The determination of the crystal structure of the metallo $\beta$-lactamase from *Stenotrophomonas maltophilia* and mechanistic studies.

Jakir Hussain Ullah.

A thesis submitted in partial fulfillment of the requirements of the University College of London for the Degree of Doctor of Philosophy.


Division of Protein structure.
National Institute for Medical Research.
Mill Hill, London.
This work is dedicated to my family.
Abstract.

The discovery of penicillin and its subsequent development for clinical use against bacterial infections was initially hailed as a magic bullet. However, it was soon apparent that this was not the case because of the appearance of penicillin resistant bacteria. β-lactam antibiotics, such as penicillin, function by inhibiting cell wall synthesis in growing bacterial cells. The major mechanism of resistance in bacteria to the β-lactams is mediated by β-lactamase production. These enzymes bind β-lactams with high affinity and hydrolyse the lactam ring, thus causing inactivation.

Metallo β-lactamases are a potential problem to the efficiency of current β-lactam antibiotics. This is because they have a broader spectrum of activity than the more common serine β-lactamases and there are, as yet, no clinically useful inhibitors of the metallo β-lactamases. Currently, metallo β-lactamases have been found in only a few species that are pathogenic. However, recent reports of the plasmid mediated transfer of metallo β-lactamase genes between bacteria suggest that these genes may spread to pathogens that are more serious and thereby, severely compromise current antibiotic therapy.

The work presented here describes the crystal structure of the metallo β-lactamase L1 from *Stenotrophomonas maltophilia*, which was determined to 1.7Å resolution. The structure was solved using multiwavelength anomalous dispersion from data collected at the absorption edge of the intrinsic zinco's and substituted seleno-methionine residues. Phasing was carried out by treating the data at different wavelengths as a special case of multiple isomorphous replacement with anomalous scattering (MIRas).

The crystal structure showed that L1 was a tetramer, which was also confirmed by sedimentation equilibrium studies. The overall fold is composed of two roughly equivalent domains consisting of a β-sheet surrounded by α helices. The active site is formed in a grove above the interface of the core β sheets and binds two zinc ions. Molecular modelling of substrates suggests that binding maybe facilitated by both the
active site zins. We propose that the substrate is bound to the enzymes via the zinc ions; hydrophobic residues near the active site then contribute to the binding of the various side chains of the β-lactams. The reaction is likely to proceed by nucleophilic attack by a zinc bound hydroxide ion on the β-lactam carbonyl to form the transition state. This complex is subsequently degraded by the donation of a proton, probably from a well positioned water molecule, to the β-lactam nitrogen.
Acknowledgements

I would like to acknowledge my supervisor, Dr Steven Gamblin and Dr James Spencer for their invaluable contributions and advice throughout this work. I would also like to acknowledge Dr Ian Taylor for his help in regards to the ultracentrifugation and kinetic studies that were carried out. I also would like to acknowledge the contribution of Dr C S Verma who carried out the molecular modelling studies.

I thank Dr Timothy Walsh for providing the recombinant plasmid, β-lactam substrates and for allowing me to inherit this project. I would also like to thank Dr Christopher Schofield for providing the tri-fluoromethyl ketone inhibitors used in this work. I also thank Dr Steven Smerdon, Dr Phillip Walker and Dr Emanuel Skordalakes for their assistance during synchrotron trips and Dr John Corrie for helpful discussions regarding the inhibition studies. I also thank the NIMR photographic department for their help in the preparation of the figures in this report.

I thank my family and all my friends, especially B T Hussain, for all the support and encouragement over the years. I also thank the MRC who provided the funding and the facilities for this work.
Contents.

Abstract 3
Acknowledgements 5
Contents 6
List of figures 10
List of tables 11
Abbreviations 12

Chapter 1  Introduction to β-lactamases.

1.1 The discovery and development of β-lactam antibiotics 15
1.2 The mechanism of β-lactam toxicity 17
1.3 Bacterial resistance to β-lactams: The β-lactamases 20
1.4 The serine β-lactamases 22
1.5 Metallo β-lactamases 23
1.6 Catalytic mechanism of enzymes using the OH nucleophile 29
  1.6.1 Active site serine enzymes 29
  1.6.2 Active site aspartic enzymes 33
  1.6.3 Metallo enzymes 33
1.7 Metallo β-lactamase: binding and catalysis 36
1.8 Inhibition of the metallo β-lactamase 39
1.9 The metallo β-lactamase L1 from S. maltophilia 41
1.10 Project aims 43

Chapter 2  Introduction to structure determination of biological macromolecules by X-ray crystallography.

2.1 Introduction 45
2.2 Crystallisation 46
Chapter 3  Methods and Materials.

3.1 Production and preliminary characterisation of native L1 60
3.2 Production of seleno-methionine incorporated L1 61
3.3 Purification of native and Se-met incorporated L1 63
3.4 Assessment of purification strategy and quality 64
3.5 Determination of the relative molecular mass of L1 65
3.6 Kinetic studies 67
3.7 Crystallisation of native and Se-met L1 68
3.8 Preparation of crystals for exposure to X-rays 69
3.9 Data collection of native L1 69
3.10 Multiwavelength Anomalous dispersion data collection 70
3.11 Preparation of enzyme-inhibitor complex and data collection 71
3.12 Data analysis procedures 71
3.13 MIR using CCP4 software 74
3.14 SOLVE 75
3.15 Density Modification and Phase extension 76
3.16 Refinement 77
3.17 Computing and additional software 78

Chapter 4  Results. Protein purification, characterisation and crystallisation.

4.1 Purification 81
4.2 Dynamic light scattering 83
4.3 Determination of molecular mass 84
4.4 Crystals of L1 87
Chapter 5  Results. Crystallographic analysis.

5.1 Native data  
5.2 MAD data  
5.3 Phasing  
5.4 Comparison of HASSP and SOLVE in phasing

Chapter 6  Results. The structure of *S. maltophilia* L1 and mechanistic studies.

6.1 The structure of L1  
6.2 The structure of the active site  
6.3 The L1 tetramer  
6.4 Inhibition of L1

Chapter 7  Discussion.

7.1 Discussion of phasing procedures  
7.2 CCP4 and SOLVE approaches to phase determination  
7.3 Structural features of L1  
7.4 Inhibition studies  
7.5 Substrate binding  
7.6 Mechanism of catalysis

References.
List of figures.

1 The structure of penicillin G 15
2 The structure of a clinically used carbapenem 16
3 Bacterial cell walls 18
4 The peptidoglycan layer (from *S. aureus*) 19
5 Sequence alignment of known metallo β-lactamase sequences 26
6 The structures of ccrA and BCII 28
7 A schematic of the general mechanism proposed for serine β-lactamases 32
8 Mechanism proposed by Concha *et al.* 37
9 Mechanism proposed by Bounaka *et al.* 38
10 A BPT molecule bound at the active site of ccrA 40
11 Crystallisation diagram 47
12 Hanging drop method 49
13 MAD phasing 55
14 Structure determination 73
15 SDS PAGE of the peak fractions from gel filtration 82
16 Subunit mass of L1 from ESMS 85
17 *M* of native L1 by sedimentation equilibrium 86
18 L1 crystals 87
19 Hydrolysis of nitrocefin 88
20 Fluorescence scan of Zn close to the absorption edge 93
21 Fluorescence scan of Se close to the absorption edge 93
22 Harker section z=1/3 95
23 Ramachandran plot 98
24 Electron density map at 1.7 Å 99
25 Subunit structure of L1 104
26 Topology of L1 105
### List of Tables.

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conserved sequence elements in the serine ( \beta )-lactamase family and PBPs</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Sequence similarity amongst the known metallo ( \beta )-lactamases</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Examples of atoms which could be used for MAD phasing</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>Purification of native L1</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>Mean values from dynamic light scattering analysis</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>Crystal size, lattice and unit cell dimensions</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>Native dataset to 2.4 ( \AA )</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>Native dataset to 1.7 ( \AA )</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>Data processing statistics for the MAD datasets</td>
<td>92</td>
</tr>
<tr>
<td>10</td>
<td>MLPHARE refined anomalous scattering sites</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>Phasing statistics</td>
<td>97</td>
</tr>
<tr>
<td>12</td>
<td>Refinement</td>
<td>97</td>
</tr>
<tr>
<td>13</td>
<td>Comparison of HASSP and SOLVE in the determination of anomalous</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Scattering sites</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Phase errors</td>
<td>101</td>
</tr>
<tr>
<td>15</td>
<td>Zinc co-ordination</td>
<td>108</td>
</tr>
<tr>
<td>16</td>
<td>Hydrogen bonds involved in the formation of the tetramer</td>
<td>111</td>
</tr>
<tr>
<td>17</td>
<td>The affect of the TFK inhibitor on ( K_m ) and ( V_{max} )</td>
<td>116</td>
</tr>
<tr>
<td>18</td>
<td>Kinetics of meropenem hydrolysis</td>
<td>117</td>
</tr>
</tbody>
</table>
Abbreviations

ml Millilitre \((10^{-3} \text{ litres})\)

µl Microlitre \((10^{-6} \text{ litres})\)

mg Milligrams \((10^{-3} \text{ grams})\)

µg Micrograms \((10^{-6} \text{ grams})\)

mM Millimolar \((10^{-3} \text{ molar})\)

µM Micromolar \((10^{-6} \text{ molar})\)

nM Nanomolar \((10^{-9} \text{ molar})\)

Mm Millimetre \((10^{-3} \text{ metres})\)

nm Nanometres \((10^{-9} \text{ metres})\)

Å Angstroms \((0.1 \text{nm})\)

Da Daltons

KDa Kilo Daltons

s seconds

KJ Kilo Joules

C Centigrade

K Kelvin

rpm revolutions per minute

T temperature

pI Isoelectric point

K_m Michaelis constant

V_{max} Maximal velocity

K_{cat} Turnover number

K_i Inhibition constant

K_d Dissociation constant

EU enzyme units

UV Ultra-violet

Zn zinc
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se</td>
<td>Selenium</td>
</tr>
<tr>
<td>Se_met</td>
<td>Seleno-methionine</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td>SIR</td>
<td>Single isomorphous replacement</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MAD</td>
<td>Multi-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>(lambda) wavelength</td>
</tr>
<tr>
<td>F</td>
<td>Structure factor</td>
</tr>
<tr>
<td>(</td>
<td>F</td>
</tr>
<tr>
<td>( f )</td>
<td>Atomic scattering factor</td>
</tr>
<tr>
<td>B</td>
<td>Temperature factor</td>
</tr>
<tr>
<td>BTP</td>
<td>Biphenyl tetrazole</td>
</tr>
<tr>
<td>MTE</td>
<td>Mercaptoacetic acid thiol esters</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropylthiogalactoside</td>
</tr>
<tr>
<td>TFK</td>
<td>Trifluoromethyl ketone</td>
</tr>
<tr>
<td>BCII</td>
<td>metallo ( \beta )-lactamase from <em>Bacillus cereus</em></td>
</tr>
<tr>
<td>CcrA</td>
<td>metallo ( \beta )-lactamase from <em>Bacteroides fragilis</em></td>
</tr>
<tr>
<td>L1</td>
<td>metallo ( \beta )-lactamase from <em>Stenotrophomonas maltophilia</em></td>
</tr>
</tbody>
</table>
Chapter 1.

Introduction to $\beta$-lactamases.
1.1 The discovery and development of $\beta$-lactam antibiotics

Penicillin was discovered by Alexander Fleming, largely by chance, in 1928. Fleming noticed that a bacterial culture plate which had become contaminated with the mould *Penicillium notatum*, displayed regions of cell lysis in the vicinity of the contaminating mould. Fleming was able to show that a substance secreted by the mould had caused the observed bacterial cell lysis. He named the substance penicillin (Fleming, 1929). Penicillin was the first antibiotic to be discovered; its structure consisted of a thioloazidine ring fused to a $\beta$-lactam ring with a variable side chain attached to the $\beta$-lactam by a peptide bond (figure 1).

