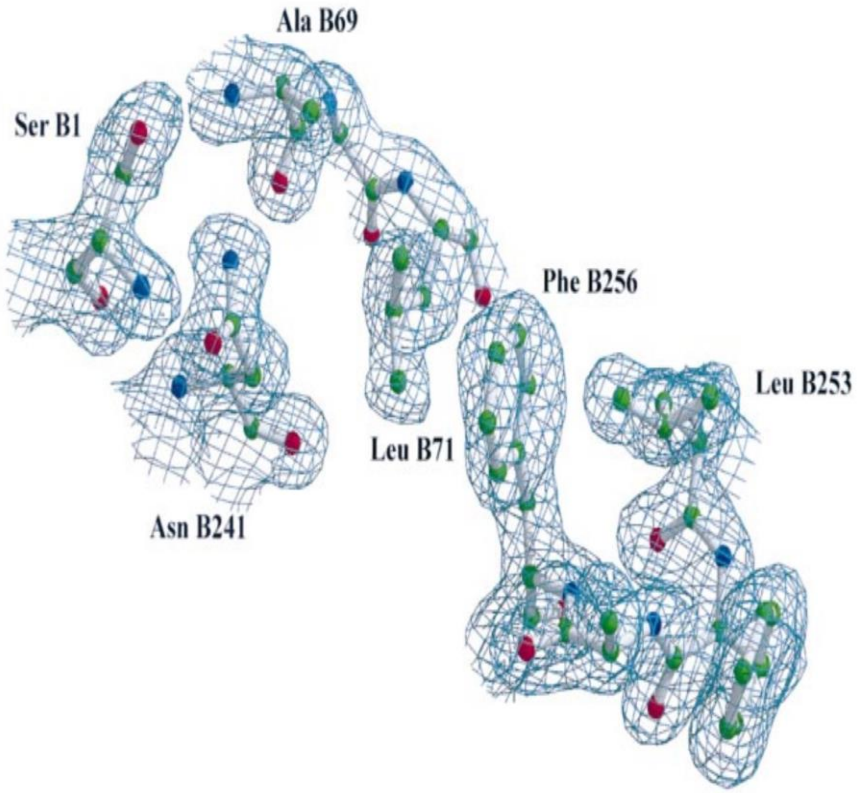


Figure 6.5 The final $2F_o–F_c$ electron density map contoured at 1σ showing the refined atomic positions of residues around the conformational change of the Leu^{B71} mutant.

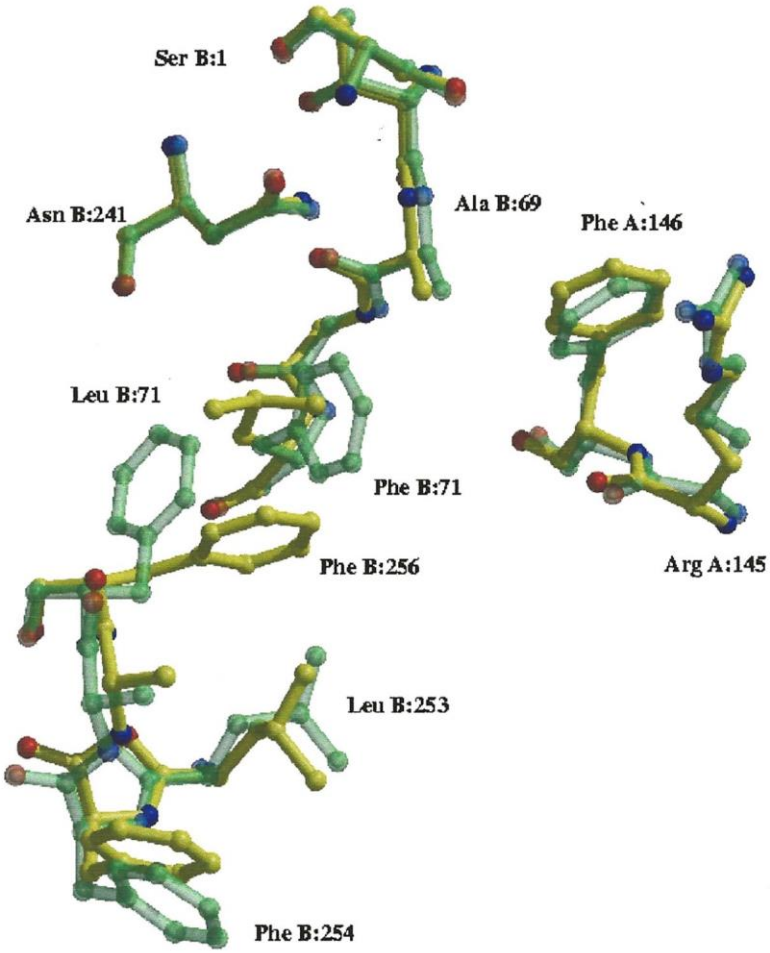


Figure 6.6 Overlap of Leu^{B71} (yellow) with wild-type PGA (green) to show conformational changes within active site.

It is probable that this leads to changes in the surface complementarity between enzyme and substrate rather than altering a specific interaction which provides a rationale for the mutation at this same point in the chain selected by different substrate challenges.

6.3.2 Structural evidence for the kinetic properties of the altered specificity mutants

It is difficult to rationalise the improved stability of the mutants as there seems to be no obvious structural explanation. However the greater stability was measured in the presence of NIPAB and perhaps the mutation offers greater stability when NIPAB is bound in contrast to wild type PGA. A similar result has been described for sarcosine oxidase (Nishiya and Imanaka, 1994) where the mutation F103V showed altered specificity and exhibited *N*-methylvaline oxidase activity with a catalytic efficiency 15 times higher than wild type. They also showed that the pH profile is shifted (from pH 7.5 $\rightarrow$ 8.5) as a result of the mutation although the charge is not altered. Generation of mutants at regions coding for amino acids predicted to be at the surface of *E .coli* PGA resulted in a mutant with enhanced stability at alkaline pH (del Ri*o et al*., 1996; del Ri*o et al*., 1995). Characterisation of this mutant, Trp B:431 Arg, shows a near two-fold increase in half-life at pH 8.5 as compared with wild type enzyme (del Ri*o et al*., 1995). Trp B:431 is located on a surface helix and mutation to an arginine could well result in a favourable hydrogen bond to a glutamine residue on a neighbouring helix.

The three main features that distinguish the altered specificity mutant from the wild type enzyme are:

1. **Some main chain movement of Ala B:69 which forms part of the oxyanion hole.**

Reports of altered specificity mutants where the mutation had been identified showed that the amino acid changes were confined to residues Gly B:70 (Gly B:70 Asp, Niersbach *et al*., 1995b) and Phe B:71 (to either leucine, cysteine, this lab; valine, Roa *et al*., 1994). The close proximity of the mutations to the proposed main chain oxyanion hole residue Ala B:69 may influence the improvement in k_{cat} for NIPAB hydrolysis reported by several groups. Subtle changes at the oxyanion hole invoked by the mutations at B:71 could influence catalysis.

2. **Movement of Phe B:256.**

Without the structural determination of the mutant Leu^{B71} one could not have predicted the consequent change in active site conformation. The loss of aromatic interaction between Phe B:71 and Phe B:256 and ring flip of the latter, consequently changes both pH and thermal stability. This also enhances substrate affinity for NIPAB and other phenylacetate derivatives in addition to the novel activity against glutaryl-(L)-leucine.

3. **Narrowing of the entrance to the active site pocket.**

The narrowing of the active site pocket by the conformational ring flip of Phe B:256 (see Figure 6.7) is perhaps consistent with the improved K_M values (5-10 fold higher for phenylacetate derivative compounds). Superimposition of the Ala^{B241}-PG complex described in Chapter 5 with the Leu^{B71} mutant adds further support to the tighter binding observed for NIPAB possibly by introducing an extra hydrophobic interaction with the substrate (Figure 6.8).

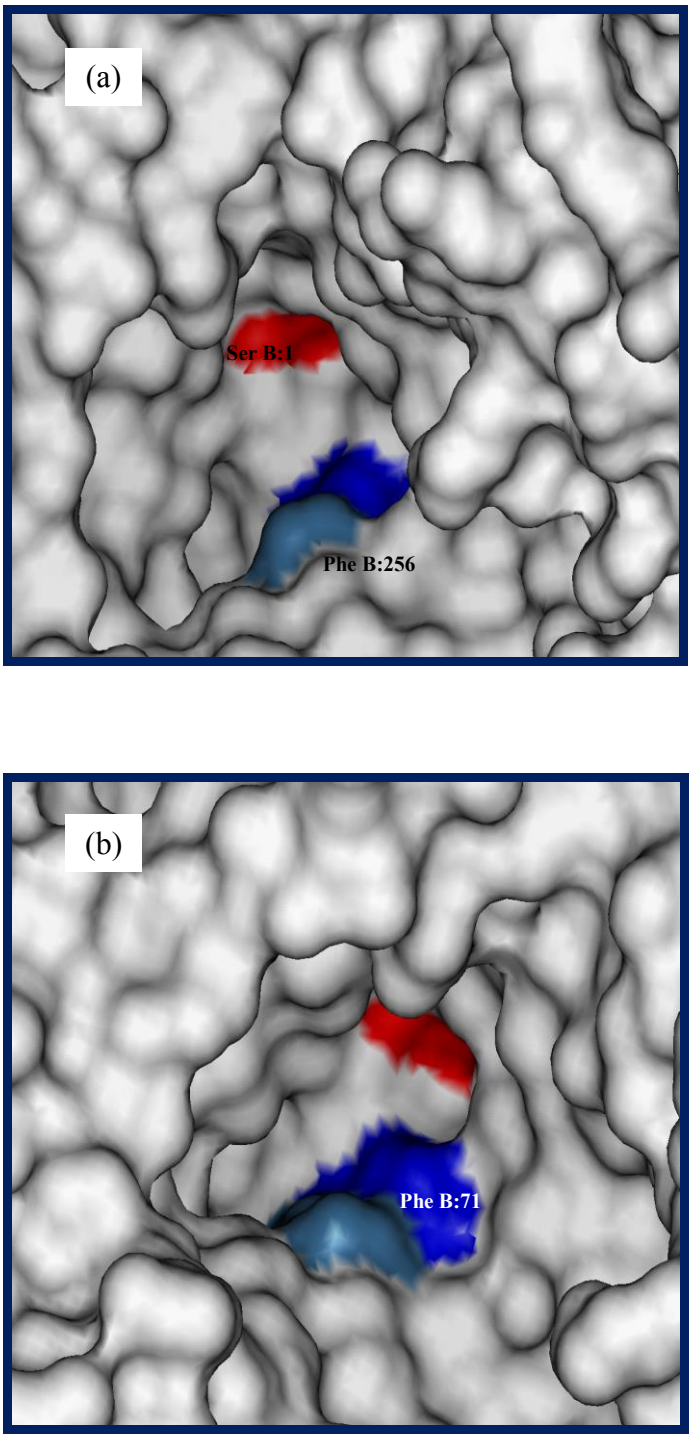


***Figure 6.7 Surface representation of active site to show narrowing of* substrate pocket.** (a) The active site pocket of Leu^{B71}. (b) The active site pocket of wild type PGA. The active site Serine is coloured in red, Phe B:256 is coloured light blue and Phe/Leu B:71 is coloured dark blue.

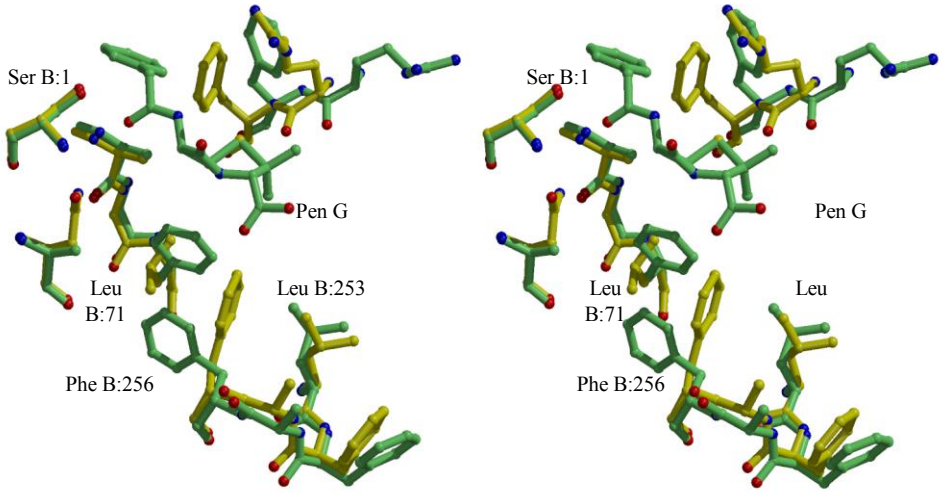


Figure 6.8 Stereoview of the altered specificity mutant Leu^{B71} (yellow) superimposed with the Ala^{B241}-PG complex (green).

6.4 Conclusions

The substrate specificity of penicillin G acylase was previously altered by selective pressure to hydrolyse a substrate mimic of cephalosporin C, namely glutaryl-(L)-leucine (Forney *et al*., 1989a), while no more than 1 % of this activity was shown by the wild type enzyme (Morillas, 1996). However, the catalytic efficiency is still insufficient to be viable for industrial applications. What was most surprising is that the altered specificity variants are in a specific mutagenic hotspot, Phe B:71. The results of extensive activity, specificity and stability experiments on these and closely related mutants are available.

Mutant Leu^{B71} was selected for 3-D analysis. The crystals obtained were closely similar to the triclinic crystals of the wild type enzyme, and its structure has been refined at a resolution of 2.0 Å. Comparison with that of the wild type allows some conclusions regarding the specificity of both the mutant and wild type enzymes. Phe B:71 forms one face of the S1' specificity pocket, and is a neighbour of the oxyanion residue Ala B:69.

This was not expected since the altered specificity was targeted towards the S1 specificity pocket.

B:71 acts in synergy with A:146 to clasp the substrate ready for catalysis. The 3-D structure of Leu^{B71} emphasises the importance of hydrophobic interactions within the binding site. The conformational changes in the active site pocket in the mutant would appear to result in a change in complementarity of the enzyme surface to substrate, based on comparison with the complexes described in Chapter 5. A major feature is the ring flip of Phe B:256 which provides an additional potential hydrophobic interaction to the substrate. This extra contract would explain the improved affinity of the mutant for NIPAB and rationalise the mutants' kinetic properties towards novel substrates as well as the established substrate spectrum for PGA. It also explains why mutant selection with a variety of substrates has led to alterations at the same site, as the new interactions are not highly specific, but cannot be formed by the wild type enzyme.

Some directed evolution experiments resulted in mutations in the S1 pocket and targeted Met A:142, at the back of the pocket, making structural sense for the observed altered specificity. The contrasting results observed for the altered specificity mutants suggests that PGA is perhaps not an ideal candidate for protein engineering to give a viable cephalosporin acylase and that a different enzyme may provide a better starting point, one possible target being glutaryl-7-ACA acylase. Nevertheless, these preliminary results do provide some impetus in the pursuit of using an engineered acylase to hydrolyse cephalosporins.

Selective pressure can be used to modify the substrate specificity of PGA, but, owing to the synergistic binding of both acyl and amine moieties, an amino acid that could favour binding of new acyl side chains might well alter the broad range of amine moieties recognised by the wild type enzyme. An altered specificity mutant Phe B:71 Val, from *K. citrophila* PGA was characterised as result of positive selection towards adipyl-(L)-leucine and was also able to hydrolyse glutaryl-(L)-leucine with greater efficiency than wild type enzyme (Ro*a et al*., 1994). However the mutant was not able to hydrolyse glutaryl-7-aminocephalosporanic acid and had lost 90% and 50% of activity on penicillin G and phenylacetyl-(L)-leucine while retaining its original activity with NIPAB. It was suggested that the binding specificity of PGA by the acyl and amine moieties of the substrate are not independent phenomena (Ro*a et al*., 1994).