**Figure 1** The structure of penicillin G

It was not until the late 1930’s that penicillin was isolated and exploited as the basis of first generation $\beta$-lactam antibiotics for medicinal use (Nathwani and Wood, 1993 and references therein). Since then other $\beta$-lactam antibiotics have been developed and have made a huge impact on medical practice. They are mass-produced by the pharmaceutical industry with a global turnover of several billion pounds. $\beta$-lactam antibiotics currently administered, in cases of simple or complex bacterial infection
(i.e., infection due to several bacterial strains), are classified according to their structure. Some of these are listed below (alphabetical):

- Carbapenems
- Cephalosporins
- Cephamycins
- Monobactams.
- Penicillins.

The carbapenems are a relatively new member of this family and show the widest spectrum of activity amongst the β-lactam antibiotics and are also stable to most β-lactamases (Wise, 1986; Livermore D, 1993; Rasmussen and Bush, 1997). The carbapenems differ in their chemical structure to penicillins in having an unsaturated bond between position 2 and position 3 and also a carbon replaces the sulphur at position 1 (Moellering et al., 1989). Other non β-lactam antibiotics are available such as the aminoglycosides, tetracyclins, sulphonimides, and the quinolones but as these are not targets for the β-lactamases they will not be discussed here. The interested reader is referred to relevant reviews (Plotkin, 1968; Plotkin, 1975; Neu, 1987; Neu, 1992).

**Figure 2** The structure of a clinically used carbapenem (meropenem)
1.2 The mechanism of β-lactam toxicity.

Naturally occurring β-lactams have been found in several types of microorganisms, including eukaryotic fungi and bacteria, as part of the host defence mechanism (Abraham E P, 1977; Schonberg-Albers et al., 1978). A large number of analogues and derivatives of the β-lactams have been developed either from a complete synthesis approach, or by partial synthesis from naturally occurring precursor β-lactams (Abraham E P, 1977; Leanza et al., 1979; Kahan et al., 1983; Sunagawa et al., 1987; Nathwani and Wood, 1993). β-lactams only exert their toxic affects on growing bacterial cells and they do this by interfering with the normal process of cell division. β-lactam antibiotics recognise and bind a class of proteins, collectively known as penicillin binding proteins (PBPs) because of their high affinity for penicillins (Spratt, 1975; Ghuysen, 1991; Frere, 1995), that are important in cell wall synthesis. The sequestration of PBPs leads to cell death.

Bacterial cells are coated with a rigid cell wall that provides mechanical support. This coat allows them to survive in hypotonic environments that would otherwise cause them to burst due to their high internal osmotic pressure. Bacteria are classified according to whether they can take up Gram stain (Gram positive) or cannot (Gram negative). Gram positive bacteria have a thick cell wall (≈250Å) surrounding their plasma membrane, whereas Gram negative bacteria have a thinner cell wall (≈30Å) covered by a complex outer membrane (see figure 3). Both types of cell wall contain a peptidoglycan layer. The peptidoglycan layer is composed of linear strands of glycan, consisting of the alternating sugars N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), crosslinked by short peptides (figure 4). The final cross-linking step is catalysed by a DD-transpeptidase.
Figure 3  Bacterial cell walls

**Gram positive cell wall**

- Peptidoglycan layer (~250 Å)
- Inner membrane

**Gram negative cell wall**

- Outer membrane
- Peptidoglycan layer (~30 Å)
- Periplasmic space
- Inner membrane
The basic unit of peptidoglycan is the N-acetylglucosamine (NAG)-N-acetylmuramic acid (NAM)-peptide structure (a). In *S. aureus*, a pentapeptide (L-Ala, D-Gln, L-Lys, D-Ala, D-Ala) is attached to the NAM sugar as shown below. A second peptide (consisting of 5 glycines) is then attached via the L-Lys of the first peptide and is referred to as the glycine bridge.

![Diagram of peptidoglycan structure](image)

The above units are used to generate polymeric glycan strands by forming alternating β1-4 glycosidic bonds between adjacent NAM-NAG sugars. Parallel glycan strands are then crosslinked via the terminal glycine of the glycine bridge (upper strand) and the penultimate D-Ala (lower strand) as shown in (b). This step is catalysed by a DD-transpeptidase and results in the removal of the terminal D-Ala.

![Diagram of crosslinking](image)

In growing cells, the inhibition of the DD-transpeptidase by the β-lactam antibiotics results in the formation of uncrosslinked glycan strands and hence a defective cell wall. This leads to a series of events that results in cell lysis. This figure was adapted from Stryer 1990.
During cell division and growth, the peptidoglycan layer is systematically broken down and regenerated for remodelling of the cell. Hydrolysis of the peptidoglycan layer is catalysed by DD-carboxypeptidases and its regeneration after elongation or remodelling is catalysed by the DD-transpeptidases. Both of these proteins are penicillin sensitive (Wise et al., 1965).

The mechanism of inhibition of the DD-transpeptidase is fairly well understood. Inhibition was found to result in the accumulation of uncross-linked glycan in *S. aureus* (grown in the presence of sublethal concentrations of penicillin (Wise et al., 1965). It was found that the inactivation of the DD transpeptidase was due to the acylation of a serine at its active site by penicillin. This occurs because penicillin is a structural analogue of the DD-transpeptidase substrate, the D-Ala-D-Ala C-terminus of the NAG-NAM-peptide unit (Tipper et al., 1985). The role of the DD-carboxypeptidases in bacterial cell death is less clear. In some strains it has been found that inhibition of its activity by penicillin is lethal, whereas in other cases these enzymes have been shown to be 'inessential' to bacterial cell growth (Frere, 1995).

1.3 Bacterial resistance to β-lactams: The β-lactamases.

Bacteria have acquired mechanisms to overcome the threats posed to their survival by the β-lactam antibiotics. How do bacteria adapt so quickly? Bacteria are predisposed for rapid adaptability because they are able to accept foreign DNA and have rapid multiplication rates. Resistance to β-lactam antibiotics can arise in several ways, one of which is alteration of the PBP’s so that they no longer recognise or have lower binding affinities for β-lactam antibiotics (Dever et al., 1991). Resistance in gram negative bacteria can also arise from reduced membrane permeability, usually mediated
by porin loss or modification (reviewed in Frere, 1995; and Nord and Hedberg, 1996). However, the major cause of resistance in bacteria is due to β-lactamase production.

β-lactamases are produced by bacteria so that they can withstand and survive in environments that would otherwise be toxic (Abraham et al., 1940; Acar, 1986; Brook, 1987; Frere, 1995). In gram negative bacteria these enzymes are located in the periplasmic space but in gram positive bacteria they are extracellular (Frere, 1995). β-lactamases are either produced constitutively or are induced by the presence of β-lactams (Paton et al., 1994; Frere, 1995; Rossolini et al., 1996; Walsh et al., 1997). Genes coding for β-lactamases can be either chromosomal or plasmid borne (Watanabe et al., 1991; Bandoh et al., 1992; Ito et al., 1995; Minami et al., 1996; Jacoby et al., 1997; O’Hara et al., 1998).

β-lactamases have been classified on both a molecular and functional basis. The molecular approach is based on amino acid sequence similarity (Ambler R P, 1980) and has led to the definition of four molecular classes A-D. Classes A, C, D all utilise a serine as the catalytic residue and thus could be said to belong to the family of serine β-lactamases. β-lactamases belonging to molecular class B require a divalent metal ion, usually zinc, for activity and are thus referred to as the metallo β-lactamases. Functionally, β-lactamases have been classified into four groups 1-4. Groups 1, 2 and 4 consist of the serine β-lactamases and group 3 the metallo β-lactamases (Bush K, 1989(a); 1989(b); 1989(c); Bush et al, 1995). Gram positive bacteria only produce β-lactamases from the molecular classes A and B whereas all four classes are produced in gram negative bacteria (Frere, 1995).
1.4 The serine β-lactamases.

Crystal Structures for the class A and C serine β-lactamases have been determined by x-ray crystallography and show considerable similarities (Herzberg O. 1991; Jelsch et al.; 1992). The overall fold consists of an all α and a αβ domain and the active site is situated in a groove between the two domains. Although there is little overall sequence homology between the different classes of serine lactamases, careful sequence alignments have revealed the existence of three conserved sequence elements surrounding the active site that are located in similar positions within the known structures. These elements contain identical or chemically similar residues and are found in all members of serine β-lactamase family and also the PBP’s (see table 1).

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Element 1</th>
<th>Element 2</th>
<th>Element 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>SXXK</td>
<td>SDN</td>
<td>KTG/KSG/RSG</td>
</tr>
<tr>
<td>Class C</td>
<td>SXXK</td>
<td>YAN</td>
<td>KTG</td>
</tr>
<tr>
<td>Class D</td>
<td>SXXK</td>
<td>YGN</td>
<td>KTG</td>
</tr>
<tr>
<td>PBP’s</td>
<td>SXXK</td>
<td>SXN/SXC/YGN</td>
<td>KTG/KSG</td>
</tr>
</tbody>
</table>

Table 1 Conserved sequence elements in the serine β-lactamase family and PBP’s.

Class A, C and D are serine β-lactamases. This table was adapted from Galleni M et. al.; 1995.

The first element is part of a helix and contains the active site serine and a conserved lysine (see next section). The helical nature of the this element results in the side chain of this lysine pointing into the active site, where it forms a hydrogen bond with the active site serine hydroxyl. Element 2 is a short loop from the all α domain. The side chains of the first and third residues point into the active site and form one wall of the
active site cavity, while the second residue is buried in the core of the protein. The third element is provided by the inner most sheet of the α/β domain and forms the opposite wall of the active site cavity. In addition the Class A enzymes have a fourth conserved element known as the omega loop which contains an invariant glutamate.

Although a structure has not been determined as yet for the class D enzyme, sequence alignments have identified three structural elements corresponding to those observed in Class A and C. The elucidation of the structures of several PBP’s has also revealed the presence of equivalent elements that superimpose very well on those of the class A and C β-lactamases (Kelly et al., 1986; Samraoui et al., 1986; Ghuysen et al., 1991; Jamin et al., 1994; Kelly et al., 1998). These findings support the initial hypothesis that the serine β-lactamases have evolved from the PBPs (Strominger and Tipper, 1965).

The serine β-lactamases are a significant threat to the use of β-lactam antibiotics. However the development of inhibitors and more stable antibiotics has helped to circumvent the problems posed by the serine β-lactamases, (Chen et al., 1992; Livermore, 1994). With a few exceptions, serine β-lactamases are unable to inactivate carbapenems (Rasmussen and Bush, 1997).

### 1.5 Metallo β-lactamases.

Bacteria that possess a class B β-lactamase, with one exception, are characterised by exhibiting resistance to virtually all known β-lactam antibiotics, including the carbapenems. The broad spectrum of resistance may be conferred by a class B enzyme alone (Felici et al., 1993; Rasmussen et al., 1994; Felici et al., 1995) although
the literature suggests that they are produced in conjunction with β-lactamases from the other groups (Rasmussen and Bush, 1997). The exception is the class B enzyme from the *Aeromonas* family, which appears to be a specific carbapenemase (Walsh et al., 1996).

Metallo β-lactamas cannot inactivate monobactams but they are not inhibited by them either (Livermore, 1993; Yang et al., 1996). It was initially thought that moxalactam (a semi synthetic β-lactam related to cephalosporin and cephemycins) and cefoxitin (a cephemycin) were inhibitors of the metallo β-lactamas. However, it was later found that inhibition only occurred at extremely high concentrations that were not clinically feasible (Felici et al., 1995). The metallo β-lactamas are also unaffected by serine β-lactamase inhibitors (Yang et al., 1992; Felici et al., 1993). The requirement of a divalent transition metal ion for activity was confirmed by inactivation of these enzymes by EDTA and EGTA (Davies et al., 1974; Bandoh et al., 1991; Yang et al., 1992; Paton et al., 1994). Genes coding for metallo β-lactamas are usually chromosomally mediated but in some cases can be plasmid borne (see earlier references, page 21). Gene expression can be either constitutive or induced (see earlier references, page 21).