Forney examined the possibility of using PGA in the synthesis of ampicillin and cephalexin by altering the substrate specificity of PGA to hydrolyse α-aminophenylacetyl-(L)-leucine, a substrate mimic, using positive selection with a leucine auxotroph (Forney and Wong, 1989b). Forney does point out that the use of a leucine at the P1' site of the selection substrate does not preserve the ability to effectively bind penam or cephem nuclei, he suggests replacing leucine with an amino acid such as tryptophan which has structural features more closely resembling those of β-lactam nuclei. This is borne out by Forney's results which showed a 2-fold drop in affinity for cephalexin and 10-fold drop in specific activity.

One can predict a similar conformational change for the Cys^{B71} mutant. However the presence of a polar (Cys) instead of a non polar (Leu) side chain may influence the interactions between the cysteine at B:71 and the phenylalanine side chain of B:256. The mutation does make a small difference to the active pocket as supported by the kinetic parameters when compared to Leu^{B71}.

A substrate complex with an altered specificity mutant is needed to add further support to the experimental evidence provided here, both structural and kinetic, to aid in our understanding for the observed altered specificity. A complex with NIPAB would be appropriate as it has good affinity with the mutants in contrast with other substrates. Expression of a second knockout mutation to enable binding without turnover of substrate needs further investigation. Preliminary results of a double mutant Ala^{B241}/Leu^{B71} show reasonable expression of a processed enzyme.

Chapter Seven Probing substrate specificity by amber suppressor mutagenesis

7.1 Introduction

The work described in Chapter 5 shows that Phe A:146 plays a pivotal role in substrate specificity and binding and was selected for investigation using a novel mutagenesis approach. The ability to substitute specific amino acids is a key requirement for studying protein structure-function relationships, ideally to introduce each of the remaining 19 amino acids at a site of interest and attempts to achieve this have applied a variety of approaches such as degenerate primers and selecting mutant from a random pool or a more laboured method, to make each mutant by separate mutagenesis (Wilson and Agard, 1991; Zoller, 1991). However, it would be desirable to study multiple amino acid substitutions from a single mutation experiment.

The approach selected here involved the use of *in vivo* amber suppression, for which a brief background now follows. Miller and coworkers (Mill*er et a*l., 1989; Norman*ly et a*l., 1986b; Norman*ly et a*l., 1986a) were the first to exploit the phenomena of nonsense suppression by changing the identity of a transfer RNA (tRNA). Nonsense suppressors have mutations that alter the anticodon of tRNAs, allowing insertion of an amino acid in response to one of the nonsense triplets, UAG, UAA, or UGA. To employ this method, genes containing amber, ochre, or opal mutations (resulting in UAG, UAA, or UGA chain terminating codons respectively) are suppressed in mutant strains that produce suppressor tRNAs capable of inserting an amino acid in response to the nonsense mutation. Amber suppressor tRNAs have been constructed which allow insertion of any one of twelve amino acids (Figure 7.1) at an amber (TAG) stop codon (Klei*na et a*l., 1990b; Kleina and Miller, 1990a; Norman*ly et a*l., 1990). Although not comprehensive, it provides a representative sample of amino acid properties including size, charge and hydrophobicity.

Using a set of strains containing these various suppressor tRNAs, it is possible to determine the effect of substituting each of the twelve amino acids at a particular site

from a single amber mutation. Systematic survey of the effects of single amino acid substitutions in both *lac* repressor and T4 lysozyme has been used to generate a functional map describing which are deleterious, allowing a detailed analysis of a protein's substitution tolerance. Amber mutations were introduced into every codon of bacteriophage T4 lysozyme gene resulting in 2015 amino acid substitutions (Renne*ll et a*l., 1991). For *lac* repressor some 4000 variants were analysed (Markiewi*cz et a*l., 1994; Suck*ow et a*l., 1996).

The method has been taken one step further with attempts to expand the genetic code to incorporate unnatural amino acids (L*iu et a*l., 1997; Schimmel and Söll, 1997) which would enable a variety of applications, including structure-function analysis of specific sites in proteins by means of probes inserted at defined locations and the creation of proteins with new chemical activities. Liu *et al*. (1997) describe a first step towards a non-toxic *in vivo* aminoacylation system that allows the incorporation of a non-natural amino acid at a defined position in a protein of choice. Their attempts have demonstrated the complexities imposed by that part of the tRNA world that deals with the translation apparatus and the genetic code. 75% incorporation of the amino acid analogue *p*-fluoro-phenylalanine at a programmed amber stop codon in the DHFR marker protein has been demonstrated, providing fluorinated protein suitable for ^{19}F-NMR spectroscopy (Furter, 1998).

Amber suppression is here applied to PGA. The structural information gained from the substrate complexes and the altered specificity mutant in previous Chapters provides the platform for such mutagenesis experiments addressing further aspects of the enzyme specificity. The use of selective pressure, Chapter 6, to direct the evolution of PGA towards new substrates highlighted the importance of B:71 as a determinant of substrate specificity, with a rationale provided by the 3-D structure of the mutant Leu^{B71}. In addition, the structure of the Pen G complex, Chapter 5, revealed the synergy between A:146 and B:71 in substrate binding. The A chain helix which comprises in particular the residues Met A:142 to Phe A:146 forms one face of the active site. This helical region has shown structural flexibility upon substrate and ligand binding.

Phe A:146 appears to play a key role in binding and perhaps catalysis and was selected as the target for mutagenesis using the amber suppression system.

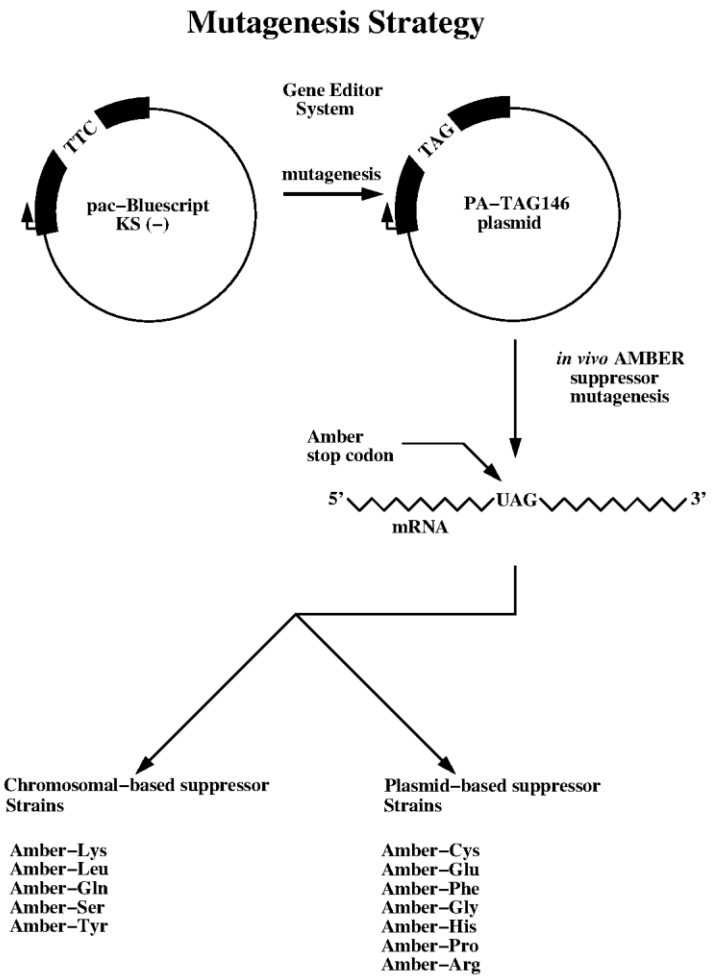


Figure 7.1 A schematic representation of the amber suppressor mutagenesis system.

7.2. Amber stop codon in vivo *mutagenesis*

This system expands the versatility of site-directed mutagenesis where a single amber stop codon mutation can be exploited to study a variety of amino acid substitutions (Figure 7.1). To measure the effect of a particular substitution on protein activity, an amber mutant is generated at the location of interest by site-directed mutagenesis. Subsequently, the plasmid containing this amber is transformed into the various suppressor bacterial strains. Mutants generated in ColE1 plasmids such as pBR322 and pUC derivatives, e.g. pBluescript KS, must be subcloned into a pACYC derived vector in order to be compatible with the plasmid-based suppressors (Table 7.1).

Strain	Amino acid inserted at TAG stop codon	Suppressor source	Antibiotic resistance*	Subcloning from ColE1 vectors
Amber-Cys	Cysteine	plasmid	amp	Yes
Amber-Glu	Glutamic acid	plasmid	amp	Yes
Amber-Phe	Phenylalanine	plasmid	amp	Yes
Amber-Gly	Glycine	plasmid	amp	Yes
Amber-His	Histidine	plasmid	amp	Yes
Amber-Pro	Proline	plasmid	amp	Yes
Amber-Arg	Arginine	plasmid	amp, tet	Yes
Amber-Lys	Lysine	chromosomal	tet	No
Amber-Leu	Leucine	chromosomal	none	No
Amber-Gln	Glutamine	chromosomal	none	No
Amber-Ser	Serine	chromosomal	none	No
Amber-Tyr	Tyrosine	chromosomal	none	No

*amp = ampicillin, tet = tetracycline

Table 7.1 Characteristics of amber suppressor strains.

It was a great advantage that the PA-TAG146 expression plasmid (p15A origin of replication) usually used for PGA was compatible with the plasmid based amber suppressors (ColE1 origin of replication). PGA activity can then be assayed in each of the strains to determine the effect of the amino acid substitution.

7.2.1 Mutagenesis

An amber stop codon introduced at position A:146 in *pac* used the GeneEditor™ system described previously. Phe A:146 was targeted to investigate further the role of this residue in substrate recognition and binding. *pac* ssDNA was prepared as described in Chapter Two (Figure 7.2) and used as a template for the mutagenesis experiment. A 40-mer oligonucleotide DNA was designed that incorporated an amber stop codon at position encoding Phe A:146. A silent mutation which introduces a *Spe*I restriction site was included in the oligonucleotide to identify positive clones containing the amber mutation.

TAG-146 non-coding sequence

5'-ATC AAT TTC ACT AGT GCT ATC AGA CTA GCG GTT TGC CAT G-3'

***Spe*I** **amber stop codon**

7.2.1.1 Incorporation of an amber stop codon at position A:146

A number of transformed colonies from the mutagenesis experiment were selected and plasmid DNA isolated. Mutant clones are predicted to incorporate a new *Spe*I site which is 2,500 bp from the vector encoded site. Restriction digest of the plasmid DNA with *Spe*I identified positive clones with the amber stop codon incorporated by the presence of a new 2,500 bp restriction fragment. Analysis of the mutagenesis experiment showed that 60% of the clones contained the amber stop codon (Figure 7.2).

7.2.2 Cloning and expression of amber suppressor mutants

The 3.24 kb *Bam*HI fragment of *pac* with an amber stop codon at A:146 (*pac-TAG146*) was subcloned from the Bluescript KS (-) vector into the *Bam*HI site of the expression vector pACYC184 to give the expression plasmid PA-TAG146. The vector pACYC184

has a p15A origin of replication, this is compatible with the plasmid based suppressor strains which carry a ColE1 origin of replication.

Following DNA ligation and transformation into competent cells (*E. coli* JM109, Promega), the colonies were screened by PCR to identify those clones which contained the mutated *pac*. Approximately 44 % contained the *pac*-TAG146 gene (Figure 7.3). Plasmid DNA was then isolated from the clones identified by PCR screening and further verified by restriction digest with *Spe*I restriction endonuclease (Figure 7.4). This clearly identified clones with opposite orientations of insert (Figure 7.5). PA-TAG146 was transformed into the amber suppressor strains as described in the manufacturers' instructions.

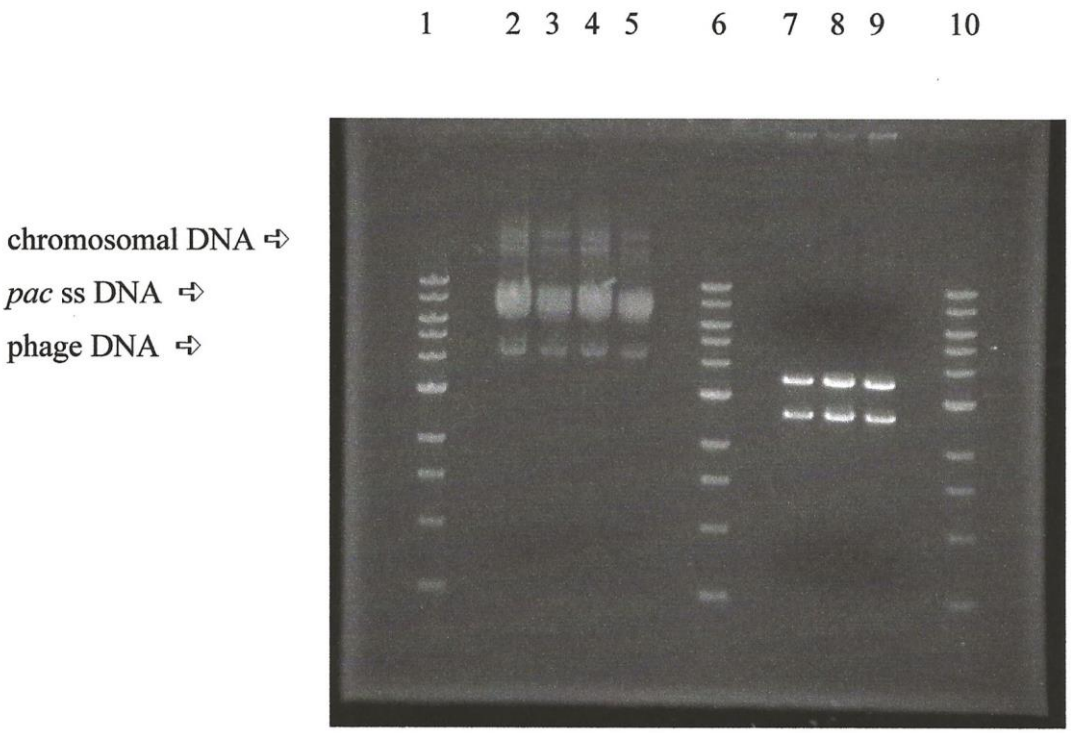


Figure 7.2 A DNA agarose gel showing PGA ssDNA and pBS-TAG146. *Lanes 1, 6, and 10: 1Kb DNA ladder (NEB), Lanes 2-5: pac ssDNA. Lanes 7-9: Positive clones of the mutant pBS-TAG146 cut with the restriction enzyme Spe*I.

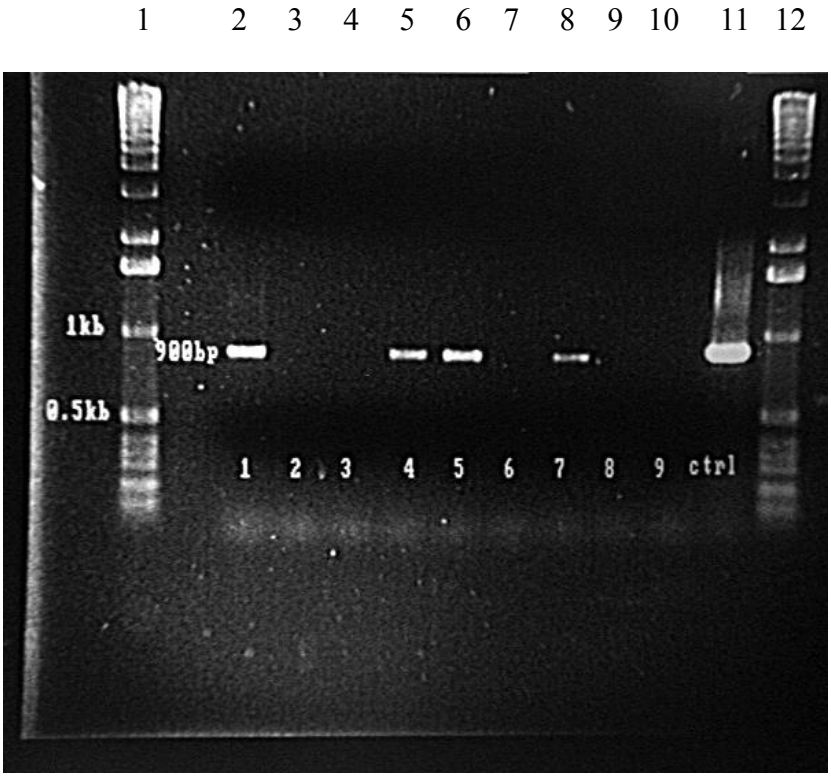


Figure 7.3 Colony PCR screening to identify positive clones containing the mutated pac gene. *Lane 1 and 12: 1kb ladder. Lanes 2-10: Colony PCR of colonies #1-9. Lane 11: Control PCR of PA-BS.*

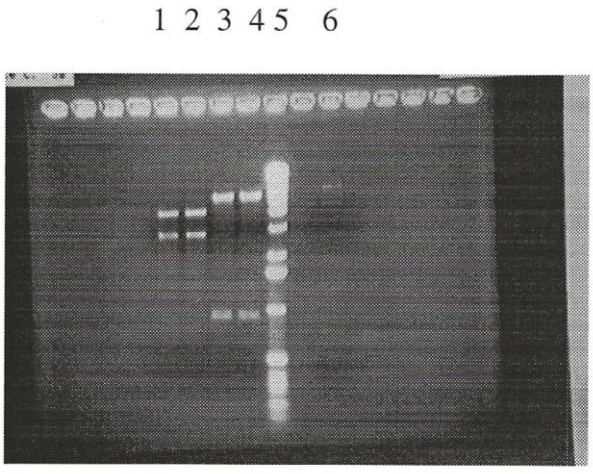


Figure 7.4 Restriction digest with ClaI/SpeI to identify positive clones containing the mutated pac gene in pACYC184. *Lane 1 : Clone #1 of PA-TAG146. Lanes 2. Clone #4 of PA-TAG146. Lane 3: Clone #5 of PA-TAG146. Lane 4: Clone #7 of PA-TAG146. Lane 5: 1 Kb DNA ladder. Lane 6: ClaI/Dra*III *digest of pA1.*

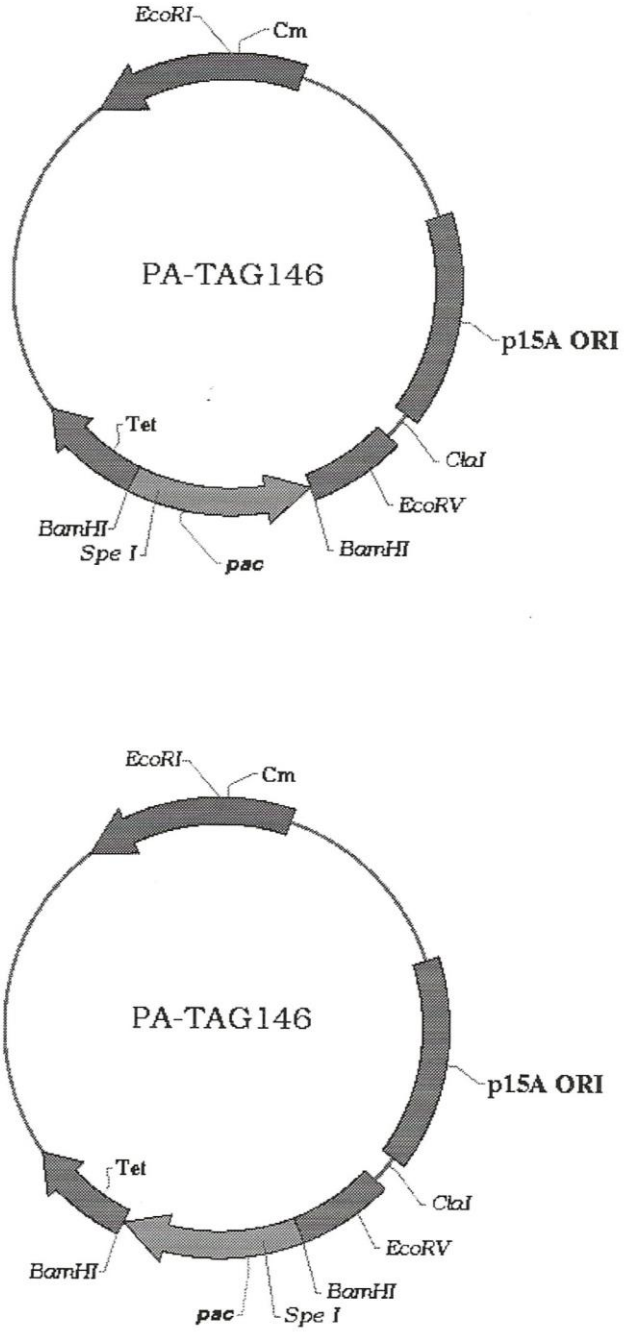


Figure 7.5 A schematic diagram of the plasmid PA-TAG146 showing the two orientations of the pac-TAG146 gene. *The Spe*I *restriction shows the position of the amber stop codon.*

7.2.2.1 Expression of suppressors mutants in *E. coli* K strain XAC

Expression of *pac* amber suppressor mutants was unsuccessful with the host *E. coli* K XAC strains provided in the mutagenesis system (Figure 7.6a). These strains contain either the chromosomal or plasmid constructs encoding the amber suppressors and has a chromosomal amber mutation in the arginine biosynthetic pathway (*arg*E-am) which allows selection for amber suppression activity when grown in minimal media. This provides a control to assure the functionality of the suppressor in these strains. The mutants were also grown in M9 minimal media to ensure that selection for suppressor activity was not to blame for lack of expression of the amber suppressed mutants (Figure 7.6b).

pac is found only in the *E. coli* W strain and expression in a K strain is incompatible and may be due to poor recognition of the promoter sequence or differing strain characteristics. SDS-PAGE analysis of the periplasmic fraction of each amber suppressor mutant showed no detectable expression of the PGA mutants (Figure 7.6). Furthermore, the periplasmic fraction from each mutant was assayed with NIPAB and examined visually for any colour change following the release of the chromophore 3-amino-6-nitrobenzoic acid. Only incorporation of the wild type phenylalanine amino acid showed any detectable activity, albeit very low, confirming the results shown by SDS-PAGE. Stresses induced by amber suppression in combination with a less than ideal host strain represent the most likely explanation for the poor expression.

7.2.2.2 Expression of plasmid based suppressors in BL21 (DE3)

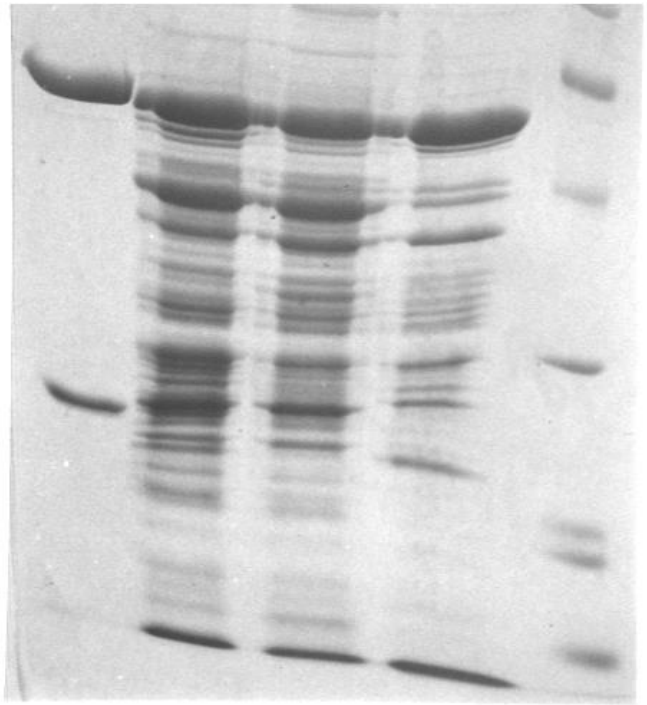
The poor expression shown by the XAC cells resulted in an attempt to express the amber mutants in the wild type host strain BL21 (DE3). The limitations to this approach were that only the synthetic constructs encoding amber suppressors on ColE1 plasmids could be used, as the remaining amber suppressors are chromosomally encoded. The DNA encoding the plasmid based amber suppressor tRNAs were isolated from the *endA*1 strain *E. coli* XAC as described in Chapter Two. The constructs are based on the plasmid pGFIB-1 (Norman*ly et a*l., 1986b) and the plasmid DNA for each amber suppressor construct is shown in Figure 7.7. The plasmids were isolated from each amber suppressor XAC strain as described. The plasmid PA-TAG146 and each of

the plasmid based amber suppressor tRNA constructs were co-transformed into the expression strain BL21 (DE3).

To determine if the expression of the mutants was possible, incorporation of the wild type phenylalanine amino acid by amber suppression was compared with expression of the non-mutated *pac* in BL21 (DE3) host cells. The level of *pac*-TAG146 was compared in both orientations (Figure 7.8). The orientation of wild type *pac* in the plasmid pA1 is the same as PA-TAG146, clones 1 and 4. The expression of the amber Phe mutant was better in the opposite orientation to that of pA1 as shown by amber-Phe expression in PA-TAG146 clone 5. The *pac*-TAG146 construct was chosen in this orientation for expression trials of further amino acid substitutions. In this expression host, there were comparable levels of PGA to wild type.

Expression of the amber suppressor protein variants is shown in Figure 7.9. The periplasmic extracts of the amber suppressor mutants were each assayed with NIPAB. The activity was poor for all the mutants, with only the His mutant measurable by spectrophotometric assay. The relative activity of the remaining mutants was assayed by adding equivalent amounts of protein to the chromogenic substrate and leaving the samples on the bench until a colour change was observed. This took approximately two hours for the Cys mutant in comparison to the Gly mutant which took overnight. The relative enzyme activity of each mutant in descending order was: His>Cys>Pro>Gly.

(a) 1 2 3 4 5



(b) 1 2 3 4 5 6 7 8

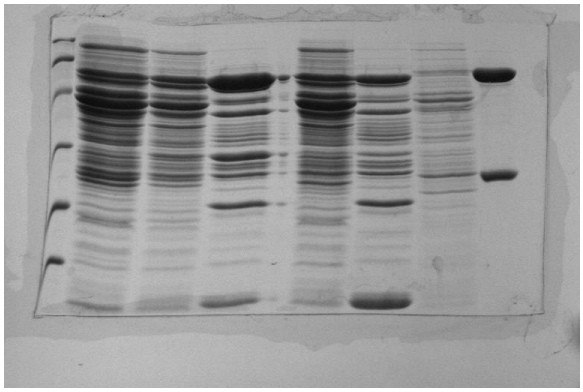


***Figure 7.6 (a) Expression of amber mutants in E. coli XAC K strain using LB media**. Lane 1: Purified PGA. Lane 2: Amber-Phe. Lane 3: Amber-Gln. Lane 4: Amber-His. Lane 5: Protein markers. **(b) Expression of amber mutants in E. coli XAC strain using minimal media**. Lane 1: Protein markers. Lane 2: Amber-Gln in M9 minimal media. Lane 3: Amber-Gly in M9 minimal media. Lane 4: Amber-Gly in LB media. Lane 5: Amber-His in M9 minimal media. Lane 6: Amber-His in LB media. Lane 7: pA1 expression in LB (24h). Lane 8: Purified PGA.*

1 2 3 4 5 6 7 8 9

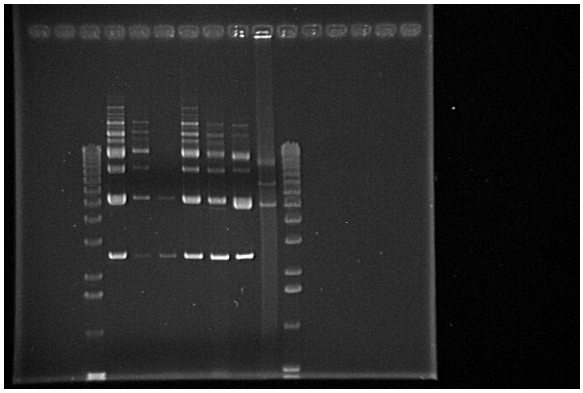


Figure 7.7 A DNA agarose gel showing the plasmid based amber suppressor tRNAs. *Lane 1 and 9: 1kb DNA ladder (Gibco BRL). Lane 2: pAMBER-Cys. Lane 3: pAMBER-Gly. Lane 4: pAMBER-His. Lane 5: pAMBER-Pro. Lane 6: pAMBER-Phe. Lane 7: PA-TAG146 + pAMBER-Phe. Lane 8: pA1.*

1 2 3 4 5 6

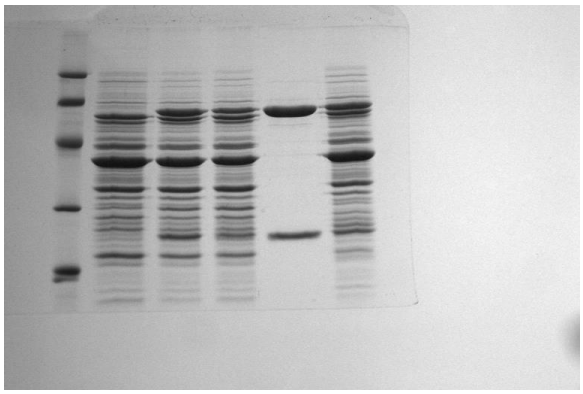


Figure 7.8 Protein expression of PA-TAG146 by amber-Phe suppression*. Lane 1: Markers (Biorad). Lane 2: Amber-Phe tRNA plasmid. Lane 3: Amber suppressed Phe mutant (+). Lane 4: Amber suppressed Phe mutant (-). Lane 5: Purified wt PGA. Lane 6: wt PGA.*

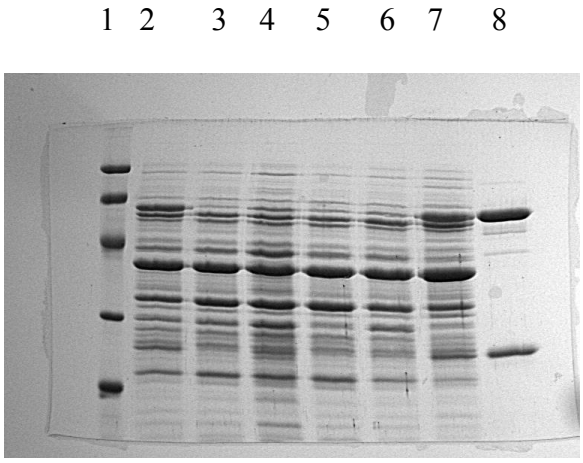


Figure 7.9 Protein expression of amber suppressed mutants*. Lane 1: Protein Markers. Lane 2: Phe suppression. Lane 3: Gly suppression. Lane 4: His suppression. Lane 5: Pro suppression. Lane 6: Cys suppression. Lane 7: wt expression. Lane 8: purified PGA.*

7.3 Purification of His A:146

The Phe A:146 His mutant (His^{A146}) was purified using a monoclonal antibody (mAb) affinity column. The mAb specific to PGA (mAb IV F 19) was immobilised on cyanogen bromide (CnBr) activated sepharose (Kasche *et al*., 1994). The antibody mAb IV F 19 was provided by Dr Volker Kasche and the mAb affinity column was prepared and provided by Dr. Lorraine Hewitt.