Presently metallo β-lactamas have been found in just 19 species. Of these, amino acid sequence information is available for the following 10 species. Hereafter, the enzymes from these 10 species will be referred to by the names given in italics in parenthesis.

- *Aeromonas hydrophilla* (*cphA*) (Massida et al., 1991)
- *Aeromonas sobria* (*ImiS*) (Walsh et al., 1998)
- *Pseudomonas aeroginosa* (*IMP-1*) (Watanabe et al., 1991)
The metallo β-lactamase found in the two *Aeromonas* species is almost identical (97% sequence homology) and thus a single sequence in sufficient for alignment purposes. Also, the metallo β-lactamases found in *Serratia marcescens*, *Klebsiella pneumonia*, and *Pseudomonas aeruginosa* are identical and therefore a single entry is again sufficient in sequence alignments. Allowing for the above exceptions the degree of sequence homology amongst the metallo β-lactamases is generally rather low as can be seen from table 2. All the sequences but one share a single common element consisting of His-X-His-X-Asp (where X can be any amino acid), which is a zinc binding motif (Vallee et al., 1990). The exception is the *Aeromonas* species where the first His is replaced by an Asn (Massida et al., 1991; Rossilini et al., 1996; Walsh et al., 1998).

**Table 2** Sequence similarity amongst the known metallo β-lactamases.

<table>
<thead>
<tr>
<th>Percent similarity</th>
<th>L1</th>
<th>ccrA</th>
<th>BCII</th>
<th>Imp-1</th>
<th>cphA</th>
<th>BlaB</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>10.8</td>
<td>12.5</td>
<td>12.2</td>
<td>12.2</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td>CcrA</td>
<td></td>
<td>28.9</td>
<td>29.3</td>
<td>20.5</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>BCII</td>
<td>28.5</td>
<td>21.7</td>
<td>29.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imp-1</td>
<td></td>
<td>18.7</td>
<td>27.9</td>
<td></td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>cphA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BlaB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5  Sequence alignment of known métallo β-lactamase sequences.

Key to métallo β-lactamase sequences: L1 - *Stenotrophomonas maltophilia*; BCII - *Bacillus cereus*; IMP-1 - *Serratia marcescens*; ccrA - *Bacteroides fragilis*; BlaB - *Chrysobacterium meningosepticum*; cphA - *Aeromonas hydrophilia*. The residues in red are those involved in zinc ligation. These are invariant except in cphA in which the first His is replaced by an Asn. The residues coloured in green are invariant whilst the residues in blue are conserved in all but a single exception amongst the known métallo β-lactamases. The alignment was carried using CLUSTALW (Thompson *et al.*, 1994) and the figure was made using Alscript (Barton G, 1993).
The crystal structures of the metallo β-lactamases, BCII (Carfi et al., 1995, 1998a; Fabiane et al., 1998) and ccrA (Concha et al., 1996, 1997, Carfi et al., 1997, 1998b), have been determined (figure 6). The overall structures of these enzymes are similar, consisting of two approximately equivalent αβ domains that form a αββα fold. The active site is situated in a groove created by the core β sheets at the interface between the two domains and is flanked by two loop regions named A and B. Loop A in BCII and ccrA (residues 32-39 in BCII and residues 27-35 in ccrA) has high temperature factors and it has been suggested that this region may act as a ‘flap’ that can move toward the active site and participate in substrate binding. Loop B contains a strictly conserved glycine (Gly179/192 in BCII/ccrA) and a largely conserved asparagine (Asn180/193 in BCII/ccrA) and may also be involved in substrate binding.

There are differences in the stoichiometry of zinc binding amongst the metallo β-lactamases. In L1 and ccrA, there are two zinc ions per molecule of enzyme (Crowder et al., 1996, 1998). In BCII there are two zinc binding sites, with differing affinities (1μM, and 25mM respectively (Davies et al., 1974)), but only one zinc ion was observed in the structure determined by Carfi et al. (1995). A later structure of BCII, by Fabiane et al. (1998), at higher resolution (1.9Å) showed that two zinc ions were bound to the active site albeit with differing B-factors (57.5Å² and 31.5Å² respectively). Finally, the Aeromonas enzymes have been found to be active with only one zinc at the active site and the addition of a second zinc leads to inhibition. (Vallaradasares et al., 1998).
The structures of ccrA and BCII.

The structure is composed of 2 core \( \beta \)-sheets flanked by two \( \alpha \) helices on either side. The active site is situated in a groove at the interface between the 2 sheets. The active site residues are provided by the loops surrounding the active site groove. The zinc ions are shown as grey spheres. BCII binds one zinc whereas ccrA binds two zinc ions.

The active sites of BCII and ccrA are quite similar with respect to the zinc coordination. In the structure of BCII, zinc 1 is tetrahedrally co-ordinated to three histidines (His86, His88 and His149) and a water molecule (Wat1) (Carfi et al., 1995, 1998; Fabiane et al., 1998). In the two-zinc structure of BCII (Fabiane et al., 1998), the second zinc is co-ordinated to Asp90, Cys168, and His210 and 2 water molecules.
Wat1 and Wat2. The geometry is trigonal bipyramidal. In the structure of ccrA, determined by Concha et al., the 2 zinc ions are separated by 3.5 Å. Zn1 is tetrahedrally co-ordinated by His99, His101, His162 and Wat1; co-ordination of zinc2 is by Asp103, Cys181, His223, and two water molecules Wat1 and Wat2 that adopt trigonal bipyramidal geometry. Wat1 is called the shared hydroxide as it is co-ordinated to both zinc ions. In contrast, in the bi-zinc structure of BCII determined by Carfi et al. (1998a), Zn2 is not co-ordinated to Wat1 and Wat2. Instead a carbonate ion is observed to act as a bidentate ligand around the two zinc ions.

1.6 Catalytic Mechanism of enzymes using the OH nucleophile.

1.6.1 Active site serine enzymes.

The role of serine in the active sites of enzymes has been best studied for the serine proteases. Similarly to the β-lactamases, the serine proteases catalyse the hydrolysis of an amide bond. In chymotrypsin, one of the best understood examples of the serine proteases, Catalysis is facilitated by three key residues, through their chemical nature and orientation within the active site, known as the catalytic triad. Typically the catalytic triad consists of a histidine, an aspartate and a serine (Neurath H. 1989 and references therein). The histidine acts as a general base to abstract a proton from the serine hydroxyl side chain thereby increasing the nucleophilicity of this serine. The third group in this triad is an acidic residue, usually aspartate, which acts to orient the histidine residue and also to neutralise the charged histidine intermediate. The mechanism involves the serine making a nucleophilic attack on the carbonyl carbon of the amide group to form an acyl enzyme. The amide bond is then cleaved by back donation of a proton from the histidine, now acting as a general acid, to the amide nitrogen. Deacylation of the enzyme then follows and involves an activated water molecule.
The catalytic machinery of the serine proteases, however, does not always consist of the catalytic triad. Indeed with certain proteases a Ser/Lys dyad has been found to be the catalytically active group (Paetzel M et al., 1997). In this mechanism, the lysine is thought to act as a general base, i.e., play a similar role to the histidine in the classical catalytic triad, and abstract a proton from the active site serine. However, for lysine to do this, its side chain must be in the deprotonated state. The pKa of lysine is normally around 10.8 and therefore it is not clear how the lysine would be able to exist in its deprotonated form at physiological pH. One possibility is that a positive charged active site may aid in the lowering of the pK of this lysine. It is also thought that hydrophobic residues in the vicinity of the active site could contribute to the lowering of the lysine's pK. Buried lysine's have been previously found to have pK's as low as 6.5 (Doa-Pin et al., 1991). The absence of an acidic residue is crucial to the dyad mechanism. The presence of a negative charge near the residues that comprise the dyad would lead to an elevated pK of the catalytically implicated lysine thus disabling its ability to act as a general base. An alternative role of the conserved lysine however, could be to increase the local positive charge in the vicinity of the active site serine so that its hydroxyl group is already deprotonated and thus removing the need for a general base in the acylation step (Moews PC et al., 1990).

The Ser/Lys dyad is present in all serine β-lactamases but the mechanism of the enzymes of the various classes appear somewhat different. With the Class A enzymes, which contain the additional element the omega loop, there were two candidates for the role of general base, namely Glu 166 (from the omega loop), and Lys 73. A pH rate profile of this enzyme revealed that two groups were important for catalysis with pK's of 5.6 and 9 (Page M et al., 1998). The general base would be expected to correspond to the lower pK of 5.6. Subsequently, the role of Lys 73 as the general base was found to be unlikely from 13C NMR studies which showed the lysine to have a normal pK (Damblon et al., 1995). However, there was conflicting
evidence was found for Glu 166 as the general base. A Glu166Asp mutant of the Bacillus cereus I enzyme (class A) showed decreased rate constants for both the acylation and deacylation steps, suggesting that this residue may playing a similar role of the histidine residue in the classic catalytic triad (Leung et al., 1994). However, other Glu 166 mutants have suggested that only the deacylation step is affected (Escobar et al., 1991). As a consequence of these findings, it is generally accepted that Glu 166 acts as a general base in the deacylation reaction, but its exact role in the acylation process remains poorly understood as is the role for Lys 73.

With the Class C β-lactamases, the absence of an equivalent residue to Glu166 is very likely to result in a more positively charged active site. Such an environemnt may indeed be able to lower the pK of the conserved lysine so that it could act as the general base. However, superimposition of the active site of a class C enzyme (from Citrobacter freundii) on the corresponding groups of chymotrypsin, it was observed that the oxygen group of the conserved tyrosine (in element 2) of the former was in a position equivalent to the proton accepting imidazole nitrogen of the catalytic triad His of the latter (Oefner et al., 1990). Hence making this tyrosine a candidate for the general base. However, the normal pK of the tyrosine hydroxyl group is 10.8 and thus would have to be considerably lowered. This was suggested to occur through electrostatic affects generated by the interaction of the side chain lysine from element 1 and a second conserved lysine with the tyrosine oxygen atom via hydrogen bonds (Damblon et al., 1995).

In summary, the mechanism of serine based proteolytic enzymes and the serine β-lactamases are similar probably due to convergent evolution. That is, there is an absolute requirement for a serine that is activated either through a general base or due to the accumulation of positive charges at the active site. The catalytic core can be presented in the form of the classical catalytic triad or the Ser:Glu/Lys/Tyr dyads. In
In the latter cases, the high normal pKa of the general base may be lowered by hydrogen bonding and hydrophobic affects. The active site provides residues that form the oxyanion hole required to stabilise the oxaynanionic tetrahedral acyl enzyme inermediate. In chymottrpsin, the oxyanion hole is comprised of the main chain NH-groups of Ser 195 and Gly 193. Similar oxyanion holes have been observed in the serine β-lactamases (Ser 70 and Ala 237 in class A and Ser64 and Ser/Ala318 in Class C) and the PBP’s (Ser 62 and Thr 301 in the S. R61 DD-peptidase). The formation of the tetrahedral intermediate requires two proton transfer steps, i.e; proton removal from the attacking serine and proton donation to the departing amide nitrogen. Deacylation also involves two proton steps, i.e; removal of a proton from a water molecule by the general base and then donation of a proton to the departing serine.

**Figure 7** A schematic of the general mechanism proposed for serine β-lactamases.

The general base is depicted as B. The mechanism has been discussed in detail in the text. This figure was adapted from Page et. al. 1998.
1.6.2 Active site aspartic enzymes.