Protein expression and the extraction of the periplasmic fraction of the His^{A146} mutant was as described in Chapter 2. The addition of ampicillin to maintain selection of the Amber-His tRNA plasmid construct was particularly important especially since the cells were grown to high cell densities where the ampicillin levels can fall below the concentration sufficient to maintain selection for the plasmid carrying the Amber-His suppressor. Ampicillin was added (100 μg/ml) twice daily when using this plasmid. The protein was buffer exchanged into 100 mM potassium phosphate buffer pH 7.0 and loaded onto the bioaffinity column equilibrated in the same buffer (1 ml/min). The

column was then washed in three column volumes (9ml) of the buffer and eluted in 100 mM potassium phosphate buffer pH 4.0 as 3 ml fractions into 200 mM potassium phosphate buffer pH 8.0 (2 ml).

Purification of previous mutants and wild type PGA followed a 3-step procedure. Following isolation of the periplasmic cell fraction the protein is purified by ammonium sulphate fractionation, hydrophobic interaction chromatography and finally anion exchange chromatography. Loss of protein through several purification steps is often unavoidable. Since the expression of the amber suppressor was low, it was important to optimise the purification and final protein yield if further protein characterisation was to be feasible. Bioaffinity chromatography provided a one-step purification of the amber suppressor mutant and hence maximised the yield of protein obtained. Purification of wild type PGA gives typical yields of 7 mg/L of cell culture. From the gels, the level of mutant expression appeared to be about ten-fold lower but the final yield after purification of this mutant was 2.5 mg/L, similar to that of the Ala:B^{241} in Chapter 4, and only three times lower than wild type, reflecting the advantage of the antibody column purification. .

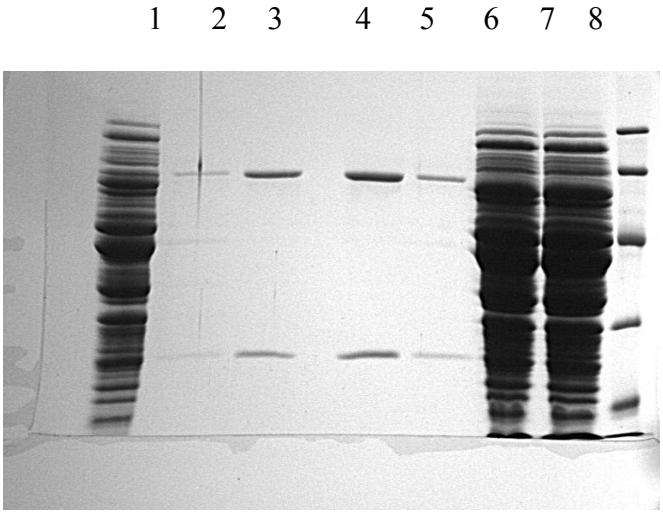


Figure 7.10 Purification of the Histidine amber suppressor mutant by bioaffinity chromatography. *Lane 1: Periplasmic extract of HisA146. Lane 2 - 5: Eluted fractions of HisA146. Lane 6 and 7: Flow through of unbound protein. Lane 8: Protein markers.*

7.4 Kinetic data analysis

Kinetic analysis was carried out making the usual steady-state assumptions. To ensure that a steady-state condition exists, initial velocity measurements were made when the velocity was linear both with time and enzyme concentration. The initial velocity of an enzyme catalysed reaction is related to the substrate concentration by the Michaelis-Menten equation [1]:

$$v = \frac{[S]V_{\max}}{[S] + K_m} \qquad [1]$$

$[S]$ = substrate concentration

$V_{\max}$ = maximal rate

K_m = Michaelis-Menten constant

v = initial rate of reaction

The Michaelis-Menten equation is defined mathematically as a rectangular hyperbola. The K_m and V_{max} parameters define the rectangular hyperbola relating v to $[S]$. The enzyme kinetic parameters V_{max} and K_m were determined by following the hydrolysis spectrophotometrically of the chromogenic substrate NIPAB (Kutzbach and Rauenbusch, 1974). Penicillin acylase hydrolyses NIPAB to yield phenylacetic acid and 3-amino-6-nitrobenzoic acid which has a molar absorption coefficient of 9090 M^{-1} cm^{-1} at λ = 405nm. The rate of reaction for the hydrolysis of NIPAB by both wild type and His^{A146} were determined as follows: 10 µl of enzyme solution (~1µg of wild type PGA; ~15.8 µg of His^{A146}) was added with immediate mixing, to a cuvette containing 20 µl of NIPAB solution (freshly prepared in 50 % ethanol) in a 1 ml volume containing 100 mM potassium phosphate buffer pH 7.5. The increase in absorbance at λ = 405nm was measured using a Shimazdu UV-160A spectrophotometer and the reaction monitored for 60 sec. The procedure was repeated for ten different substrate concentrations in the range 10 - 200 µM, approximately 0.5 to 5 K_M. The derived value for V_{max} with the units Δ Abs/min, was then converted using the Beer-Lambert law (ε = 9090 M^{-1} cm^{-1}, [2]) to give units of M^{-1} s^{-1}. From this the first order rate constant k_{cat} was calculated using [3].

$$A = \varepsilon.c.d \qquad [2]$$

$$k_{cat} = V_{\max}/[E]_o \qquad \text{Where } [E]_o = \text{total enzyme concentration} \qquad [3]$$

7.4.1 Calculation of kinetic parameters

The kinetic data were fitted to the Michaelis-Menten model using non-linear least-squares regression analysis with the program Origin v2.8 (MicroCal Software Inc), the only reliable means to obtain kinetic parameters of an enzyme catalysed reaction. This avoids the problems of exaggeration of error associated with linear plots, e.g., Lineweaver-Burk plot. Values of k_{cat} and K_m were determined from experimental data measured in triplicate.

7.4.2 Kinetic data for the His A:146 mutant

To make a valid kinetic analysis of the mutant His^{A146}, kinetic parameters were determined for wild type PGA under similar experimental conditions. The steady-state kinetic constants determined for wild type PGA are in agreement with results published elsewhere (Morillas *et al.*, in preparation) and a simple rate versus substrate concentration plot is presented in Figure 7.11. However as an active site titration of PGA was not carried out, the value of $[E]_o$ is not a true estimate of active enzyme and this introduces additional error into the estimate of k_{cat}.

This does not detract from making a valid comparison with the His^{A146} mutant. 15.8 μg of His^{A146} were required to ensure steady-state conditions. The initial reaction velocity versus varying substrate concentration is presented in Figure 7.12. The kinetic parameters V_{max}, K_M, and k_{cat} for PGA and the mutant are presented in Table 7.1.

The k_{cat}/K_M is a measure of how rapidly an enzyme works at low substrate concentration and allows comparison of the activity of enzyme and mutants for defined substrates. The k_{cat}/K_M value for NIPAB for the His^{A146} mutant shows a 20-fold drop in specificity for this substrate in comparison with wild type enzyme. The major contribution to this decrease is the 10-fold decrease in the turnover number (k_{cat}) which suggests an increase in the activation energy required for formation of the transition state. The 2-fold

decrease in K_M also suggests that mutation of Phe A:146 decreases specificity towards the P1' subsite.

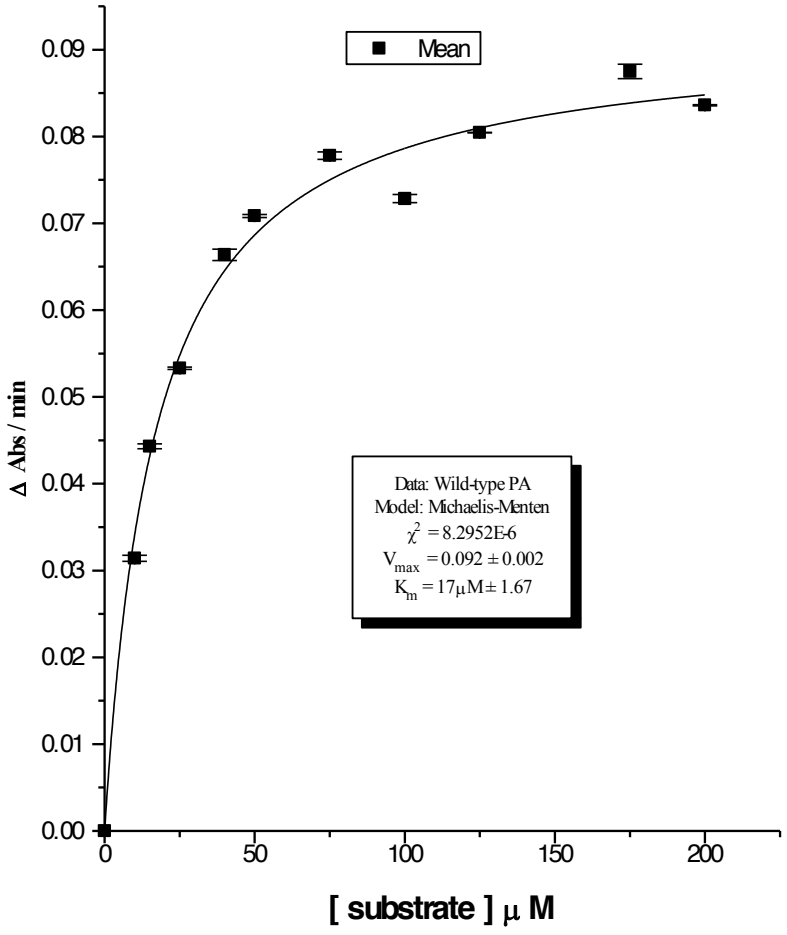


Figure 7.11 The initial reaction velocity against varying substrate concentrations for wild type PGA.

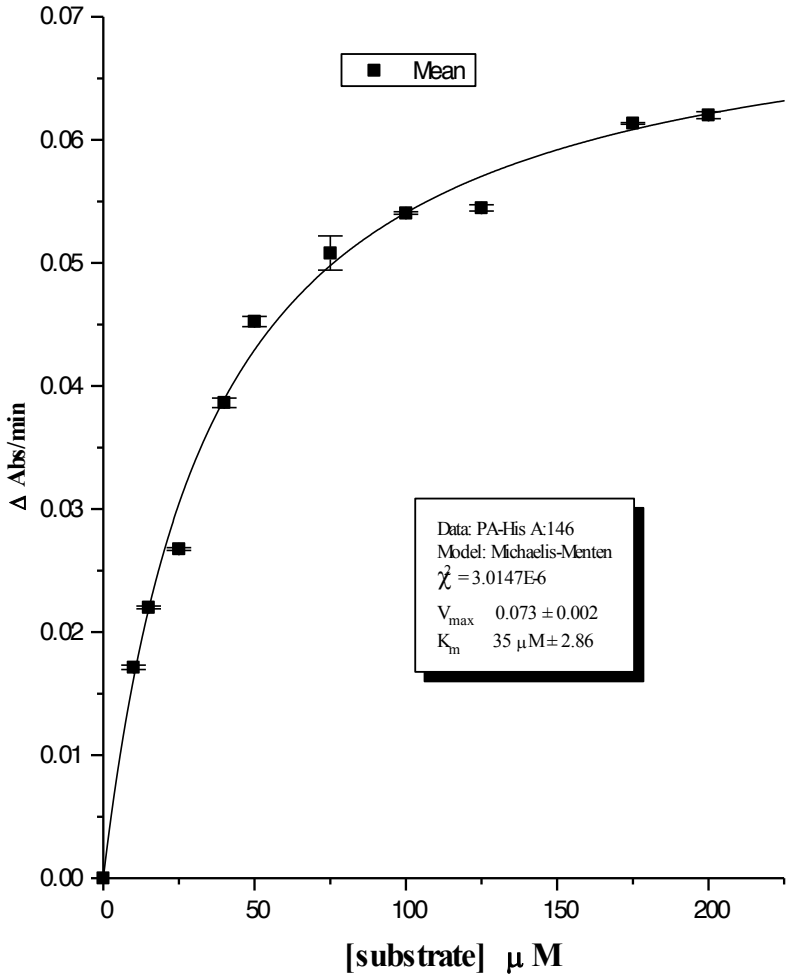


Figure 7.12 The initial reaction velocity with varying substrate concentration for the mutant His^{A146}.

Penicillin G acylase	V_{max} (Δ Abs/min)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
wild type	0.092 ± 0.002	17 ± 2	14.15	8.3 x 10^5
His^{A146}	0.073 ±0.002	35 ± 3	1.46	4.2 x 10^4

Table 7.2 Steady-state kinetic parameters of NIPAB hydrolysis by wild type and mutant PGA. *Initial velocity measurements were made at 24 °C with NIPAB in 100 mM phosphate buffer, pH 7.5 with wt PGA and the mutant His^{A146}.*

7.5 Results and Discussion

Amber suppression was employed to screen a range of mutants to dissect the role of Phe A:146 in substrate specificity, on the basis of information gleaned from the crystallographic analysis of the ligand complexes, described in Chapters 5. The present results provide further support for the role of A:146 in determining substrate specificity.

In principle, amber suppression *in vivo* mutagenesis provides an easy scheme to analyse site directed invariants. The method should allow the introduction of any of twelve amino acids at a selected point. The inability to develop efficient tRNA suppressors for all 20 has restricted the choice to 12 of the naturally occurring amino acids for substitution. While the twelve are generally representative of the properties of the full 20, this is nevertheless a limitation of the present technology. The results here show both the advantage and the weakness of the approach in its present form.

For PGA, the system available from Promega was further limited by the *E. coli* K strain provided as the expression of variants was extremely low, even for the amber Phe suppressed variant corresponding to wild type. An important factor was the compatibility of the plasmid based expression system normally used with PGA with that

applied with the amber suppressor system. The strain was therefore changed to *E. coli* B already used successfully for PGA expression.

However this further reduced the possible amino acid substitutions to those seven encoded on plasmid constructs since the host strain provided was unsuitable for expression of the other five encoded on chromosomal DNA. This left seven possible substitutions one of which was the wild type Phe. In this strain, the Phe suppression was as good as the usual wild type expression providing evidence that amber suppression did not add further stresses to the cell's protein expression.

Four were selected for further study on the basis of their different and representative side chain characteristics. However, for all four mutants the expression levels were low, as has been previously observed for mutants obtained by more conventional means. Even wild type PGA is a only expressed at relatively low levels and any further demands such as mutations can be expected to have adverse effects on protein production. This reflects the complexity of PGA maturation, involving translation as a precursor with signal peptide, translocation into the periplasmic space and cleavage of the spacer peptide between the A and B chain of the active enzyme.

Crude cellular extracts allowed a rough estimate of kinetic activity, however the best, His, was the only one sufficiently expressed to be taken further for detailed kinetic analysis. His^{A146} retained significant activity on the substrate NIPAB although the activity and binding were reduced compared to wild type. This supported the hypothesis that Phe A:146 is important for activity, but leaves questions about the structure of the His and other mutants for future studies.

The k_{cat}/K_M is a measure of how rapidly an enzyme works at low substrate concentration. It is a means of comparing and studying the specificity of an enzyme for a particular substrate. A comparison of the k_{cat}/K_M for wt PGA and the mutant His^{A146} shows a 20-fold drop in catalytic effectiveness for the substrate NIPAB by this mutant. The major contribution to this decrease is the 10-fold decrease in the turnover number (k_{cat}) which suggests an increase in the activation energy required to form the transition state. The decrease in activity can be summarised by a weaker binding of the Michaelis-Menten complex (2-fold) and by weaker binding of the transition state (10-fold).