Aspartic proteases are commonly found in both prokaryotes and eukaryotes and include the mammalian enzymes pepsin, gastricin, renin, the lysosomal enzyme cathepsin D and the retroviral proteases (eg HIV-1 protease). The aspartic proteases share considerable sequence and structural homology and are characterised by having two aspartyl residues at the active site. The structure of the aspartic proteases consist of 2 similar domains, that are related by an approximate two fold symmetry axis, linked by a third linker domain (Suguna K et al., 1987 and references therein; Lin et al., 1994). In the HIV protease, the two domains are identical and the enzyme is actually a homodimer (Abelardo S et al., 1996 and references therein). The active site is formed in a large cleft between the two domains. The catalytic aspartyl groups are located on domain 1 and 2 (monomer 1 and 2 in the case of HIV-1) respectively and are related by an approximate two fold symmetry axis. A water molecule is found between the two aspartyl residues that is within hydrogen bonding distances. This water molecule was found to be tightly bound and is proposed to be the reaction nucleophile.

The exact mechanism of proteolysis by these enzymes is not clearly understood but thought to involve a general acid general base mechanism. The general consensus is that the bridging water molecule is activated by a general base mechanism to enable it to carry out a nucleophilic attack on the substrate carbonyl. As with the serine proteases, a general acid mechanism is then expected to result in the protonation of the peptidyl amide group and hence facilitate peptide bond lysis (Suguna K et al., 1987; Meek T et al., 1994; Abelardo S et al., 1996).

1.6.3 Metallo enzymes.

The metallo enzymes that utilise the OH nucleophile include, amongst others, the enzyme carbonic anhydrase, the metallo proteases and the metallo β-lactamases. The
Metal ion involved is usually zinc. The structures of these enzymes are unrelated and do not appear to share a similar catalytic mechanism outside closely related family members. These enzymes may bind one or two zins at the active site (Coleman J E, 1998); the mechanistic importance of this and other features found in the various groups of metalloenzymes are described in the following text.

**Carbonic anhydrase (CA).**

CA catalyses the hydration/dehydration of carbon dioxide (Keilin et. al., 1940) and thus is not functionally related to the zinc proteases and the zinc metallo β-lactamases. The similarity lies in the utilisation of a zinc bound water molecule as the reaction nucleophile. Indeed CA is considered the prototype for enzymes that use a zinc activated water molecule (or hydroxide) as the reaction nucleophile.

There are in total seven isoforms of CA and crystal structure are known for four of these. All the known structures have a similar fold and contain central ten stranded β-sheet that is the major secondary structure element. The active site is located in a cone like cavity and contains a single zinc that is coordinated to three histidine residues and the oxygen atom of a water molecule. The zinc bound water is incorporated into a hydrogen bond network that includes Thr199 and Glu106. Similar to the zinc proteases, the zinc bound water is first activated by a general base mechanism to produce the more nucleophilic hydroxide species. In CA, the general base is proposed to be the zinc ligand His-64 (Lindskog S. 1997 and references therein; Coleman J. E. 1998).

**Zinc proteases.**

The zinc proteases have been classified according to sequence into about 30 families. These families however can be amalgamated into fewer groups on the basis of their zinc binding motif. Approximately half of all the known metalloproteases have the
HEXXH motif. This motif can occur sometimes by coincidence in a sequence and it has been found that actually only 20% of protein containing the motif are metallo proteases. Subsequently, this motif has been redefined as abXHEbbHbc. Here, a is commonly a Val or a Thr; b is any uncharged residue; c is any hydrophobic residue and X is any residue except proline as it has been found that the above motif forms part of a helix. The HEXXXH motif has been found in thermolysin, matrilysin, collagenase and autolysin to name but a few examples. In all of the above, metal ion is usually zinc and the coordination sphere around the metal is completed by a water molecule. The catalytic mechanism in these enzymes involves the activation of zinc bound water molecule, usually by a glutamic acid residue acting as a general base, to carry out nucleophillic attack on the substrate carbonyl (Coleman J. E. 1998).

Metallo proteases that bind zinc without the HEXXXH motif include the carboxypeptidase family of enzymes, the zinc D-Ala-D-Ala carboxypeptidase and aminopeptidases (Coleman J.E. 1988). In these enzymes, the zinc ligands are primarily His residues. The catalytic mechanism of these are nevertheless similar to the HEXXXH enzymes, and is proposed to involve a glutamate residue that activates a zinc bound water molecule.

**Bi-nuclear zinc enzymes.**

As mentioned earlier, the metallo β-lactamases contain two zinc binding motifs and generally bind two zins in the active site. Binuclear zinc complexes have also been observed at the active sites of several aminopeptidases (Strater et. al., 1995) and phosphotriesterases (Vanhoucke et. al., 1996). The zinc ions in all these enzymes are usually around 3Å apartand are usually bridged by a water/hydroxide molecule. The need for a general base mechanism to activated the shared water seems unlikely in these enzymes as the pK of the di-zinc bound water is likely to be sufficiently low for it to exist as a readily nucleophillic hydroxide (Bounaga et. al., 1998).
β-lactamases are the main focus of this thesis, a more detailed discussion of substrate binding and catalysis is given in the next section.

1.7 Metallo β-lactamases: binding and catalysis.

Currently, for the metallo β-lactamases, little is known in regards to the catalytic mechanism and substrate binding. Several non-physiological inhibitors of these enzymes do exist and a structure of ccrA in complex with one of these has been determined (Toney et al., 1998). No structural information is available for the enzyme substrate complex, however molecular modelling techniques have been used to study binding and investigate the catalytic mechanism.

In substrate modelling studies Concha and co-workers (1996) made 3 assumptions, based on the structure of ccrA, regarding substrate binding. The first assumption was that Wat1, the proposed nucleophilic group, would be close to the carbonyl carbon of the β-lactam ring in order to initiate nucleophilic attack. Secondly, an electrostatic interaction was proposed to occur between the side chain amino group of a largely conserved lysine (Lys184) and the carboxyl moiety of the β-lactam substrate. The final assumption was that the carbonyl oxygen of the β-lactam ring forms an electrostatic interaction with the largely conserved asparagine (Asn193), and a more distant interaction with the tetrahedrally co-ordinated zinc (Zn1). Together the interactions involving Asn193 and Zn1 would provide an oxyanion hole that enhances the polarisation of the carbonyl oxygen of the β-lactam. Three β-lactam antibiotics, ampicillin (a penicillin), ceftazidime (a cephalosporin) and imipenem (a carbapenem) were modelled into the active site of ccrA. The modelling suggested that in addition to the assumed interactions, additional contacts would occur between the hydrophobic
susbtituents of the substrate and loop A. It was suggested that this loop could fold over to favour substrate binding. A catalytic mechanism consistent with this model of substrate binding was proposed (figure 8).

Figure 8  Mechanism proposed by Concha et al. (1997).

Concha et al. (1997) propose that the zinc bound hydroxide carries out nucleophilic attack on the substrate carbonyl. This carbonyl is likely to be polarised by the positively charged environment at the active site. The nucleophilic attack leads to the formation of a tetrahedral oxyanion species that is stabilised by the oxyanion hole. This is thought to be the transition state. Substrate hydrolysis is then facilitated by proton transfer from Wat2 to the β-lactam nitrogen.
A somewhat different mechanism was proposed by Carfi et al., (1995), based on modelling work with the structure of BCII with a single zinc bound at the active site. In their mechanism, the substrate is proposed to make similar interactions and the reaction nucleophile is the zinc bound water. However, they propose that the conserved aspartate (Asp88) acts as first as a general base and deprotonates Wat1, to produce the OH nucleophile, and subsequently as a general acid by protonating the β-lactam nitrogen. A similar mechanism has been proposed by Bounaga et al., (1998), for BCII. However, the latter authors suggest that Wat1 will already be present as a hydroxide ion at physiological pH (the pK of this water may be lowered in the positively charged active site). The conserved aspartate is instead proposed to deprotonate the transition state to form a dianionic species. The aspartate subsequently relinquishes its proton to the β-lactam nitrogen as before (Figure 9).

Figure 9  Mechanism proposed by Bounaga et al. (1998).
1.8 Inhibitors of the metallo β-lactamases.

Metallo β-lactamases are inhibited by EDTA and EGTA but inhibition is non-specific and involves sequestration of the active site metal ions and is therefore of limited clinical utility. Fortunately, numerous other small molecules have been discovered that can reversibly or irreversibly inhibit the metallo β-lactamases in a more specific manner.

Fitzgerald et al., (1998), found that 2-(N-Morpholino) ethanesulfonic acid (MES) was a competitive inhibitor of the class B enzyme, ccrA, from Bacteroides fragilis. Kinetic studies indicated that the inhibition constant was 23mM. The crystal structure of ccrA with MES bound at the active site was determined and showed that binding involved loop A which undergoes a conformational change from the substrate-free form. The information from the MES-metallo β-lactamase complex was used in modelling studies, in parallel with high throughput screening methods, to identify other possible inhibitors. This led to the discovery of the biphenyl tetrazoles (BPT’s) which are potent competitive inhibitors of ccrA and the IMP-1 metallo β-lactamase (see reference below). Importantly, inhibition is accompanied by increased susceptibility of imipenem resistant bacteria to imipenem.

The structure of the complex of a BPT and ccrA was recently solved by Toney et al., (1998), (figure 10). The BPT molecule sits in the active site and is capped by loop A. The positions of the zins were unchanged but the co-ordination was altered; the shared hydroxide (Wat1) was absent and the N1 atom of the tetrazole moiety replaced the apical water (Wat2). Interaction between the inhibitor and loop A mainly involves a tryptophan residue (Trp32) and the second biphenyl ring. This tryptophan residue is also present in IMP-1 encoded metallo β-lactamases but absent from BCII, L1 and

39
the *Aeromonas* enzymes. This may explain why BPT’s are unable to inhibit the enzymes from the latter species.

**Figure 10**  A BPT molecule bound at the active site of ccrA (Toney *et al.*, 1998)

The structure of the active site region of ccrA is shown as a space filling model. The inhibitor, shown in yellow, is docked into the active site. Loop A has folded over to cover the active site and is involved in inhibitor binding. The red colour shows regions of negative electrostatic potential.
Another class of metallo β-lactamase inhibitors are the mercaptoacetic thiol ester derivatives (MTE’s) and thiol compounds (Payne et al, 1997). These compounds are effective inhibitors of the IMP-1 encoded metallo β-lactamases, BCII, L1 and the *Aeromonas* enzymes but do not inhibit ccrA. Inhibition is irreversible in BCII but reversible with the other enzymes. The structure of such an enzyme-inhibitor complex is yet to be determined.

Tri-fluoromethyl ketone (TFK) and alcohols (TFA) are another class of inhibitors of the metallo β-lactamases. These molecules have also been shown to be inhibitors of serine proteases (Brady *et al.*, 1990) and the zinc dependent carboxypeptidase A (CPA) (Christianson *et al.*, 1986). In the latter case, the crystal structure of the complex of CPA and a TFK has been determined to 1.7Å. Several TFK’s and TFA’s with inhibitory affects on class B enzymes have been synthesised (Walter *et al.*, 1996) and have been shown to be competitive inhibitors of L1, BCII and IMP-1 (Walter *et al.*, 1997). The *Aeromonas* enzymes are inhibited irreversibly but ccrA is not affected. The mechanism of inhibition in all these cases is unknown.

1.9 *The metallo β-lactamase L1, from S. maltophilia.*

*S. maltophilia* is an opportunistic pathogen and is ubiquitous in nature. The organism can survive and prosper in many environments including water, soil, sewage, raw milk, frozen fish, human faeces, sinks and soap solutions. It has also been found in hospital disinfectant solutions, water units in ICU wards and respiratory therapy units (Jang *et al.*, 1992). *S. maltophilia* does not colonise skin or the gastro-intestinal tract but is commonly found in the respiratory organs. Men appear to be more prone to infection than women and it is also more frequent in patients with cancer, HIV, cystic fibrosis and those recovering from organ transplants (Liang *et al.*, 1995; Garrison *et al.*, 1996;
Lipton et al., 1996; Denton et al., 1997; Elsner et al., 1997; Penzak et al., 1997). Risk factors associated with infection include broad-spectrum antibiotic therapy and hospitalisation. There is also a strong correlation between exposure to imipenem and *S. maltophilia* infections (Khadori et al., 1990).

In *S. maltophilia* β-lactam resistance is mediated by two chromosomally co-ordinated β-lactamases named L1 and L2 (Saino et al., 1982; Dufrense et al., 1988; Bicknell et al., 1985; Paton et al., 1994; Crowder et al., 1998). L1 is a metallo enzyme belonging to molecular class B with a pI of 6.9 and molecular mass of 114KD (Ullah et al., 1998; also see later) composed of four subunits of approximately 28KD each. L2 belongs to molecular class C and behaves as a cephalosporinase. Both β-lactamases are inducible and appear to share the same regulatory mechanism (Paton et al., 1994). L1 has a wide spectrum of activity that includes virtually all the β-lactam antibiotics (Felici et al., 1993, 1995). The exceptions are the monobactams, however, *S. maltophilia* is not susceptible to these substances because they are good substrates for L2 (Paton et al.; 1994). Although the enzyme is tetrameric, there is no evidence for allostery (Felici et al., 1993, 1995). The crystal structure of L1 has been determined in this work and is discussed in later chapters.
1.10 Project aims.

The objectives of this work were to carry out structural and mechanistic studies on the metallo β-lactamase L1 from *Stenotrophomonas maltophilia*. The aims were thus to:

- Determine the three dimensional structure of the metallo β-lactamase by X-ray crystallography.

- Investigate the mechanism of action and possible modes of substrate binding, possibly by using mutagenesis and kinetic studies.

- Study enzyme-inhibitor complexes by X-ray crystallography.

The first of these aims was achieved and is described in this thesis. The latter two objectives were attempted but not completed and is work that is being continued in the lab.
Chapter 2.

Introduction to structure determination of biological macromolecules by X-ray crystallography.
2.1 Introduction

X-rays can be used to study the three-dimensional structures of biological macromolecules because the range of wavelengths of the radiation involved is of the same order of magnitude as the separation of atoms in the molecules. This property enables molecules like proteins to act as diffraction gratings. However the intensity of diffracted X-rays from a single protein molecule is negligible and no structural information can be obtained from such an experiment. The science of protein crystallography was born by the discovery that proteins could be crystallised. The crystal phase is that between the liquid and the solid phase; that is, it is at a state of minimum solubility. At this point, proteins are driven to form a 3D crystal lattice in order to reach an energy minimum. A crystal is composed of identical unit cells each related to each others by a translation vector. Each unit cell may contain a single or several protein molecules, related by the symmetry of the crystal. The basic unit that can form the unit cell by application of the crystal symmetry is called the asymmetric unit.

Diffraction of X-rays from a crystal can be observed because it serves to amplify the diffraction pattern of a single molecule. Diffraction occurs when certain conditions, described by the Bragg equation (equation 1), are satisfied. The principles of X-ray diffraction have been extensively reviewed and will not be discussed here (Glusker and Trueblood, 1985; Helliwell J R, 1992; McRee D E, 1993; Methods of Enzymology, 1997).

\[
\text{Bragg's Law : } n\lambda = 2d \sin \theta \quad [1]
\]

Where \( n \) is any integer, \( \lambda \) is the wavelength, \( d \) is the interplanar spacing and \( \theta \) is the angle of incidence of the x-rays on the atom plane.
The scattering of X-rays by atoms may be described in terms of the structure factor equation (equation 2). The complex structure factor is regarded as being made up of two components. The first is the amplitude of each reflection, which can be determined directly from the diffraction pattern and secondly the phase of each reflection. For macromolecular crystals the phases cannot be directly determined and a number of procedures have been developed to address this problem (see later).

\[ F(hkl) = \sum f_j \exp(2\pi i(hx_j + ky_j + lz_j)) \]  \[2\]

Where \(f_j\) is the atomic scattering factor of the \(j\)th atom and is chosen so as to account for isotropic thermal vibration. The \(2\pi i(hx+ky+lz)\) component is the phase of the scattered wave from the set of planes \(hkl\) relative to a scattered wave in the same direction from the origin of the unit cell. The amplitude of the reflection \((hkl)\) is given by \(|F(hkl)|^2\).

### 2.2 Crystallisation

If a protein can be produced in large quantities and at high levels of purity, it may be a candidate for structure determination by X-ray crystallography. However the process is not without difficulty. Essentially there are two major rate limiting steps; the first, and the most important, is to obtain crystals of the protein of interest. The principles of crystallisation have been extensively studied but it remains difficult to predict the conditions under which a protein will crystallise. There are many factors that can affect crystallisation including pH, temperature, sound and vibration, protein purity and concentration, and the solvents used to induce crystallisation. These have been extensively reviewed (McPherson A Jr, 1982).

The principle behind a crystallisation experiment is to induce a state of minimum solubility and thereby encourage the protein to form a 3D crystal lattice. Crystallisation kits based on the precipitating properties of salts (e.g. \((NH_4)_2SO_4\)) and organic solvents (e.g. Polyethylene glycol) have been developed and are commonly
used to screen for conditions that result in crystallisation. The conditions required for nucleation and for crystal growth are usually very narrow and not necessarily the same. The figure below (figure 11) shows a crystallisation diagram. Once a condition is found, fine screening by adjusting certain parameters can be carried out to streamline the crystallisation so as to obtain better crystals.

**Figure 11** Crystallisation diagram.

![Crystallisation diagram](image)

P - precipitation zone, N - nucleation zone, M - metastable zone and S is the supersaturation curve. A crystal will not be formed beneath the S curve or in the precipitation zone.
In any crystallisation experiment, crystals will not be formed underneath the saturation curve (S) or in the precipitation zone (P). If the experiment results in the formation of amorphous precipitate, the conditions have induced super-saturation too fast or pushed it beyond the minimal solubility state. Most crystals will form in the nucleation (N) phase. Here, spontaneous crystallisation may occur, however, a drawback of N phase is that many nuclei may form thus limiting the growth of the crystals. In such cases it may be necessary to set up a similar experiment but adjust the conditions so as to be at the metastable phase (M). Crystals from the original experiment can be transferred to the new experiment, by a process referred to as seeding, in order to obtain bigger crystals. This is because in the M phase spontaneous nucleation cannot occur, but protein molecules in solution can crystallise by adding to pre-existing crystalline nuclei. A popular method for inducing crystallisation is the vapour diffusion technique. The experimental set up can be of two types, the sitting drop or the hanging drop (see figure 12), but the principle in both cases is the same.

In the hanging drop method, a certain volume (e.g. 0.5ml) of the crystallisation solvent (a buffer at a certain pH, and a precipitant (e.g. PEG)) is transferred to a crystallisation well. A small amount (e.g. 1μl) of concentrated protein is placed on a clean siliconised cover slip; to this protein droplet an equivalent amount of the crystallisation solvent is added. The cover slip is then used to seal the well in the crystallisation tray with the aid of some oil or vacuum grease. The experiment is now airtight and under these conditions a concentration gradient is established between the precipitant in the well and the droplet. Equilibrium is reached by the distillation of water out of the protein droplet. As a result, the protein molecules become more concentrated and at the point of minimum solubility the proteins molecules may arrange themselves on a 3D crystal lattice.
Figure 12  Hanging drop method.

The coverslip is sealed to the top of the well using oil or vacuum grease to provide an airtight environment.
2.3 **Determination of structure factor phases.**

Macromolecular structure factor phases may either be determined by some variation of heavy atom replacement or by molecular replacement (MR). As MR was not successful in this project we shall be concerned here only with heavy atom based phasing procedures. Heavy atoms can be introduced into the protein crystal either by soaking the crystal in solutions containing heavy metals in the millimolar concentration range or by substitution of derivatised amino acids. It is often difficult to get high occupancy substitution of heavy atoms and even then the resulting crystal may not be isomorphous with the native crystal. When a successful derivative is found, there will be a difference in the intensity of the diffraction pattern between the native and derivative crystals, which is due to the bound heavy atom i.e.;

\[ F_H = F_{PH} - F_P \]

\( F_P \) is the structure factor corresponding to the native protein; \( F_{PH} \), the structure factor corresponding to protein-heavy atom derivative; \( F_H \), describes the structure factor of the heavy atom alone. The differences in the magnitudes of \( F_H \) and \( F_{PH} \) can be used in a Patterson function to locate the position of the heavy atoms.

A Patterson map is an originless vector map of all the atoms in the unit cell. It can be calculated directly from the diffraction data using just the \( |F|^2 \) terms of each reflection as no phase information is required. The Patterson Function is defined in equation 3.

\[
P(u,v,w) = \frac{1}{Vc} \sum \sum_{h,k,l} |F|^2 \cos 2\pi (hu + kv + lw) \tag{3}
\]
The patterson function can be regarded as a convolution of the electron density at all points \(xyz\) in the unit cell with the electron density at \(x+u, y+v\) and \(z+w\). Peaks will be observed in the Patterson map if \((u,v,w)\) corresponds to a vector that relates 2 atomic positions.

A difference Patterson is calculated by squaring the differences between the structure factor amplitudes for the native protein and the derivatised protein. The resulting Patterson map can be solved by either visual inspection of the Harker sections (see below) or by automated procedures such as HASSP (CCP4 1994). Harker sections are special planes that arise from the crystal symmetry (Harker D, 1936). For example in space group \(P6_422\), vectors between equivalent atoms in two molecules related by the 6 fold screw axis will lie on a plane one third along and perpendicular to the six fold axis. In the same way, the crystal two-fold axis will generate another Harker section perpendicular to the twofold passing through the origin. Harker sections contain peaks due to vectors between symmetry related heavy atoms; the position of these peaks reveal the location of the heavy atoms. Once the atomic positions are known, phases can be calculated from the native and derivative amplitudes. With just a single derivative two equally probable phases are found. This ambiguity can be resolved by using another heavy atom derivative in the MIR case (Multiple isomorphous replacement) or by including anomalous scattering information; this is often referred to as MIRas (MIR with anomalous scattering).

2.4 Multiwavelength Anomalous Dispersion (MAD).

Another method of phase determination is by the Multiwavelength Anomalous Dispersion (MAD) technique. This is a variation of the heavy atom replacement method but unlike MIR, all the phasing information can be obtained from a single
crystal. This method was used in this work for phase determination and is briefly described in the following paragraphs. MAD phasing is reliant on the two sets of information that arise from the dispersive and the anomalous contributions to anomalous scattering. The MAD dispersive signal is obtained from changes in the real component of the scattering from the anomalous centre that arises between the same reflections measured at different wavelengths. The anomalous signal manifests itself as the Bijvoet differences between Friedel mates.

The theoretical possibility of MAD was understood long before it became a practical technique. With the emergence of tuneable synchrotron sources, which can provide very intense and finely focused monochromatic X-rays, methodologies were developed and successfully used by Hendrickson (1991) and others (Smith J L, 1991) to overcome the critical phase problem in macromolecular crystallography. The advantages of MAD over isomorphous replacement (MIR, SIR) methods include the removal of non-isomorphism as all the necessary data can be collected from a single crystal. MAD offers better phasing potential at high resolution as unlike scattering from a heavy atom derivatised crystal, the anomalous scattering intensity does not fall off as a function of the resolution.

To carry out a MAD experiment the protein being studied must contain an atom that exhibits significant anomalous scattering at a wavelength between 0.8-1.8Å. As such, there are only a few atoms that can be exploited by MAD for protein phase determination. Examples are given in table 3, which shows the more commonly used heavy atoms in protein crystallography. Elements that have measurable L edges would be the most useful in MAD as a large anomalous signal could be attained, however, it is not always possible to use these atoms because they are not normally found as intrinsic atoms in proteins. These atoms can be added to the crystals as in isomorphous replacement (MIR) but as such, are prone to the problems associated
with MIR except for isomorphism (although it is often said that there are three problems with MIR; lack of isomorphism, lack of isomorphism and lack of isomorphism). The anomalous component of the atoms commonly found in macromolecules (e.g. H, C, O, N, S, and P) is negligible and does not affect the experiment (Hendrickson et al., 1991).