Preliminary kinetic and mutagenesis studies at Phe A:146 were previously carried out to study the substrate specificity in *E. coli* PGA (Janss*en et a*l., 1998). The mutant Phe A:146 Tyr was shown to have higher affinity for α-substituted phenylacetic acid derivatives than the wild type enzyme, and stereospecificity was not affected by the mutation. Kinetic characterisation of the Phe A:146 Tyr mutant showed an increased K_m (4 mM) and decreased k_{cat} (0.8 s^{-1}) for ampicillin hydrolysis but was inactive in the synthesis reaction of ampicillin. They concluded that Phe A:146 plays an important role both in binding of the side chain and in binding of the β-lactam nucleus. This is confirmed by the present results for His^{A146}.

7.6 Conclusions

Expression of PGA is not trivial even in its optimised set-up for the wild type and this poses severe challenges for the amber suppressor approach. Nevertheless the system has already proved useful both in deducing any effects the mutation may have on enzyme processing, which can be a stumbling block to mutagenic analysis of enzyme function and in providing a simple means to provide a functional description of an active site residue.

Determining the structure of His^{A146} may shed light on the observed kinetic data. In addition, the pH profile for the activity of the mutant should be determined. A detailed kinetic analysis of a range of mutants and their 3-D structure at A:146 may further our understanding of the role of this residue in substrate specificity.

Chapter Eight Summary and Future Work

Penicillin acylase is an important enzyme in the production of semi-synthetic penicillins. A wealth of information on this key enzyme was already available at the start of this thesis, including its biochemical properties, the structural organisation of its gene and the 3-D crystal structure of the native enzyme and several complexes. The background was given in Chapter 1.

The 3-D structure had provided major insights and revealed a novel catalytic mechanism, forming a sound starting point for further analysis through protein engineering. A variety of approaches, involving both 'rational' design and natural or randomly induced mutants has been successfully combined to study PGA structure : function relationships. PGA was shown to be a member of the then recently identified family of Ntn hydrolases.

The present studies were primarily directed towards an improved understanding of the mechanism and specificity of the mature enzyme. The techniques used were briefly introduced in Chapter 2.

PGA is a large enzyme with molecular weight 86 kDa. The structure had originally been determined to 1.9 Å resolution at room temperature using protein purified by conventional methods. The first objective, described in Chapter 3 was to obtain a crystal form amenable to cryogenic freezing, this was successfully achieved using ethylene glycol in the mother liquor and led to a wild type structure refined at 1.3 Å resolution. Streak seeding was an essential factor in maintaining a reproducible supply of crystals less expensive in terms of the amount of protein required, while cross seeding provided a simple method for crystallising mutant PGA variants. The structure at 100 K was determined using state-of-the-art techniques, including synchrotron radiation and maximum likelihood refinement. It is very accurate for a protein of this size.

A substantial number of ethylene glycol molecules were included in the refined model, several of these occupying the active site of the enzyme. While the presence of these

might be expected to cause problems in the binding of ligands, this was clearly not important for strongly binding moieties such as Pen G and PGSO, see below.

Chapter 4 described experiments designed to produce inactive mutants of PGA which would allow the binding of substrates or substrate analogues to probe the geometry of the active site and shed light on the mechanism. Two residues were chosen as the target for mutagenesis, the catalytic Ser B:1 and a residue in the oxyanion hole, Asn B:241. Successful mutation of Ser B:1 to cysteine provided mature and correctly folded enzyme with zero activity. Unfortunately structural studies of this mutant showed that it had also lost the ability to bind substrate perhaps through oxidation of the cysteine sulphur. Attention was therefore concentrated on the Asn B:241 Ala mutant, which also expressed reasonably well and showed negligible activity.

Crystals of the inactive Ala B:241 mutant were isomorphous to those of the native enzyme. Chapter 5 describes the 3-D structure determination of this mutant, its complexes with Pen G and PGSO, together with that of the wild type PGSO complex. The structures bring out a number of key points. Firstly, Asn B:241 in the oxyanion hole is important for positioning the substrate in the active site as well as in its presumed role of stabilising the tetrahedral intermediate. Secondly, the position of the penam moiety is well defined in spite of its rather non specific interactions with the enzyme. It is only poorly defined in the wild type PGSO complex where partial hydrolysis may have occurred. Thirdly, residues Met A:142 to Phe A:146 are distributed between two alternative conformations in the various structures and this is almost certainly related to ligand binding. Of these residues Phe A:146 is most directly involved and occupies a vital position at the neck of the active centre, between subsites S1 and S1'. Fourthly, the phenylacetyl moiety is buried in the same pocket previously observed for PAA, but its exact orientation varies between the wild type and inactive mutant structures.

Together with the covalent benzylboronate structure determined previously (Done, 1996) these results identify the likely position of substrate in the active site and allow some speculation on the mechanism. However a question mark remains over the existence of a water in a position appropriate for a catalytic role. Chapter 6 describes the expression and 3-D structure of mutant Leu B:71 with altered specificity. The

altered kinetic properties are rationalised by the structure, although structures of complexes of this mutant would be required to further clarify the role of this residue.

In Chapter 7, a good candidate was selected for further mutagenesis. Phe A:146 was subjected to multiple mutation using the amber suppression approach. Only six of the hoped-for 12 amino acids could be expressed at this point, all at low levels other than the naturally occurring phenylalanine. The kinetics of the histidine mutant alone were characterised in some detail and showed a reduced activity consistent with the importance of this site. The results show the difficulty in applying the amber suppression technique to a low expression protein such as PGA. Modifications in the plasmids and strains used would be necessary to advance this analysis and make it more easily applicable.

In pursuit of a cephalosporin acylase – fact or fantasy

The work described has covered many aspects of the protein engineering design cycle. The attainment of an enzyme to hydrolyse cephalosporins is still a long way off. When I began my PhD the goal was to engineer a PGA that would have the capacity to hydrolyse cephalosporin. In order to engineer new substrate specificity in PGA my first objective was to fully understand the substrate specificity already instilled within the catalytic framework of PGA. The work described in this thesis has led to a better understanding of the structural basis of substrate specificity, but further work is required to achieve the goal of rational engineering of an enzyme capable of cephalosporin hydrolysis.

Future Work

- A different crystallisation approach is needed to avoid competitive binding of cryosolvents. This is an absolute requirement to study substrates with lower binding affinities
- Further analysis of substrate binding to an altered specificity mutant is needed to understand the structural basis of the altered specificity of the mutants. The

structure of a glutaryl-(L)-leucine complex with an inactive double mutant Leu B:71/Ala B:241 will help elucidate improving specificity towards cephalosporin mimics when B:71 is mutated compared to wild type PGA. Preliminary investigation of a double mutant has shown reasonable protein expression for structural studies.

- The flexibility of the helical region around A:142 to A:146, in concert with Phe B:71, provide a rationale for the broad specificity of PGA for the S1' substrate subsite. In contrast its specificity is uncompromising for the S1 subsite and is tailored for the phenylacetyl moiety. Phe A:146 may be the focal point which ties down the specificity pocket. Mutation screening at this residue by amber suppressor mutagenesis shows that enzyme activity is severely disabled by amino acid substitutions at this site. A detailed examination of mutations at residue A:146 and at Met A:142 may provide the key to relaxing the specificity of the S1 subsite and both biochemical and structural characterisation is needed.
- Chemical modification of the α-NH_2 group of the catalytic serine should provide further avenues of research to probe the catalytic mechanism and may also provide an inactive enzyme for studying substrate binding. One suitable reagent is acetic anhydride which chemically acetylates free amino groups which would minimise disruption of the active site.

Appendix

Media

LB	Luria-bertani broth	
	Bacto-tryptone	10g/L
	Yeast extract	5g/L
	Sodium chloride	5g/L
2TY	Bacto-tryptone	16g/L
	Yeast extract	10g/L
	Sodium chloride	5g/L
M9 minimal media	Na_2HPO_4	6g
	KH_2PO_4	3g
	NaCl	0.5g
	NH_4Cl	1g

Add H_2O to 1 litre volume. Autoclave and cool to 50 °C before aseptically adding the following sterile filtered solutions:

1 M $MgSO_4$	2 ml
1 M $CaCl_2$	0.1 ml
20 % glucose	10 ml
0.1 M thiamine-HCl	10 ml
0.1 M methionine	10 ml
0.1 M proline	10 ml

Buffers & Solutions

1M MOPS pH 7.2

3-*N*-morpholinopropanesulfonic acid
pH adjusted with 0.5M NaOH.

PEG/NaCl soln - (100ml)

Polyethylene glycol 6K	20.0g
Sodium chloride	14.6g

made up to 100ml in deionised water
and sterile filtered with a 0.2µ amicon filter.

1M TRIS

Tris base

TE

10 mM Tris-HCl (pH 8.0), 1mM EDTA pH 8.0

x50 TAE

Tris base	242g
Acetic acid	57.4ml
0.5M EDTA	20ml

x6 DNA sample buffer

0.25% bromophenol blue
0.25% xylene cyanol FF
30% glycerol in water
store at 4 °C

x4 SDS-PAGE sample buffer stock

TRIS	6.05 g
SDS	8 g
EDTA	7.44 g
glycerol	40 ml
final volume	100 ml

sample buffer stock	1 ml
β-mercaptoethanol	200 µl
bromophenol blue (sat. soln)	50 µl

Bacterial Strains

BMH 71-18 *mut*S
thi, *sup*E, Δ(*lac-pro*AB), [*mut*S:Tn10] [F′, *pro*AB, *lacI*^{q}ZΔM15]

JM109
*e*14-(McrA-), *end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R1117 (rk-, mk+), *rel*A1, *sup*E44, lambda-, Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *lacI*^{q}ZΔM15]

BL21 (DE3)
F-, *omp*T, *hsd*S_B, (r_B-, m_B-), *dcm*, *gal*, λ(DE3)

XAC (Amber suppressor strain)
ara, Δ(*lacpro*), *gyr*A, *arg*E-am, *rpo*B, *thi*

XL-1 Blue MRF′
Δ(*mcr*A)183, Δ(*mcr*CB-*hsd*SMR-*mrr*)173, *end*A1, *sup*E44, *thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac* [F′, *pro*AB, *lacI*^{q}ZΔM15,Tn10 (Tet^{r})]

E. coli pac sequence

```
 G N I G Y V H T G A Y P D R Q S G H D P R L P V P G T G K W D W K G L L P F E M (514)
GGCAATATTGGTTATGTTCATACTGGTGCTTATCCAGATCGTCAATCAGGCCATCATCCGCTTCCTGGTACGGGAAAATGGGATCTGAAAGGGCTATTGCCTTTGAAATG 1920

 N P K V Y N P Q S G Y I A N W N N S P Q K D Y P A S D L F A F L W G G A D R V T (554)
AACCCTAAGGTGTATAACCCCAGTCGGGATATATTGCTAGTGGAACAATTCTCCCGAAAAGATTATCCGCTTCAGATCTGTTTGGGGTGGTGCAGATCGGCGTTACG 2040
                                                                  Bgl II
 E I D R L L E Q K P R L T A D Q A W D V I R Q T S R Q D L N L R L F L P T L Q A (594)
GAGATCGACCGACTGCTTGAGCAAAAGCCACGCTTAACTGCTGATCAGGCATGGGATGTTATTCGGCAAACCAGTCGTCAGGATCTTAACCTGAGGCTTTTTTACCTACTCTGCAAGCA 2160

 A T S G L T Q S D P R R Q L V E T L T R W D G I N L L N D D G K T W Q Q P G S A (634)
GCGACATCTGGTTTGACACAGAGCGATCCGCGTCAGTTGGTAGAAACATTAACAGGTTGGGATGGCATCAATTTGCTTAATGATGATAAAACCTGGCAGCAGCCAGGCTCTGCC 2280

 I L N V W L T S M L K R T V V A A V P M P F D K W Y S A S G Y E T T Q D G P T G (674)
ATCCTGAACGTTTGGCTGACCAGTATGTGAAGCGTACCGTAGTGGCTGCCGTACCTATGCCATTTGATAAGTGGTACAGCGCCAGTGGCTACGAAACAACCCAGGACGGCCCAACTGGT 2400

 S L N I S V G A K I L Y E A V Q G D K S P I P Q A V D L F A G K P Q Q E V V L A (714)
TCGCTGAATATAAGTGTTGGAGCAAAAATTTGTATGAGGCGGTGCAGGGAGACAAATCACCAATCCCAGCAGGCGGTTGATCTGTTTGCTGGGAAACCACAGCAGGAGGTTGTTGTTGGCT 2520

 A L E D T W E T L S K R Y G N N V S N W K T P A M A L T F R A N N F F G V P Q A (754)
GCGCTGGAAGATACCTGGGAGACTCTTCCAAAGCGCTATGGCAATAATGTCAGTAACTGGAAAACACCTGCAATGGCCTTAACGTTCCGGGCAAATAATTTCTTTGGTGTACCGCAGGCC 2640

 A A F E T R H Q A E Y Q N R G T E N D M I V F S P T T S D R P V L A W D V V A P (794)
GCAGCGGAAGAACGCGTCATCAGGCGGAGTATCAAACCGTGGAACAGAAACGATATGATTGTTTCTCACCAACGACAAGCGATCGTGCTTGGTGGGATGTGGTCGCACCC 2760
                 MIu I
 G Q S G I A P D G T V D K H Y E D Q L K M Y E N F G R K S L W L T K Q D V E A (834)
GGTCAGAGTGGGTTTATTGCTCCGATGGAACGTTGATAAGCACTATGAAGATCAGCTGAAAATGTACGAAAATTTTGGCCGTAAGTCGCTCTGGTTAACGAAGCAGGATGTGGAGGCG 2880

 H K E S Q E V L H V Q R ***                                          (846)
CATAAGGAGTCGCAGGAAGTGTTGCACGTTCAGAGATAATTAAGCCCCGAAAGCCCTCATATCTGGAGGAATATGAGGGTGATAAAGTCTCAGGTATTGTGATGTTGGAGGGCTGGTTAC 3000

TTTTTGAGTCCCGCAAACGCGCCGCAATCTCATCCTCCGGGAACTGAATCCCGATAAACACCATCGTGCTATGCGGTGTTTCATCGCCCCATGGCCTGTCCAGTCGGCGCTGTAGAGG 3120

CGTGGACCGCTGGAACAGCAGGCGGTTAGGTTCGCCGTCAATCCACAGGATCCCTTTGTAACGCAGCAGTTATATCCGCCGATTCCAGCAGGAGGTTTTCCACTCACCGGGAAACTTCGC 3240

TGATATCTACCGGGTAATCCAGTTCCACCACAATCGACGAAATATCGTTTGTTATCCGCGATAAAGTGGAAACCGCGGTTTGGTGCTGACGACGTTTCTTCCAGCATAAAACCGTTGG 3360
```

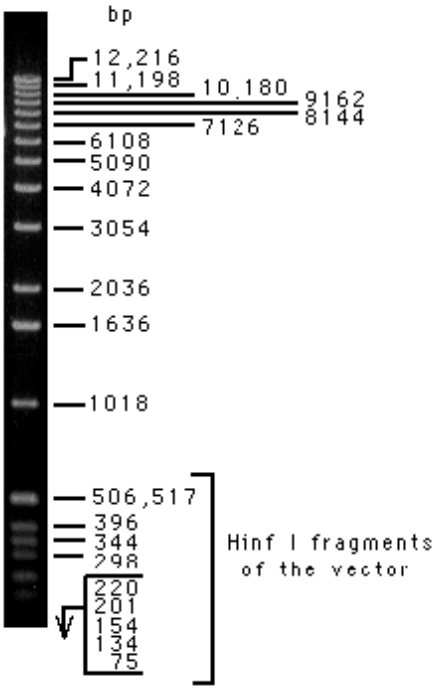
E. coli pac sequence

```
TGTTGAACAGCAGGCCCAGGTCGATCTCGCCGTGGGTGACGGTGTAGACCGGTGCACGGGCGTTGATGCGCGCCAGGCGTTCACGCAGTTTTTCTGCTTCGCCTGCGACGTCGGTTTAG 3480
                    inserted BamH I site
                    ....CGGATCC....
TCAGCAGATACGGTCGGCGTACCCGACCTGGCACTGGGCGATGGTCAACTGGTTCATCTGCTTATTGGCATGTACCGCATCCACCAGCGCAATCACGCCGTCCAGCAGGTAGCGCTGGCA 3600

TAAAATCTCATGGCAGAAAAAGGTTTGAATAATCGGGCCAGCATCGGCCATGCCCGGTGCATTTCAATGACCAGCGATCGAACTGAATATTGCCCTTGTCGACATTATCCAGCAGGTGCAG 3720

TAGCGCGTCCTCCAGCTCGTTGGAGCGCGAACAGCAGATGCAGCCGTTGGTCAGCGTTTTCATTCTCGTTTGCGCGATCACCAATTTCAATTTGATCATCAACGGAGACTTCGCCCGAATTC 3841
                                                                                                                    EcoRI
```

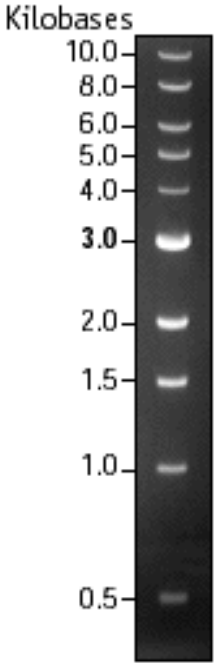
E. coli pac sequence

1 Kb DNA ladder sizes

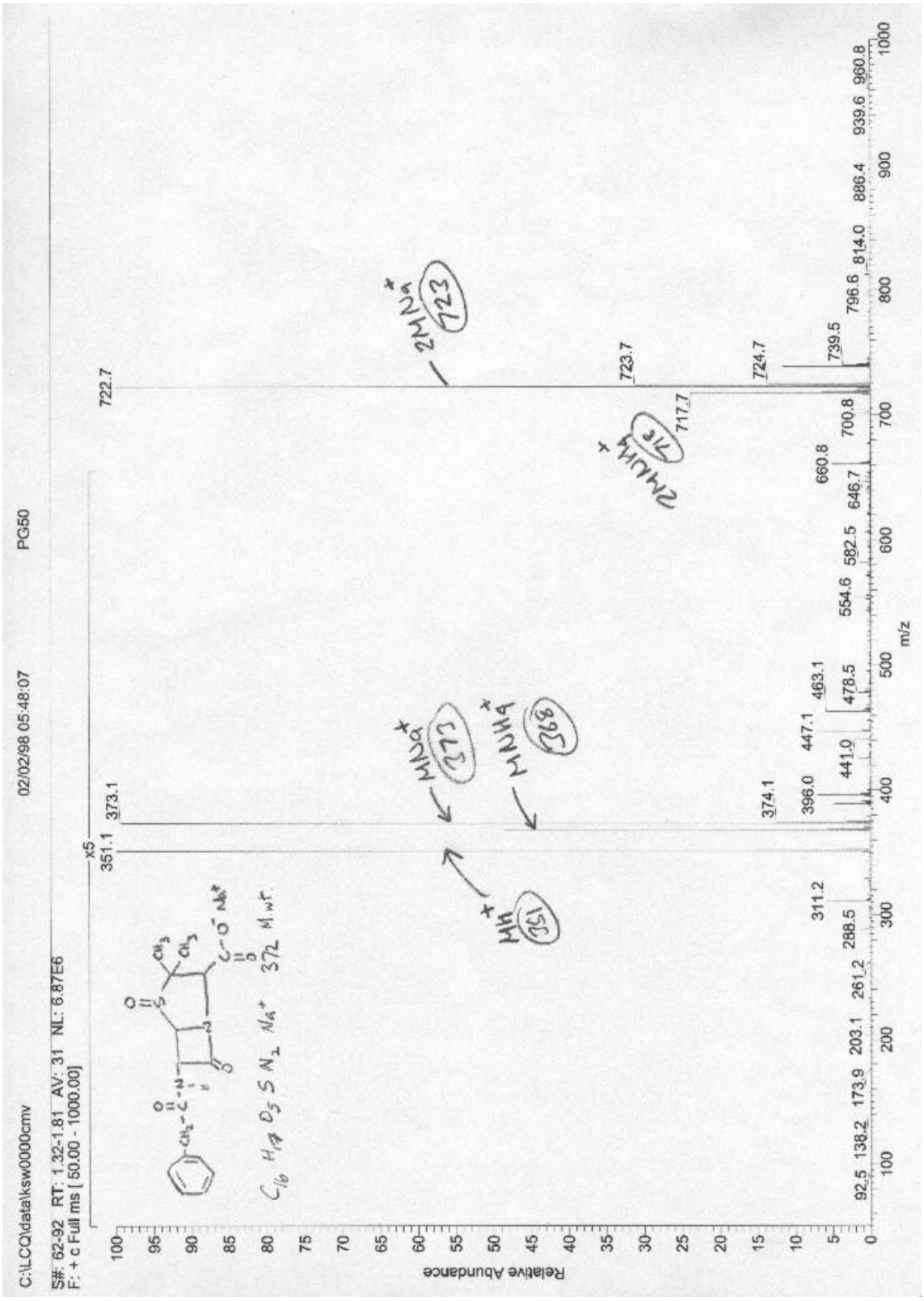
Gibco BRL 1 Kb DNA ladder

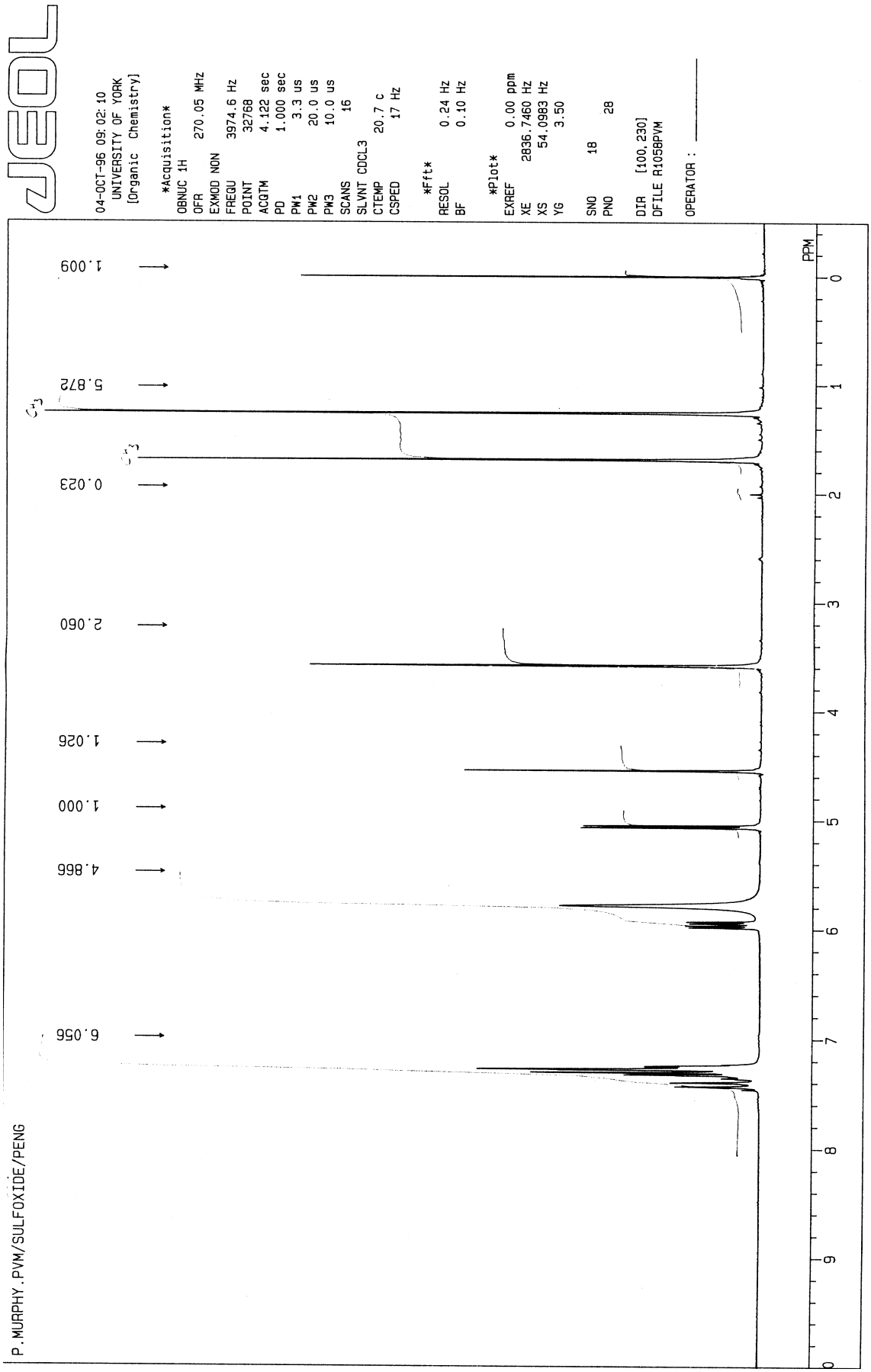


NEB 1 Kb DNA ladder

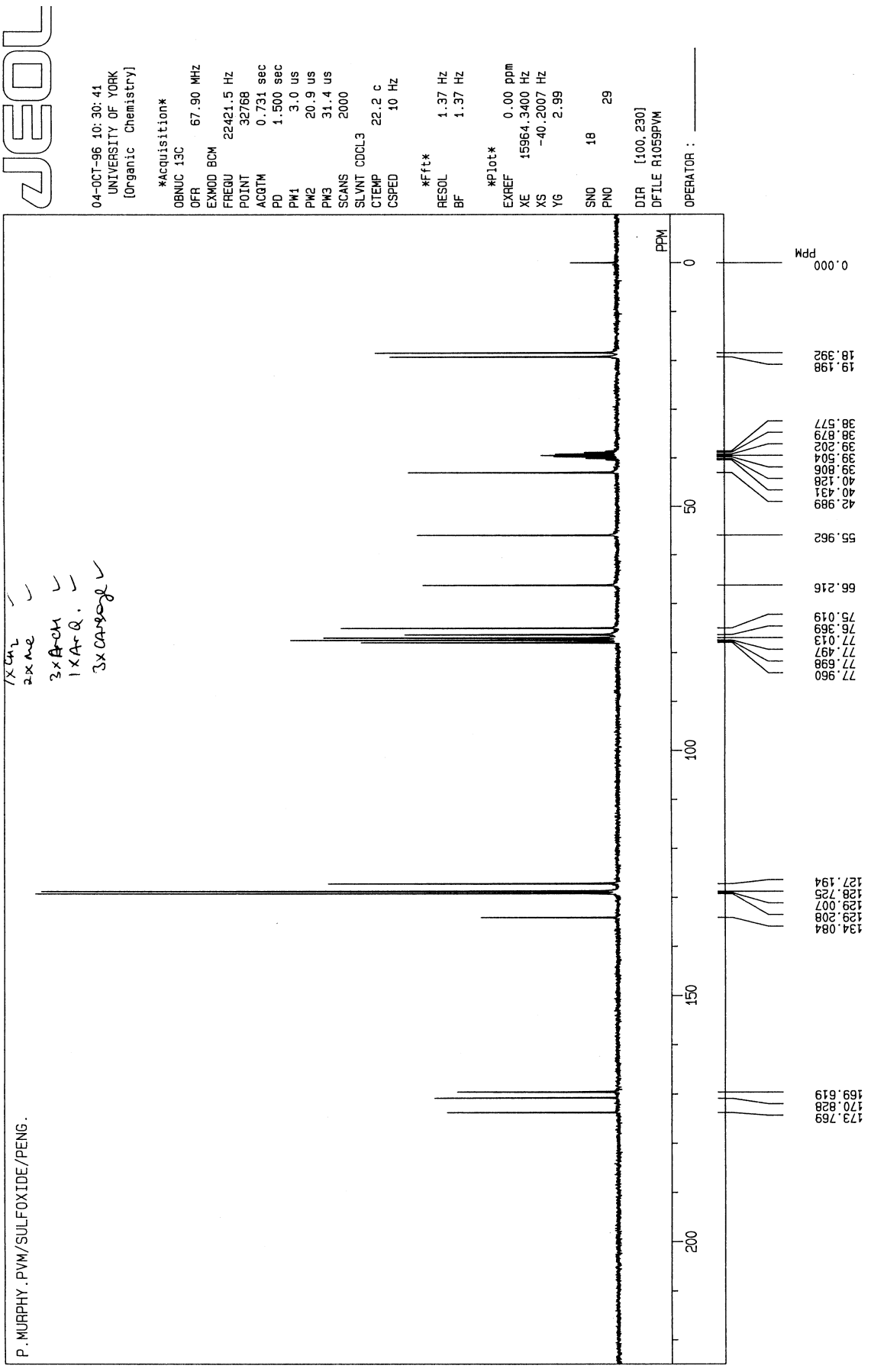


Mass spectra of PGSO

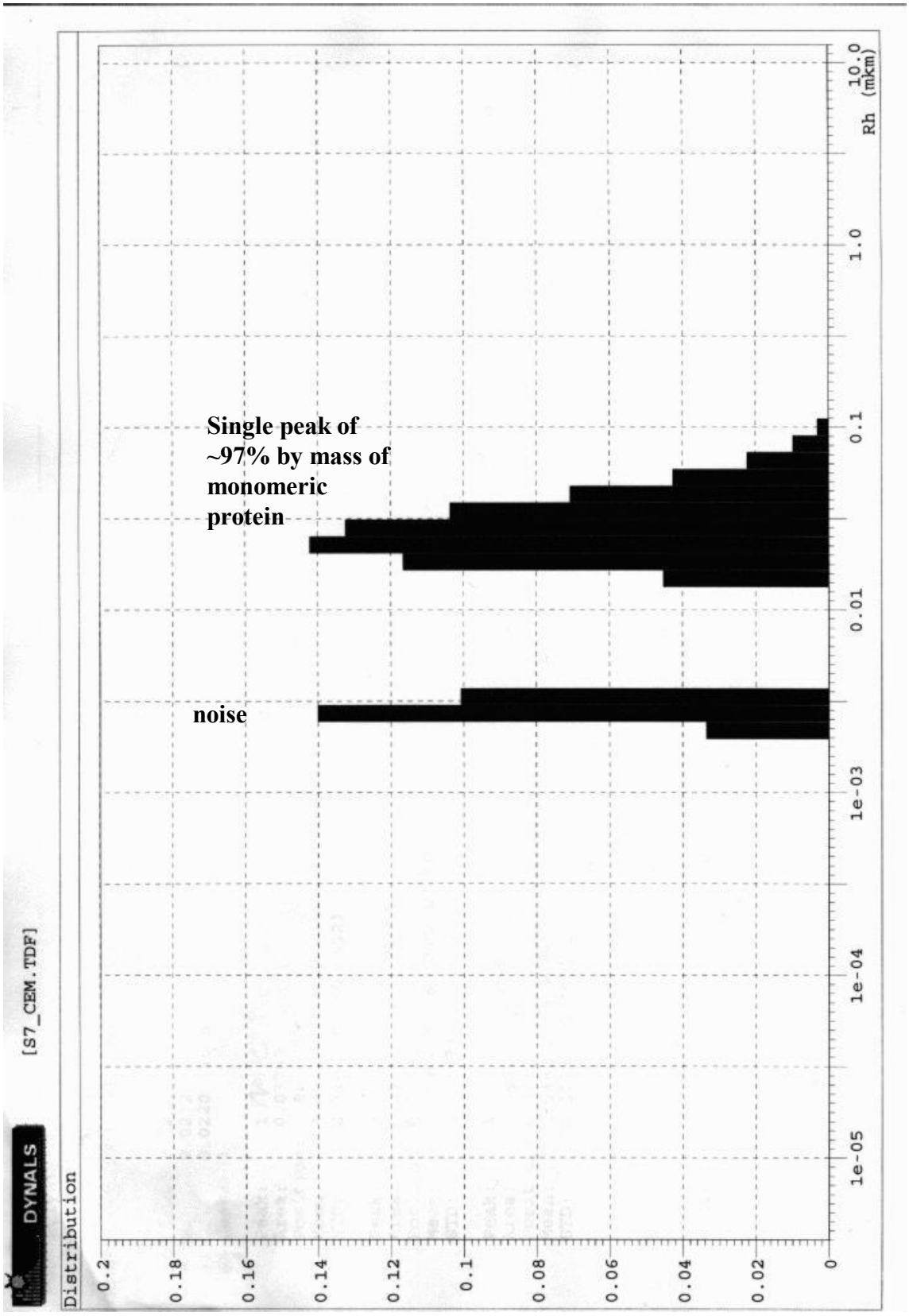




^{1}H NMR spectra of PGSO



^{13}C NMR spectra of PGSO



Dynamic Light Scattering (DLS) of Cys^{B1} protein

Bibliography

Alfani, F., Cantarella, M., Cutarella, N., Gallifucco, A., Golini, P., and Bianchi, D. (1997). Enzymatic conversion of cephalosporin C into glutaryl 7-aminocephalosporanic acid. A study in different reactor configurations. *Biotechnol. Letters* **19**(2), 175-178.