Table 3 Examples of atoms that could be used for MAD phasing.

<table>
<thead>
<tr>
<th>Atomic properties</th>
<th>Absorption edges (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element</td>
<td>Atomic No.</td>
</tr>
<tr>
<td>Co</td>
<td>25</td>
</tr>
<tr>
<td>Zn</td>
<td>30</td>
</tr>
<tr>
<td>Se</td>
<td>34</td>
</tr>
<tr>
<td>Cd</td>
<td>48</td>
</tr>
<tr>
<td>Pt</td>
<td>78</td>
</tr>
<tr>
<td>Hg</td>
<td>80</td>
</tr>
</tbody>
</table>

This table was reproduced from Helliwell (1992).

Anomalous scattering is dependent on wavelength and thus, only results when the energy of the X-ray beam is near the absorbance edge of the anomalous scatterer to be exploited. Often the signal from anomalous scattering is quite small i.e.; the difference between the overall intensity of normal data set and a data set with an anomalous component may only be a few percent. Therefore it is important to collect data of appropriately high quality.

In normal situations, i.e., at an energy above or below the absorption edge, X-rays are scattered ‘normally’; that is, the electrons behave as free entities and scatter the X-
rays with a constant phase change of 180°. At the absorbance edge of a suitable anomalous scatterer, part of the energy of the X-ray is scattered normally, and part of it is absorbed by the atom. This absorption results in the transition of inner shell electrons to a higher energy orbital in the continuum. However, the transferred electron still behaves like an inner shell electron and act as a secondary scattering source; this leads to an anomalous scattering increment \( \Delta f \) to the normal scattering \( f^\circ \).

When the electron is ejected from the inner shell to higher energy state, another electron falls back to fill the void. In the process of dropping to the lower energy level, the electron emits the excess energy as fluorescence.

Normal scattering \( (f^\circ) \) is real and independent of wavelength but decreases as a function of resolution. In the presence of an anomalous scatterer, normal scattering is composed of the scattering from the protein \( (f^\circ p) \) and the anomalous scattering atom \( (f^\circ a) \); thus total normal scattering \( f^\circ \) is composed of \( f^\circ p + f^\circ a \). The anomalous scatterer also has an anomalous scattering component, \( \Delta f \) that is complex. \( \Delta f \) is dependent on wavelength and largely independent of resolution. \( \Delta f \) is composed of \( f' \) (dispersive signal) and \( f'' \) (anomalous signal), the real and the imaginary parts respectively. These two are orthogonal to each other (figure 13). Thus total scattering \( F \) is given by equation 4.

\[
F = f^\circ (p+a) + f' + if'' .
\]  

As a result of the anomalous component, the structure factor amplitudes are no longer equal for reflections related by Friedel's symmetry i.e.; reflections \( F(h, k, l) \neq F(-h, -k, -l) \).

This difference between Friedel pairs is called the Bijvoet difference.
The position of the anomalous scattering atom can be approximately determined from an anomalous and dispersive difference Patterson. Once this is known, the magnitude of $F_A$, phase $\phi_A$ and $'f$ and $''f'$ at $x_2$ and $x_3$ can be calculated. The magnitude of $F_A$ at $x_2$ and $x_3$ can be used to draw phase circles to determine $\phi_A$. 

$\phi_A$ - phase of $F_A$
$\phi_H$ - phase of $F_H$
$\phi_T$ - phase of $F_T$
$\phi_P$ - phase of $F_P$

$F_A$ - normal scattering from the anomalous scatterer
$F_H$ - total scattering including the anomalous component
$F_T$ - total normal scattering
$F_P$ - scattering from protein

$\lambda_1$ - far from the absorption edge of the anomalous scatterer
$\lambda_2$ - $'f$ maximum
$\lambda_3$ - $''f'$ minimum
2.5 Practical aspects of the MAD method.

The fact that the class B β-lactamases contain zinc ions provides a convenient way of determining phases by the MAD method. However, it would only be possible with enzymes containing binuclear zinc centres as a single zinc would probably not provide a sufficiently strong signal. This characteristic, together with the presence of two-fold non-crystallographic symmetry, was exploited by Concha and co-workers (1996) to solve the crystal structure of the metallo β-lactamase from *Bacteroides fragillis*. A powerful method of producing proteins for structural analysis by MAD is to introduce Se into the protein by growing the microorganism on a Se-methionine substrate instead of a methionine containing substrate. Hendrickson (1990) showed that the presence of one selenium atom (Se, atomic no. 34) in a protein of not more than 150 amino acid residues should be sufficient to carry out a MAD experiment. With a larger protein additional Se atoms would be required.

To obtain sufficient information from a MAD experiment, measurements have to be taken at carefully chosen wavelengths. Data collection should be designed to fully exploit $f'$ and $f''$ signals by selecting wavelengths at which these are near their optimal values. These optima occur near the absorbance edge of the anomalous scatterer present and thus the precise character of the absorbance edge has to be determined before a MAD experiment can be carried out. This is best achieved by measuring the fluorescence spectrum of the anomalous scatterer directly from the protein crystal. Theoretical values are unreliable as the position of the absorbance edge is somewhat sensitive to the chemical environment of the anomalous scatterer.

The datasets recorded at multiple wavelengths from the MAD experiment can be treated in a similar manner to an isomorphous replacement dataset with anomalous scattering (MIRas). The wavelengths corresponding to the optimal values of the
anomalous components \( f' \) and \( f'' \) are taken as two derivatives, as in a case of MIR (see earlier). A third remote wavelength can be used as the native. The differences in the diffraction intensities due to the bijvoet differences within a single dataset and differences between datasets at different wavelengths can be used to calculate an anomalous difference and a dispersive difference Patterson respectively. From the resulting Patterson maps the position of the anomalous centres can be located and then phases calculated as illustrated in figure 13.

2.6 Refinement.

The position of any atom in space can be described if 4 parameters are known i.e.; the co-ordinates of the atom \((x,y,z)\) and the temperature \((B)\) factor. The temperature factor accounts for the vibrations from the mean position of the atom. The structure factors used to generate the initial electron density maps are composed of the diffraction intensities and the calculated phases. The quality of fit of an atomic model to the diffraction data is given by the \( R \) factor, which is measure of the difference between the observed \((F_o)\) and calculated \((F_c)\) structure factor amplitudes:

\[
R = \sum_{(hkl)} \frac{|F_o| - |F_c|}{|\Sigma|F_o|} [5].
\]

Refinement is the process by which the positions \((x,y,z)\) and \(B\) factors of atoms in the model are modified so as to improve the agreement between the observed and the calculated structure factor amplitudes, while imposing acceptable model stereochemistry. The refinement process is limited by the quality and the resolution of the data collected as well as the quality of the initial model. Refinement should provide the most accurate atomic model possible within the limitations of the observed
X-ray data. Improvements in the atomic model also produce improved phases so that poorer regions of the electron density maps may become interpretable.

A low R factor however does not necessarily correspond to a good model, as it can be superficially lowered by refinement procedures due to the large number of adjustable parameters (i.e.; xyz and B of atoms) that are used to describe the model. Therefore, for accurate model refinement a system of cross validation of the refinement procedure was introduced (Brunger A T, 1992; Kleywegt G J, 1996). In this cross validation procedure, the diffraction data are divided into two sets; a ‘working set’, consisting of approximately 90 percent of the data, and a ‘test set’ consisting of the remaining data. The working set is used in the refinement process whereas the test set is not. However, at any stage of the refinement process, an R value, referred to as the free R factor, can be calculated using the test set.

The free R factor measures the degree to which the refined model (refined using the working set) predicts the diffraction data that was excluded (the test set) from the refinement. Every reflection in a diffraction dataset contains information about the entire structure and thus, changes to the model (due to refinement) that do not improve the fit of the model to the test set leads to either an unchanged or increased value of the free R factor. On the other hand, if refinement improves the model, the free R factor decreases in a similar way as the R factor. As a general rule, free R factors of more than 0.40 indicate a grossly incorrect model. Also as an approximate rule of thumb the working R factor should be 1/10 of the highest resolution limit for a correctly built model.
Chapter 3.

Methods and materials.
3.1 Production and preliminary characterisation of native L1.

Preliminary characterisation of the metallo β-lactamase L1 from *Stenotrophomonas maltophilia* has been reported (Bicknell *et al.*, 1985; Paton *et al.*, 1994). The complete nucleotide and the amino acid sequence of L1 has been determined by Walsh and co-workers (1994). Cloning of L1 has been described elsewhere (Crowder *et al.*, 1998). In this work, L1 was expressed in BL21(DE3)pLysS (Novagen). Colonies of the transformed expression strain were grown on L agar and Terrific broth in the presence of 34 micrograms per millilitre (µg/ml) chloramphenicol (Sigma) and 50µg/ml kanamycin (Sigma). The recipe for L agar and Terrific broth are given below:

**L. agar per litre (in distilled water)**-

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto tryptone</td>
<td>10 grams (g)</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5g</td>
</tr>
<tr>
<td>NaCl</td>
<td>10g</td>
</tr>
<tr>
<td>DIFCO agar</td>
<td>15g</td>
</tr>
</tbody>
</table>

**Terrific broth per litre (in distilled water)**-

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto tryptone</td>
<td>10g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>24g</td>
</tr>
<tr>
<td>Glycerol</td>
<td>4ml</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>12.54g</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>2.31g</td>
</tr>
</tbody>
</table>

A single colony of the transformed cells was used to inoculate 50mls of terrific broth, to which kanamycin (50µg/ml) and chloramphenicol (34µg/ml) had been added, and then grown at 37°C overnight in an incubator-shaker (model G25, New Brunswick Scientific Co. inc.) at 300rpm. The resulting bacterial culture was used to inoculate a further 6
flasks containing terrific broth (750mls) and antibiotics (see above) which were then grown at 37°C as before. The growth of the cells was followed by measuring the absorption at 600nm. When a cell density of 0.6 - 0.8 absorbance units was reached, the cells were induced using 1millimolar (mM) isopropyl-β-D-thiogalactopyranoside (IPTG) (Biogene) and grown under similar conditions for a further 3 hours. The resulting bacterial cultures were spun down at 4000 revolutions per minute (rpm) for 20 minutes in a J6 Beckman centrifuge. The cells, in the form of a pellet, were resuspended in a resuspension buffer (50mM Tris (BDH) pH 8.5, 1mM ZnSO₄ (BDH), 1mM β-mercaptoethanol (βME) and also 1mM pefabloc ((Boehringer Mannheim, GmbH), a protease inhibitor). The cells were then lysed using a Branson sonifer sonicator and centrifuged (J21 Beckman centrifuge at a speed of 18,000 rpm for 35 minutes) to separate the cell debris from the protein fraction.

3.2 Production of seleno-methionine incorporated L1.

Seleno-methionine incorporated L1 was produced by expressing the L1 gene in B834(DE3) cells (Novagen). This strain is a methionine auxotroph and therefore, a source of methionine has to be supplied to support growth. In this work seleno-methionine (Se-met) was used as a substitute for methionine and thus Se-met incorporated protein was obtained. The transformed cells were grown at 37°C overnight on Se-met incorporated agar (see below). Six single colonies were selected from the resulting growth and used to streak a further six Se-met agar plates, which were grown as before.
Preparation of Se-met medium and agar (100ml)

3 grams AGAR SELECT (Sigma).

97 ml Se-met medium

5 ml M9 (20x)

0.2 ml 1M MgSO₄

1ml 40% Glucose

1ml Amino acid mix I

2ml Amino acid mix II

0.1ml Vitamin mix

0.1ml Kanamycin (50µg/ml)

33µl 2M FeSO₄

0.13ml Seleno-DL-methionine

85.43 ml distilled water.