Allen, F.H. *et al.* (1979). The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr.* **B35**, 2331-2339.

Alvaro, G., Fernandez-Lafuente, R., Blanco, R.M., and Guisan, J.M. (1991). Stabilizing effect of penicillin G sulfoxide, a competitive inhibitor of penicillin G acylase: its practical applications. *Enzyme Microb. Technol.* **13**, 210-214.

Amemiya, Y. (1997). X-ray storage-phosphor imaging plate detectors: High-sensitivity X-ray area detector, in *Macromolecular Crystallography, Part A*, C.W.C. Jr. and R.M. Sweet, Editors. 1997, Academic Press: San Diego. 233-243.

Baldaro, E., Faiardi, D., Fuganti, C., Grasselli, P., and Lazzarini, A. (1988). Phenylacetyloxymethylene, A Carboxyl Protecting Group with Immobilized Penicillin Acylase Useful in Benzyl Penicillin Chemistry. *Tetrahedron Letters* **29**(36), 4623-4624.

Barbero, J.L., Buesa, J.M., Gonzalez de Buitrago, G., Mendez, E., Pez-Aranda, A., and Garcia, J.L. (1986). Complete nucleotide sequence of the penicillin acylase gene from *Kluyvera citrophila. Gene* **49**, 69-80.

Batchelor, F.R., Doyle, F.P., Nayler, J.H.C., and Rolinson, G.N. (1959). Synthesis of Penicillin : 6-Aminopenicillanic acid in Penicillin Fermentations. *Nature* **183**, 257-258.

Betz, J.L., Smyth, P.R., and Clarke, P.H. (1974). Evolution in Action. *Nature* **247**, 261-264.

Bhattacharya, S., Gupta, V.S., Prabhune, A.A., SivaRaman, H., Debnath, M., and Ranjekar, P.K. (1993). Studies of operational variables in batch mode for genetically engineered *Escherichia coli* cells containing penicillin acylase. *Enzyme Microb. Technol.* **15**, 1070-1073.

Blow, D. *Introduction to rotation and translation functions*. in *Molecular Replacement, Proceedings of the CCP4 Study Weekend*. 1985. Daresbury, U.K.

Bock, A., R, W., Schmid, G., Schumacher, G., Lang, G., and Buckel, P. (1983b). The two subunits of penicillin acylase are processed from a common precursor. *FEMS Microbiol. Lett.* **20**, 141-144.

Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R., and Murzin, A.G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**, 416-419.

Bruggink, A., Roos, E.C., and de Vroom, E. (1998). Penicillin acylase in the industrial production of β-lactam antibiotics. *Organic Process Research & Development* **2**, 128-133.

Brunger, A.T. (1992). Free R-Value - a Novel Statistical Quantity For Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475.

Bryan, P., Pantoliano, M.W., Quill, S.G., Hsiao, H.-Y., and Poulos, T. (1986). Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Biochemistry* **83**, 3743-3745.

Brzozowski, A.M. and Tolley, S.P. (1994). Poly(ethylene) glycol monomethyl ethers - an alternative to poly(ethylene) glycols in protein crystallisation. *Acta Crystallogr. D* **50**(4), 466-468.

Burley, S.K. and Petsko, G.A. (1988). Weakly polar interactions in proteins. *Advances in Protein Chemistry* **39**, 125-189.

Buzy, A., Bracchi, V., Sterjiades, R., Chroboczek, J., Thibault, P., Gagnon, J., Jouve, H.M., and Hudryclergeon, G. (1995). Complete amino-acid sequence of *Proteus mirabilis* PR catalase - Occurrence of a methionine sulfone in the close proximity of the active-site. *Journal of Protein Chemistry* **14**(2), 59-72.

Choi, K.S., Kim, J.A., and Kang, H.S. (1992). Effects of site-directed mutations on processing and activities of penicillin G acylase from Escherichia coli ATCC 11105. *J. Bacteriol.* **174**, 6270-6276.

Claridge, C.A., Gourevitch, A., and Lein, J. (1960). Bacterial Penicillin Amidase. *Nature* **187**, 237-238.

Claridge, C.A., Luttinger, J.R., and Lein, J. (1963). Specificity of Penicillin Amidases. *Pro. Soc. Exp. Biol. Med.* **113**, 1008-1012.

Cole, M. (1964). Properties of the Penicillin Deacylase Enzyme of *Escherichia coli. Nature* **203**, 519.

Cole, M. (1969a). Hydrolysis of Penicillins and Related Compounds by the Cell-Bound Penicillin Acylase of *Escherichia coli. Biochem. J.* **115**, 733-739.

Cole, M. (1969b). Deacylation of Acylamino Compounds Other than Penicillins by the Cell Bound Penicillin Acylase of *Escherichia coli. Biochem. J.* **115**, 741-745.

Collaborative Computational Project, N. (1994). The CCP4 Suite - Programs For Protein Crystallography. *Acta Crystallogr. D* **50**(Pt5), 760-763.

Crowfoot, D., Bunn, C.W., Rogers-Low, B.W., and Turner-Jones, A. (1949). The X-ray crystallographic investigation of the structure of penicillin, in *The Chemistry of Penicillin*, H.T. Clarke, J.R. Johnson, and R. Robinson, Editors. 1949, Princeton University Press: Princeton. 310-367.

Cruickshank, D.W.J. *Protein precision re-examined: Luzzati plots do not estimate final errors*. in *Macromolecular Refinement*. 1996. Chester, U.K.

Daopin, S., Davies, D.R., Schlenegger, M.P., and Grutter, M.G. (1994). Comparison of two crystal structures of TGF-β2: the accurancy of refined protein structures. *Acta Crystallogr. D* **50**, 85-92.

Daumy, G.O., Danley, D., McColl, A., Apostolakos, D., and Vinick, F.J. (1985a). Experimental Evolution of Penicillin G Acylase from *Escherichia coli* and *Proteus rettgeri. J. Bacteriol.* **163**(3), 925-932.

Daumy, G.O., Danley, D., and McColl, A.S. (1985b). Role of protein subunits in *Proteus rettgeri* penicillin G acylase. *J. Bacteriol.* **163**, 1279-1281.

Daumy, G.O., McColl, A.S., and Apostolakos, D. (1982). Repression of penicillin G acylase of *Proteus rettgeri* by tricarboxylic acid cycle intermediates. *J. Bacteriol.* **152**, 104-110.

Dauter, Z. (1997). Data collection strategy, in *Macromolecular Crystallography, Part A*, C.W.C. Jr. and R.M. Sweet, Editors. 1997, Academic Press: San Diego. 326-344.

del Rio, G., Lopez-Munguia, A., and Soberon, X. (1996). An engineered penicillin acylase with altered surface charge is more stable in alkaline pH. *Ann. N. Y. Acad. Sci.* **799**, 61-64.

del Rio, G., Rodríguez, M., Munguía, M., Lopez-Munguia, A., and Soberon, X. (1995). Mutant *Escherichia coli* penicillin acylase with enhanced stability at alkaline pH. *Biotechnol. Bioeng.* **48**, 141-148.

Diederichs, K. and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**(4), 269-275.

Dodson, E., Kleywegt, G.J., and Wilson, K. (1996). Report of a Workshop on the Use of Statistical Validators in Protein X-ray Crystallography. *Acta Crystallogr. D* **52**, 228-234.

Done, S. (1996). Structural studies of Penicillin acylase. Ph.D thesis. Department of Chemistry, University of York, York.

Done, S.H., Brannigan, J.A., Moody, P.C.E., and Hubbard, R.E. (1998). Ligand-induced conformational change in Penicillin acylase. *J. Mol. Biol.* **284**, 463-475.

Duggleby, H.J., Tolley, S.P., Hill, C.P., Dodson, E.J., Dodson, G.G., and Moody, P.C. (1995). Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373**, 264-268.

Echols, H., Lu, C., and Burgers, P.M. (1983). Mutator strains of *Escherichia coli mutD* and *dnaA*, with defective exonucleolytic editing by DNA polymerase III holoenzyme. *Pro. Natl. Acad. Sci. USA* **801**, 2189-2191.

Engh, R.A. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392-400.

Forney, L.J. and Wong, D.C. (1989b). Alteration of the catalytic efficiency of penicillin amidase from *Escherichia coli. Appl. Environ. Microbiol.* **55**, 2556-2560.

Forney, L.J., Wong, D.C., and Ferber, D.M. (1989a). Selection of amidases with novel substrate specificities from penicillin amidase of *Escherichia coli. Appl. Environ. Microbiol.* **55**, 2550-2555.

Fowler, R.G., Degnen, G.E., and Cox, E.C. (1974). Mutational specificity of a conditional *Escherichia coli* mutator, *mutD5. Mol. Gen. Genet.* **133**, 179-191.

Furter, R. (1998). Expansion of the genetic code: Site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli. Protein Sci* **7**, 419-426.

Gamblin, S.J. and Rodgers, D.W. *Some practical details of data collection at 100K*. in *Data Collection and Processing, Proceedings of the CCP4 Study Weekend*. 1993. Daresbury, U.K.

Garman, E.F. and Schneider, T.R. (1997). Macromolecular cryocrystallography. *J. Appl. Crystallogr.* **30**, 211-237.

Gregoriou, M., Noble, M.E.M., Watson, K.A., Garman, E.F., Krulle, T.M., DelaFuente, C., Fleet, G.W.J., Oikonomakos, N.G., and Johnson, L.N. (1998). The structure of a glycogen phosphorylase glucopyranose spirohydantoin complex at 1.8 angstrom resolution and 100 K: The role of the water structure and its contribution to binding. *Protein Sci* **7**(4), 915-927.

Hall, A. and Knowles, J.R. (1976). Directed selective pressure on a β-lactamase to analyse molecular changes involved in development of enzyme function. *Nature* **264**, 803-804.

Hamilton-Miller, J.M.T. (1966). Penicillin acylase. *Bacteriol. Rev.* **30**(4), 761-771.

Henderson, R. (1970). Structure of Crystalline α-Chymotrypsin IV. The Structure of Indoleacryloyl-α-Chymotrypsin and its Relevance to the Hydrolytic Mechanism of the Enzyme. *J. Mol. Biol.* **54**, 341-354.

Hendrickson, W.A. and Konnert, J.H. (1980). Stereochemically restrained crystallographic least-squares refinement of macromolecule structures, in *Biomolecular Structure, Function, Conformation and Evolution*, R. Srinivasan, Editor. 1980, Pergamon: Oxford. 43-57.

Hewitt, L., Kasche, V., Lummer, K., Rieks, A., and Wilson, K.S. (in press). Crystallisation of a precursor penicillin acylase from *Escherichia coli. Acta Crystallogr. D* .

Huang, H.T., Seto, T.A., and Shull, G.M. (1963). Distribution and Substrate Specificity of Benzylpenicillin Acylase. *Appl. Microbiol.* **11**, 1-6.

Isupov, M.N., Obmolova, G., Butterworth, S., Badet-Denisot, M.A., Badet, B., Polikarpov, I., Littlechild, J.A., and Teplyakov, A. (1996). Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 Å crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure* **4**, 801-810.

Janssen, D.B., Floris, R., and Alkema, W. (1998). Kinetic and mutagenesis studies of *E. coli. BIOCATALYSIS-98: Fundamentals & Applications (Abstracts)* , 10.

Kachurin, A.M., Golubev, A.M., Geisow, M.M., Veselinka, O.S., Isaevaivanova, L.S., and Neustroev, K.N. (1995). Role of methionine in the active-site of alpha-galactosidase from *Trichoderma reesei. Biochem. J.* **308**(Pt3), 955-964.

Kaufman Katz, A., Glusker, J.P., Beebe, S.A., and Bock, C.W. (1996). Calcium Ion Coordination: A Comparison with that of Beryllium, Magnesium, and Zinc. *J. Am. Chem. Soc.* **118**, 5752-5763.

Kaufmann, W. and Bauer, K. (1964). Variety of Substrates for Bacterial Benzyl Penicillin-splitting Enzyme. *Nature* **203**, 520.

Keilmann, C., Wanner, G., and Bock, A. (1993). Molecular basis of the exclusive low-temperature synthesis of an enzyme in E. coli: Penicillin acylase. *Biol Chem Hoppe Seyler* **374**, 983-992.

Kleina, L.G., Masson, J.M., Normanly, J., Abelson, J., and Miller, J.H. (1990b). Construction of *Escherichia coli* Amber Suppressor tRna Genes. II. Synthesis of Additional tRNA Genes and Improvement of Suppressor Efficiency. *J. Mol. Biol.* **213**(4), 705-717.

Kleina, L.G. and Miller, J.H. (1990a). Genetic Studies of the *lac* Repressor. XIII. Extensive Amino Acid Replacements Generated By the Use of Natural and Synthetic Nonsense Suppressors. *J. Mol. Biol.* **212**(2), 295-318.

Kleywegt, G.J. and Jones, T.A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178-185.

Knowles, J.R. (1987). Tinkering with enzymes: what are we learning? *Science* **236**, 1252-1258.

Konecny, J., Schneider, A., and Sieber, M. (1983). Kinetics and mechanism of acyl transfer by penicillin acylases. *Biotechnol. Bioeng.* **25**, 451-467.

Kopps, P., GeneEditor *in vitro* Site-Directed Mutagenesis System, in *Promega Notes Magazine,*1998. No 65: 30-32.

Kutzbach, C. and Rauenbusch, E. (1974). Preparation and general properties of crystalline penicillin acylase from *Escherichia coli* ATCC 11 105. *Hoppe Seylers Z. Physiol. Chem.* **355**, 45-53.

Lamzin, V.S. and Wilson, K.S. (1993). Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129-147.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.

Lee, Y.S. and Park, S.S. (1998). Two-step autocatalytic processing of the glutaryl 7-aminocephalosporanic acid acylase from *Pseudomonas* sp. strain GK16. *J. Bacteriol.* **180**(17), 4576-4582.

Lindsay, C.D. and Pain, R.H. (1991). Refolding and assembly of penicillin acylase, an enzyme composed of two polypeptide chains that result from proteolytic activation. *Biochemistry* **30**, 9034-9040.

Liu, D.R., Magliery, T.J., Pastrnak, M., and Schultz, P.G. (1997). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo. Pro. Natl. Acad. Sci. USA* **94**, 10092-10097.