Amino acid mix I consist of all amino acid except Met, Tyr, Trp and Phe each at 2 milligrams per millilitre (mg/ml). Amino acid mix II consists of Tyr, Trp and Phe each at 2mg/ml. The Vitamin mix consists of riboflavin, niacinamide, pyridine, thiamine, and monohydrochloride each at 1mg/ml. Seleno-DL-methionine is used instead of just the L form because it is less expensive. Only the L form is available for incorporation during protein synthesis and to compensate for this, twice the amount of the DL form is used. All the materials above were obtained from Sigma. M9 is made up as a 20 times stock as below:

10g NH₄CL

30g KH₂PO₄

60g Na₂HPO₄

H₂O to 500ml.
A single colony from each plate was used to inoculate six flasks containing 50 mls Se-met medium and grown overnight in a shaker incubator at 37°C. Ten mls of each of the resulting cultures were then used to inoculate 6 further flasks containing (750mls) Se-met medium. The cells were grown as described for the native protein with the exception that post induction with IPTG, the cells were grown at 28°C. The reason for the lower temperature was to avoid problems of protein solubility. Cell harvest procedures were identical to that described for the native L1.

3.3 Purification of native and Se-met incorporated L1.

Purification of native L1 and Se-met incorporated L1 was carried out similarly and involved four steps.

(NH₄)₂SO₄ precipitation
The initial stage involved an (NH₄)₂SO₄ (BDH) precipitation step. An amount of solid (NH₄)₂SO₄ that would correspond to 25% saturation was weighed out, crushed to a powder, and then added to the crude protein extract. The protein solution was thoroughly mixed with the crushed (NH₄)₂SO₄ over a period of 30 minutes. After this, the precipitated material was then removed by centrifugation (J21 Beckman centrifuge for 30 minutes at 18,000 rpm). The supernatant was dialysed against a buffer consisting of 50 mM Tris pH7.5, 1mM βME, 1mM ZnSO₄ (hereafter referred to as Q-buffer) to remove the (NH₄)₂SO₄.

Anion exchange
The dialysed protein fraction was loaded on to a Q-sepharose anion exchange column (Pharmacia biotech) that had been pre-equilibrated with Q-buffer, using a Pharmacia
Biotech GradiFRAC system. The flow through was collected as the protein fraction of interest and then dialysed against a buffer consisting of 50mM NaAC (Sigma) pH5.0, 5mM ZnSO$_4$, 1mM βME (referred to as S-buffer). The anion exchange step exploits the neutral pI of L1 (6.9); that is, L1 does not bind to the column matrix but other proteins with acidic pI’s are bound and thus removed from the protein fraction.

**Cation exchange**

The dialysed fraction from the previous step was then loaded onto a S-sepharose cation exchange column (Pharmacia Biotech), pre-equilibrated with S-buffer. The bound fraction was eluted using a gradient of 0-500 mM NaCl (BDH) pH5.0 using the Pharmacia Biotech GradiFRAC system.

**Gel filtration**

Peak L1 fractions, identified by the hydrolysis of meropenem (Zeneca Pharmaceuticals) and by SDS PAGE, were loaded onto a S200 gel filtration column. The elution buffer consists of 50mM Tris pH7.0, 100mM NaCl, 1mM βME and 5mM ZnSO$_4$. The fractions with L1 enzyme activity from gel filtration were pooled and concentrated (to approximately 20mgs/ml using a MP4R centrifuge at a speed of 4000rpm) in centricon-50 concentrators (Amicon).

**3.4 Assessment of purification strategy and quality.**

The purification was assessed at each stage by SDS PAGE (Sodium Dodecyl-Sulphate Poly-Acrylamide Gel Electrophoresis (Schagger and Von Jagow, 1987)). Total protein assessments were carried out using the method of Bradford (1976). Standards were prepared from a 1mg/ml stock of bovine serum albumin (BSA) (Sigma). The procedure used was as described in the Bradford assay protocol booklet. Measurements were taken.
at 595nm on a Shimadzu UV-160 spectrophotometer. A calibration curve was plotted using the BSA standards from which the concentration of L1 at each stage could be approximated.

The amount of L1 that was active at each stage was determined by measuring the total activity in enzyme units (EU). An enzyme unit is defined as the amount of enzyme required to hydrolyse 1 micromole (μM) of substrate per minute. The concentration of substrate (meropenem) was 1mM and was prepared in 20mM Hepes (Sigma) pH7.0. Activity measurements were carried out at room temperature (approx. 20°C) on a Shimadzu UV-160 spectrophotometer at 330nm (ε=990 dm³ mol⁻¹). Total activity (EU) and specific activity (EU/mg) were calculated (total activity/total protein) to quantify the yield and assess the purification achieved respectively, at each step of the purification.

For crystallisation purposes, the L1 protein probably needs to be better than 95 % pure and mono-disperse. The final purity of the protein was assessed by SDS PAGE and by electrospray mass spectroscopy, using in-house facilities. The mono-dispersity of the concentrated protein solution was examined by dynamic light scattering (Dynapro Model 801, Protein solutions incorporated).

### 3.5 Determination of the relative molecular mass of L1.

Sedimentation equilibrium was used to determine the multimeric state of L1. This technique offers several advantages over other techniques such as gel filtration chromatography, light scattering, sedimentation velocity and SDS PAGE in that the shape of the molecule is not a factor and also, the protein is not denatured (SDS). Sedimentation equilibrium of a mixture (protein in salt buffer) can be established using a relatively low
speed centrifugation spin. During the centrifugation spin, a distribution of the solutes present in solution occurs so that an observable concentration gradient is established in the cell. At equilibrium, there is no net migration of the macrospecies since the rate of sedimentation resulting from the applied centrifugal field is balanced by the back diffusion induced by the concentration gradient. The distribution observed at equilibrium is dependant only on the buoyant molecular weight of the solutes present in the solution and thus, the molecular weight of the solute can then be determined using the following expression.

\[ M_r = \frac{2RT}{(1 - \nu \rho) \omega^2} \times \frac{d \ln C}{d(r^2)} \quad [7] \]

where,

- \( M_r \) - Relative molecular mass
- \( R \) - Gas constant.
- \( T \) - Temperature in Kelvin
- \( \nu \) - Partial specific volume
- \( \rho \) - Density of the solution.
- \( \omega \) - Angular velocity in radians/sec
- \( r \) - radial distance from the axis of rotation
- \( C \) - Solute concentration

Protein samples (at concentrations of 1mg/ml, 0.75mg/ml, and 0.5 mg/ml) were prepared by exhaustive dialysis against the buffer blank (50mM Tris, 200mM NaCl, 5mM ZnSO\textsubscript{4}) to remove βME added during purification. This was necessary because the absorbance of βME can lead to an overestimation of \( M_r \); that is, at the meniscus where the protein concentration is low, absorbance of βME will contribute significantly to the overall absorbance measured. 90 µl of dialysed protein and 110µl of the buffer blank were transferred to the sample and blank wells respectively in the analytical ultracentrifugation cell. The analytical cell was constructed as described in the Beckman XL-A analytical centrifuge practitioners guide.
The speed of the centrifugation run (10,000 rpm) was deduced from a calibration curve of expected molecular weight versus rpm from the aforementioned practitioners guide. After 24 hours the concentration of protein was scanned (λ=280nm) at increasing values of r (radial distance form the axis of rotation) at 2 hour intervals until there was no further change in the absorbance profile. The partial specific volume of the protein and the solvent density were estimated from tabulated values (Laue et al., 1992). The data at equilibrium was analysed using the Origin scientific data analysis program and the XL-A 2.01 data analysis supplement program.

3.6 Kinetic studies.

Kinetic studies were carried out to determine the Michealis constants \( K_m \), \( V_{max} \) and also \( K_i \), in the presence and absence of a tri-fluoromethyl ketone (TFK) inhibitor of the metallo \( \beta \)-lactamase L1. The reporter substrate, meropenem, was prepared in 20mM Hepes at pH7.0; no zinc was added as meropenem is hydrolysed in the presence of free zinc (Munn et al., 1994). The concentration of meropenem used was from 10\( \mu \)M to 250\( \mu \)M, which corresponds to just below and approximately 20 times greater than the reported dissociation constant with L1 (Felici et al., 1995). The concentrations of inhibitor ranged from 5 – 250 nM and were prepared in 5% methanol. The reported \( K_i \) of the inhibitors used in this study is around 5 nM (Walter et al., 1997).

The concentration of protein used to initiate the reaction was approximately 3nM and therefore nearly all of the enzyme was bound to the substrate at the start of the reaction. It was then possible to determine a pseudo first order rate constant for the breakdown of the complex to products. The data points for the calculation were taken within the first 10% of the reaction to minimise the affects of product accumulation. All measurements
were carried out at 25°C using a Perkin Elmar Cary 3E UV-Visible spectrophotometer at 297nm ($\lambda_{\text{max}}$ of meropenem); the extinction coefficient is $6.5 \times 10^{3}$ dm$^3$ mol$^{-1}$ (Felici et al., 1995). The data were analysed initially using Cricket Graph (Version 1.01 (1992), Computer Associates International, Inc. USA) and fitted to the Michaelis function using the program PROFIT (Version 4.2.1 (1994), Quantum Software, Switzerland).

3.7 Crystallisation of native and Se-met L1.

Crystallisation conditions of both native and Se-met incorporated protein were similar (100mM Hepes (Sigma) pH 7.75, 2.0M NH$_4$SO$_4$, 2.0% peg 400 (Fluka)) and was found from an initial screen using the crystal screen I kit; a commercially available product by Hampton research. A fine screen around these conditions showed that 1.5% peg 400 produced better crystals. Crystals were grown using the hanging drop vapour diffusion method (McPherson A Jr, 1982) at 4°C. 1µl of the crystallisation buffer was mixed with 1µl of the concentrated L1 protein on a siliconised coverslip (Hampton research) and placed over a well on a crystallisation plate (which contained 0.5mls of crystallisation buffer). The coverslip was sealed to the well using vacuum grease to provide an airtight environment. Crystals were obtained after 7-10 days.

3.8 Preparation of crystals for exposure to X-rays.

Crystals are usually sensitive to radiation damage from X-rays. Prolonged exposure would thus lead to deterioration in the diffraction of the crystal. In such situations, if a complete data set has not already been collected, a new crystal would have to be used and thus systematic errors would be introduced. To overcome this problem of crystal lifetime,
cryogenic techniques have been developed which virtually immortalise the crystal. The procedure involves a stream of nitrogen gas (at 100 K) flowing over a crystal that is mounted in the X-ray beam. However, not all crystals freeze favourably in their crystallisation solvent. It is often necessary to add certain amounts of cryoprotectants such as PEG 400 or glycerol to the crystal. With the L1 native and Se-met L1 crystals, it was necessary to add 25% glycerol (Sigma) as a cryoprotectant.

3.9 Data collection of Native L1.

The crystals were initially tested using in-house facilities i.e.; monochromatic CuKα X-rays generated by a Rikagu RU200HB (MSC) rotating anode operating at 100mV and 50kV. Initial testing identified the crystal lattice to be of the primitive hexagonal type. A native dataset was collected to 2.4Å with 1.5° oscillation at Daresbury on station 7.2. The crystal was oriented such that the six-fold symmetry axis present in the crystal was approximately parallel to the rotation axis of the goniometer. This strategy enabled most of the data to be collected from a 30° rotation of the crystal. Subsequently, a high resolution data set to 1.7Å was collected at DESY (Hamburg, Germany) using beamline X11.

3.10 Multi-Wavelength Anomalous dispersion data collection.

The MAD data were collected at the EMBL synchrotron source (DESY) in Hamburg, Germany, on the tuneable beam line, X31. Crystals are very sensitive objects and do not tolerate air travel very well. Therefore, to overcome this problem, the Se-met L1 crystals were mounted in fibre loops, frozen in liquid propane (BOC special gases) and stored in a
liquid nitrogen (CryoServices Ltd) filled dewar for transit. Propane offers the advantage of faster freezing when compared to liquid nitrogen. In addition, as liquid propane solidifies in liquid nitrogen, it allows more handling time for transferring the crystal to the experimental set-up. This means that crystals can be positioned into a stream of nitrogen whilst still in embedded in solid propane.