Ljubijankic, G., Konstantinovic, M., and Glisin, V. (1992). The Primary Structure of *Providencia rettgeri* Penicillin G Amidase Gene and its Relationship to other Gram Negative Amidases. *DNSequence-I. DNA Sequencing and Mapping* **3**, 195-200.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533-539.

Luzzati, P.V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* **5**, 802-810.

Mahajan, P.B. and Borkar, P.S. (1983). Chemical modifications of Escherichia coli penicillin acylase I, tryptophan involvement at the catalytic site. *Hindustan Antibiot Bull* **25**, 6-10.

Margolin, A.L., Švedas, V.K., and Berezin, I.V. (1980). Substrate specificity of penicillin amidase from *E. coli. Biochimica et Biophysica Acta* **616**, 283-289.

Markiewicz, P., Kleina, L.G., Cruz, C., Ehret, S., and Miller, J.H. (1994). Genetic Studies of the Lac Repressor. XIV. Analysis of 4000 Altered *Escherichia coli lac* Repressors Reveals Essential and Non-essential Residues, as well as "Spacers" which do not Require a Specific Sequence. *J. Mol. Biol.* **240**(5), 421-433.

Márquez, G., Buesa, J.M., Garcia, J.L., and Barbero, J.L. (1988). Conformational Stability of the Penicillin G Acylase from *Kluyvera citropila. Appl. Microbiol. Biotechnol.* **28**, 144-147.

Martín, J., Mancheno, J.M., and Arche, R. (1993). Inactivation of Penicillin Acylase from *Kluyvera Citrophila* by N-ethoxycarbonyl 2-ethoxy-1,2-dihydroquinoline a Case of Time Dependent Non-Covalent Enzyme Inhibition. *Biochem. J.* **291**, 907-914.

Mattos, C. and Ringe, D. (1996). Locating and characterising binding sites on proteins. *Nat. Biotechnol.* **14**, 595-599.

McDonald, I.K. and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.

McPhalen, C.A., Strynadka, N.C.J., and James, M.N.G. (1991). Calcium Binding Sites in Proteins: A Structural Perspective, in *Metalloproteins: Structural Aspects*, C.B. Anfinsen, J.T. Edsall, F.M. Richards, and D.S. Eisenberg, Editors. 1991, Academic Press: San Diego. 77-144.

McVey, C.E., Tolley, S.P., Done, S.H., Brannigan, J.A., and Wilson, K.S. (1997). A new crystal form of Penicillin acylase. *Acta Crystallogr.* **D53**, 777-779.

Merino, E., Balbas, P., Recillas, F., Becerril, B., Valle, F., and Bolivar, F. (1992). Carbon regulation and the role in nature of the *Escherichia coli* penicillin acylase (pac) gene. *Mol. Microbiol.* **6**, 2175-2182.

Miller, J.H., Kleina, L.G., Masson, J.M., Normanly, J., and Abelson, J. (1989). Protein Engineering With Synthetic *Escherichia coli* Amber Suppressor Genes. *Genome* **31**(2), 905-908.

Morillas, M. (1996). Catalytic pathway and conformational stability of penicillin G acylase. Ph.D thesis. Department of Biochemistry, University of Newcastle upon Tyne, Newcastle.

Morillas, M., McVey, C.E., Brannigan, J.A., and Virden, R. (in preparation). Change of substrate specificity on substitution of phenylalanine B:71 in penicillin G acylase from *Escherichia coli ATCC 11105.* .

Mortlock, R.P. (1982). Metabolic acquisitions through laboratory selection. *Annu. Rev. Microbiol.* **36**, 259-284.

Murao, S. (1955). Penicillin amidase. III. Mechanism of penicillin amidase on sodium penicillin G. *J. Agric. Chem. Soc. Jap.* **29**, 404-407.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**(Pt3), 240-255.

Navaza, J. and Saludjian, P. (1997). AMoRe: An automated molecular replacement program package, in *Macromolecular Crystallography (Part A)*, C.W.C. Jr. and R.M. Sweet, Editors. 1997, Academic Press: San Diego. 581-594.

Nayler, J.H.C. (1991a). Early discoveries in the penicillin series. *TIBS* **16**, 195-234.

Nayler, J.H.C. (1991b). Semi-synthetic approaches to novel penicillins. *TIBS* **16**, 234-237.

Newton, G.G.F. and Abraham, E.P. (1954). *Biochem. J.* **58**, 103-111.

Niersbach, H., Kuhne, A., Tischer, W., Weber, M., Wedekind, F., and Plapp, R. (1995a). Improvement of the catalytic properties of penicillin G acylase from *Escherichia coli* ATCC 11105 by selection of a new substrate specificity. *Appl Microbiol Biotechnol* **43**, 679-684.

Niersbach, H., Tischer, W., Weber, M., Wedekind, F., and Plapp, R. (1995b). Isolation and Mapping of a Mutant Penicillin G Acylase with Altered Substrate Specificity from *Escherichia coli. Biotechnology Lett.* **17**(1), 19-24.

Nishiya, Y. and Imanaka, T. (1994). Alteration of substrate specificity and optimum pH of sarcosine oxidase by random and site-directed mutagenesis. *Appl. Environ. Microbiol.* **60**(11), 4213-4215.

Normanly, J., Kleina, L.G., Masson, J.M., Abelson, J., and Miller, J.H. (1990). Construction of *Escherichia coli* Amber Suppressor tRNA Genes. III. Determination of tRNA Specificity. *J. Mol. Biol.* **213**(4), 719-726.

Normanly, J., Masson, J.M., Kleina, L.G., Abelson, J., and Miller, J.H. (1986b). Construction of two *Escherichia coli* amber suppressor gene: tRNA CUA-Phe and tRNA CUA-Cys. *Pro. Natl. Acad. Sci. USA* **83**(17), 6548-6552.

Normanly, J., Ogden, R.C., Horvath, S.J., and Abelson, J. (1986a). Changing the identity of a transfer RNA. *Nature* **321**(6067), 213-219.

Oh, S.J., Kim, Y.C., Park, Y.W., Min, S.Y., Kim, I.S., and Kang, H.S. (1987). Complete nucleotide sequence of the penicillin G acylase gene and the flanking regions, and its expression in Escherichia coli. *Gene* **56**, 87-97.

Ohashi, H., Katsuta, Y., Hashizume, T., Abe, S.N., Kajiura, H., Hattori, H., Kamei, T., and Yano, M. (1988). Molecular cloning of the penicillin G acylase gene from *Arthrobacter viscosus. Appl. Environ. Microbiol.* **54**, 2603-2607.

Oinonen, C., Tikkanen, R., Rouvinen, J., and Peltonen, L. (1995). Three-dimensional structure of human lysosomal aspartylglucosaminidase. *Nat. Struct. Biol.* **2**(12), 1102-1108.

Oliver, G., Valle, F., Rosetti, F., Gomez-Pedrozo, M., Santamaria, P., Gosset, G., and Bolivar, F. (1985). A common precursor for the two subunits of the penicillin acylase from *Escherichia coli* ATCC11105. *Gene* **40**, 9-14.

O'Reilly, M., Watson, K.A., Schinzel, R., Palm, D., and Johnson, L.N. (1997). Oligosaccharide substrate binding in *Escherichia coli* maltodextrin phosphorylase. *Nat. Struct. Biol.* **4**(5), 405-412.

Otwinowski, Z. and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscilllation Mode, in *Macromolecular Crystallography (Part A)*, C.W.C. Jr. and R.M. Sweet, Editors. 1997, Academic Press: San Diego. 307-326.

Paetzel, M. and Dalbey, R.E. (1997). Catalytic hydroxyl/amine dyads within serine proteases. *TIBS* **22**, 28-31.

Pauling, L. (1946). Molecular Architecture and Biological Reactions. *Chemical and Engineering News* **24**(10), 1375-1377.

Plaskie, A., Roets, E., and Vanderhaeghe, H. (1978). Substrate specificity of penicillin acylase of *E. coli*. *J. Antibiot.* **31**, 783-788.

Prabhune, A.A. and Sivaraman, H. (1990). Evidence for involvement of arginyl residue at the catalytic site of penicillin acylase from *Escherichia coli*. *Biochemical and Biophysical Research Communications* **173**(1), 317-322.

Prieto, I., Martin, J., Arche, R., Fernandez, P., Perez-Aranda, A., and Barbero, J.L. (1990). Penicillin acylase mutants with altered site-directed activity from *Kluyvera citrophila*. *Appl. Microbiol. Biotechnol.* **33**, 553-559.

Prieto, M.A., Diaz, E., and Garcia, J.L. (1996). Molecular Characterisation of the 4-Hydroxyphenylacetate Catabolic Pathway of *Escherichia coli* W: Engineering a Mobile Aromatic Degradative Cluster. *J. Bacteriol.* **178**(1), 111-120.

Rennell, D., Bouvier, S.E., Hardy, L.W., and Poteete, A.R. (1991). Systematic Mutation of Bateriophage T4 Lysozyme. *J. Mol. Biol.* **222**, 67-87.

Roa, A., Castillon, M.P., Goble, M.L., Virden, R., and Garcia, J.L. (1995). New insights on the specificity of penicillin acylase. *Biochem. Biophys. Res. Commun.* **206**, 629-636.

Roa, A. and Garcia, J.L. (1997). Identification of the pac promoter from *Kluyvera citrophila*. *FEMS Microbiol. Lett.* **151**, 9-16.

Roa, A., Garcia, J.L., Salto, F., and Cortes, E. (1994). Changing the substrate specificity of penicillin G acylase from *Kluyvera citrophila* through selective pressure. *Biochem. J.* **303**(Pt 3), 869-875.

Rodgers, D.W. (1994). Cryocrystallography. *Structure* **2**, 1135-1140.

Rolinson, G.N., Batchelor, F.R., Butterworth, D., Cameron-wood, J., Cole, M., Eustace, G.C., Hart, V., Richards, M., and Chain, E.B. (1960). Formation of 6-Aminopenicillanic Acid from Penicillin by Enzymatic Hydrolysis. *Nature* **187**, 236-237.

Sakaguchi, K. and Murao, S. (1950). A Preliminary report on a new enzyme, penicillin amidase. *J. Agric. Chem. Soc. Jap.* **23**, 411-414.

Schechter, I. and Berger, A. (1967). On the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.

Schimmel, P. and Söll, D. (1997). When protein engineering confronts the tRNA world. *Pro. Natl. Acad. Sci. USA* **94**, 10007-10009.

Schumacher, G., Sizmann, D., Haug, H., Buckel, P., and Bock, A. (1986). Penicillin acylase from *E. coli*: unique gene-protein relation. *Nucleic Acids Res.* **14**, 5713-5727.

Sheldrick, G.M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr.* **A46**, 467-473.

Shewale, J.G. and Sivaraman, H. (1989). Penicillin acylase: Enzyme Production and its Application in the Manufacture of 6-APA. *Process Biochemistry* **24**, 146-154.

Sizmann, D., Keilmann, C., and Bock, A. (1990). Primary structure requirements for the maturation in vivo of penicillin acylase from *Escherichia coli* ATCC 11105. *Eur. J. Biochem.* **192**, 143-151.

Slade, A., Horrocks, A.J., Lindsay, C.D., Dunbar, B., and Virden, R. (1991). Site-directed chemical conversion of serine to cysteine in penicillin acylase from Escherichia coli ATCC 11105. Effect on conformation and catalytic activity. *Eur. J. Biochem.* **197**, 75-80.

Smith, J.L., Zaluzec, E.J., Wery, J.-P., Niu, L., Switzer, R.L., Zalkin, H., and Satow, Y. (1994). Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science* **264**, 1427-1433.

Stura, E.A. and Wilson, I.A. (1990). Analytical and Production Seeding Techniques. *Methods: A Companion to Methods in Enzymology* **1**(1), 38-49.

Stura, E.A. and Wilson, I.A. (1991). Applications of the streak seeding technique in protein crystallisation. *Journal of Crystal Growth* **110**, 270-282.

Suckow, J., Markiewicz, P., Kleina, L.G., Miller, J., Kisterswoike, B., and Mullerhill, B. (1996). Genetic Studies of the Lac Repressor. XV. 4000 Single Amino acid Substitutions and Analysis of the Resulting Phenotypes on the Basis of the Protein Structure. *J. Mol. Biol.* **261**(4), 509-523.

Suresh, C.G., Pundle, A.V., Rao, K.N., SivaRaman, H., Brannigan, J.A., McVey, C.E., Verma, C.S., Dauter, Z., Dodson, E.J., and Dodson, G.G. (in press). Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat. Struct. Biol.* .

Švedas, V.K., Savchenko, M.V., Beltser, A.I., and Guranda, D.F. (1996). Enantioselective penicillin acylase-catalyzed reactions. Factors governing substrate and stereospecificity of the enzyme. *Ann. N. Y. Acad. Sci.* **799**, 659-669.

Szentirmai, A. (1965). Properties of Penicillin Acylase Isolated from *Escherichia coli*. *Acta Microbiologica Academiae Scientiarum Hungaricae* **12**, 395-405.

Teng, T.-Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallogr.* **23**, 387-391.

Valle, F., Balbas, P., Merino, E., and Bolivar, F. (1991). The role of penicillin amidases in nature and in industry. *TIBS* **16**, 36-40.

Valle, F., Gosset, G., Tenorio, B., Oliver, G., and Bolivar, F. (1986). Characterization of the regulatory region of the *Escherichia coli* penicillin acylase structural gene. *Gene* **50**, 119-122.

Vandamme, E.J. and Voets, J.P. (1974). Microbial Penicillin Acylases. *Adv. Appl. Microbiol.* **17**, 311-369.

Verhaert, R.M., Riemens, A.M., van der Laan, J.M., van Duin, J., and Quax, W.J. (1997). Molecular cloning and analysis of the gene encoding the thermostable penicillin G acylase from *Alcaligenes faecalis. Appl. Environ. Microbiol.* **63**, 3412-3418.

Waldmann, H. (1988). The Phenylacetyl (PhAc) group as Enzymatically Removable Protecting Function for Peptides and Carbohydrates. Selective Deprotections with Penicillin Acylase. *Leibigs Ann. Chem.* **12**, 1175-1180.

Waldmann, H., Heuser, A., Braun, P., Schulz, M., and Kunz, H. (1992). New Methods for the Selective Fuctionalisation of Carbohydrate Derivatives, in *Microbial Reagents in Organic Synthesis*, S. Servi, Editor. 1992, Kluwer. 113-122.

Waldmann, H. and Sebastian, D. (1994). Enzymatic Protecting Group Techniques. *Chemical Reviews* **94**(4), 911-937.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

Watenpaugh, K.D. (1991). Cryocrystallography. *Curr. Opin. Struct. Biol.* **1**, 1012.

Weber, P.C. (1997). Overview of Protein Crystallization Methods, in *Macromolecular Crystallography (Part A)*, C.W.C. Jr. and R.M. Sweet, Editors. 1997, Academic Press: San Diego. 13-22.

Williams, J.A. and Zuzel, T.J. (1985). Penicillin G Acylase (E.C.3.4.1.11) Substrate Modification by *in vitro* Mutagenesis. *J. Cellular Biochem.* **sup. 9B**, 99.

Wilson, C. and Agard, D.A. (1991). Engineering substrate specificity. *Curr. Opin. Struct. Biol.* **1**, 617-623.

Zographos, S.E., Oikonomakos, N.G., Tsitsanou, K.E., Leonidas, D.D., Chrysina, E.D., Skamnaki, V.T., Bischoff, H., Goldmann, S., Watson, K.A., and Johnson, L.N. (1997). The structure of glycogen phosphorylase b with an alkyl- dihydropyridine-dicarboxylic acid compound, a novel and potent inhibitor. *Structure* **5**(11), 1413-1425.

Zoller, M.J. (1991). New molecular biology methods for protein engineering. *Curr. Opin. Struct. Biol.* **1**, 605-610.