A by product of the excitation of inner shell electrons of heavy atoms to higher energy states is the emittance of fluorescent radiation as another electron loses energy to fill the vacated orbitals. The fluorescence can be measured and used to determine the absorption edge of a particular heavy atom in a protein. Thus, on a prefrozen crystal mounted on the goniometer in a stream of nitrogen (using a Oxford cryosystem cryostream), fluorescence scans were collected around both the selenium and zinc absorption edges. The scans were then interpreted to obtain values of the $f'$ and $f''$ wavelengths for Zn and Se in this experiment.

The data were collected in the following order, $f''$, $f'$, at the absorbance edges of Se and then Zn. A final dataset was collected at a wavelength below the absorbance edges of both Se and Zn to act as a remote dataset. All data were collected from one crystal to a resolution limit of 3.0 Å with a high degree of redundancy. Data acquisition was controlled using the MAR Research control software, and the data were recorded using 180mm MAR Research image plate.

**3.11 Preparation of enzyme–inhibitor complex and data collection.**

Two strategies were used in attempts to obtain crystals of the complex of L1 and a TFK molecule. The first was to soak the inhibitor into a L1 crystal in a manner analogous to the preparation of derivatives (see chapter 2); the second approach involved setting up crystallisation experiments with L1 complexed to a TFK. Since only a small quantity of
the inhibitor was available, a screening process was not feasible and therefore the experiment was based around the native conditions. Crystallographic data for both types of crystals were measured using both in-house and synchrotron facilities (Daresbury Station 7.2 and 9.6 and Hamburg beamline X31).

In both approaches to obtain crystals of the inhibited complex, a vast excess of inhibitor (5mM) to protein (500μM) was used. The inhibitor was dissolved in the crystallisation buffer (see section 3.7) with the addition of methanol (to give a final concentration of 5% methanol after incubating with enzyme in solution or crystal drop but not in the well). With co-crystallisation, the enzyme was incubated with inhibitor overnight prior to the setting up of the crystallisation experiment. In the soaking experiments, the inhibitor was added to drops with crystals after 7 days of growth. 1μl of inhibitor (concentration - 15mM) in crystallisation buffer was added to the 2μl crystal drops.

3.12 Data analysis procedures.

The MAD experimental data were initially analysed using the CCP4 programs and post priori using SOLVE. With the CCP4 suite of programs, the MAD data were treated as a special case of MIR. The programs used in the analysis are shown in figure 14.0. Detailed background regarding the theoretical basis of the programs is not discussed but the interested reader is referred to a review article (CCP4, 1994) and references therein. Here, a brief description of the overall process is given focusing on aspects of the procedure that were particularly relevant to this project. Below is a description of the initial processing of the diffraction data, which is common to both CCP4 and SOLVE approaches to phase calculation.
A single diffraction image is initially processed using DENZO (Otwinowski and Minor, 1997). Denzo performs a robust autoindexing procedure to determine the most likely crystal symmetry and the unit cell dimensions. With this information, Denzo is able to index the image and carry out refinement of the predicted spot positions. Denzo uses the predicted spot position to place an integration box around each reflection and calculate the intensity of that reflection. The output from DENZO is a file containing the miller indices and the integrated and profile fitted intensity of each reflection for each of the diffraction images making up the dataset. The integrated dataset is then scaled, merged and postrefined using SCALEPACK (Otwinowski and Minor, 1997). The program SCALEPACK calculates an isotropic scale and B-factor for each image. It also calculates and refines the partiality of the reflections prior to adding the partial measurements from adjacent images. This postrefinement procedure refines the unit cell dimensions, crystal orientation and crystal mosaicity. If the refined values for mosaicity and the unit cell dimensions are significantly different from that used to initially integrate the data, the data should be re-integrated. Postrefinement is an important procedure for obtaining high quality data.

SCALEPACK produces statistics which indicate the quality of the data, including the I and σI columns as well as $R_{sym}$. These values are used as a criterion to determine the limit of the resolution of the data. Generally, the data is said to be good to a resolution limit were the $I/\sigma I$ is at least 3 for 65% of the data. The $R_{sym}$ column is also a good indicator of the quality of the data as it describes the differences in the intensities between equivalent (symmetry related) reflections. The overall $R_{sym}$ for a data set should generally not be more than 10%.
Figure 14  Structure determination.

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure

Initial Diffraction Image

Xdisplay

DENZO

SCALEPACK

Format Conversion

CAD

SCALEIT

Experimental Datasets

DENZO

HASSP

Zn Sites

MLPHARE

FFT, MAPMAN

DM ARP

Electron density map

REFMAC

Final Structure
3.13 MIR using CCP4 software.

The five datasets from the MAD experiment (2 Zn, 2 Se and a remote) were integrated and scaled as described above. The output from SCALEPACK was converted to mtz format for use as input to the subsequent programs. The mtz files were combined using CAD (CCP4, 1994) and then each dataset was individually scaled against the remote wavelength dataset using the program SCALEIT (CCP4, 1994). This program puts two (or more) datasets on the same arbitrary scale and analyses the agreement between their respective amplitudes. One of the statistics reported is the $R_{\text{merge}}$ value. A flat $R_{\text{merge}}$ is indicative of isomorphism. The datasets for the $f'$ and $f''$ of both the selenium and zinc edges were treated as being analogous to two derivatives in a MIRas experiment, and the remote dataset was treated as the native. The same remote dataset was used for both the Zn and Se edge data.

With the MAD data there were, in total, 8 anomalous centres to be identified (2 Zn and 6 Se). The strategy used was to calculate an anomalous difference Patterson using the Zn edge data only. The reason for this was the ease of interpretation of a two site Patterson solution. The positions of the anomalous scattering sites can be determined either, by inspection of the Harker sections, or by an automated approach using the program HASSP; the latter method was used in this work.

The Program MLPHARE (Otwinowski, 1991), which is a phase refinement program, takes the sites found by HASSP and calculates phases for the anomalous scattering atom. In this work, the initial phases from MLPHARE, from the two zinc sites, were used to calculate a dispersive difference Fourier. The amplitude coefficients for this were the differences between the remote dataset and the data collected at the $f'$ minimum of the Se edge. The resulting Fourier showed peaks in the electron density that could be attributed to the Se atoms. This procedure enabled us to identify 5 of the possible 6 Se sites. These
sites were then used in MLPHARE, in addition to the zinc sites, to calculate more accurate phases. The quality of the phases calculated by MLPHARE is indicated by the figure of merit. The phases are calculated as a probability distribution and the figure of merit is an approximation to the cosine of the phase error. Thus a FOM value of 1.0 indicates perfect phases. However, an overall FOM value of 0.5 is normally sufficient to give an interpretable electron density map.

The phases from MLPHARE were used to calculate the electron density maps using FFT (CCP4, 1994). A graphics format file was then created, covering just the molecule, using EXTEND and MAPMAN (CCP4, 1994). Several types of electron density maps may be produced. Once the initial model is available the most useful map is the $2F_o - F_c$ map. Calculated structure factors ($F_c$) are determined from the back transform of the coordinates of the model built into the initial electron density map. The observed structure factors ($F_o$) are those obtained from the diffraction data. A $2F_o - F_c$ map highlights the errors in the model and indicates where it must be adjusted in order to agree more closely with the electron density map. In this way a more accurate model can be produced. Electron density maps can be improved by phase extension, density modification and automated refinement; these methods are described later.

3.14 SOLVE

The SOLVE approach is an automated method of phase determination (Terwillinger T C, 1994a; 1994b; Terwillinger et al., 1983; 1987; 1997). SOLVE identifies potential heavy atom sites and carries out refinement of these sites with subsequent phase calculation. Essentially, it is a black box procedure, i.e. SOLVE reads the hkl file from SCALEPACK and produces an electron density map at the other end.
SOLVE uses the scaled data from SCALEPACK and calculates difference Pattersons to determine potential anomalous scatterers or heavy atoms. Each potential site is tested and given a score to reflect the chance that the site is correct. The scoring system is based on the following criteria.

- Agreement with Patterson.
- "Free" difference Fourier.
- Distinguishable solvent and protein regions in the electron density map.
- The figure of merit of the calculated phases.

In addition to these criteria, SOLVE analyses the relative scores of each criterion for a particular solution and gives a higher score to those that are uniformly good. The maps produced by SOLVE can then be used to build a model. The primary advantage of SOLVE is its automation and the fact that proteins with many heavy atom sites can be identified.

3.15 Density Modification and Phase extension.

The initial electron density map from MLPHARE was further improved and extended to higher resolution using density modification techniques including solvent flattening (Wang B C, 1985) and histogram matching (Cowtan K D, 1994) using the program DM (CCP4, 1994).

Solvent flattening

The electron density map of the unit cell contains both protein and solvent regions. If the electron density map has been generated using very good phases (FOM~0.9), the density
distribution of the protein regions will be varied whilst the solvent regions will be relatively flat. These features in a map are highly related to the quality of the original phases and thus poorer phases can in turn be improved by applying the constraint of solvent flatness. The program DM uses the algorithm of Wang to identify the protein and solvent regions. DM then assigns a mask to cover the region, which it assumes to be protein and then imposes the constraint that the rest of the cell must have the same (i.e. flat) density. The back transform of this modified map will then give a new set of phases. This procedure can often lead to improvement in the phases by about 20-30°. The limitation of this method is that one must have relatively good phases (FOM ~ 0.5) to start with and the power of the procedure is approximately proportional to the solvent content.

Histogram matching.

The electron density distributions in accurately determined protein structures are very similar (for a given resolution) and therefore it can be summarised in the form of the ideal histogram. For a map that has been calculated from very good phases (FOM~0.9), this distribution will closely follow the ideal histogram. However, the density histogram of maps generated using poorer phases (FOM~0.5) will deviate from the ideal distribution. DM calculates a histogram of the distribution of the electron density values from an empirical Fourier map and determines scale factors to apply to the observed histogram to improve the agreement with the ideal histogram. As with solvent flattening, the improvement in the phases is relatively small.

3.16 Refinement.

An initial model was built into the electron density map calculated from the DM modified MLPHARE phases. The program PROTIN (CCP4, 1994) was used to create an input
file that contained the ideal and the observed atomic parameters such as bond lengths and bond angles. PROTIN calculates the restraints that will be applied during refinement based on differences between the ideal and observed atomic parameters. The ideal parameters are taken from a dictionary of geometry, whereas the observed atomic parameters are those calculated from the initial model (i.e. x,y,z of atom, the temperature (B) factor of the atom, and the bond lengths and angles from the initial model). Phase extension was carried out by refining the initial model against the 2.4Å native data using REFMAC (Murshudov et al., 1997). The lower resolution native was used in the phase extension because it was found to be more isomorphous with the MAD data than the 1.7Å native. Automated refinement was coupled with rebuilding (using O) to correct gross errors in the model and to interpret previously poorly ordered regions.

Intermittent with REFMAC, the program ARP (Lamzin and Wilson, 1993) was used to add water molecules. ARP identifies unaccounted regions of density and adds water molecules in these positions. By this means, ARP is able to improve and extend the current phase estimates. The 2.4Å model was then refined (REFMAC) against the high resolution (1.7Å) native dataset but using the DM-modified MLPHARE phases to 3Å as additional restraints. The inclusion of the additional phase information enabled the REFMAC/ARP refinement to converge accurately from the starting model. This refinement did not work if the phase restraints were not included.

3.17 Computing and Additional software.

The Ramachandran distribution of the refined structure was calculated using the program PROCHECK (CCP4, 1994). Cartoons of the structure and the active site were produced using MOLSCRIPT (Kraulis P J, 1991), QUANTA (Molecular Simulations Inc.), GRASP (Nicholls et al., 1991) and RASMOL (Glaxo-Wellcome). Chemical structures
and reaction schematics were drawn using CHEMDRAW (Cambridge Scientific Computing Inc.). All CCP4 programs were run on X-terminals on a UNIX platform. This report was written using Microsoft Office 98 and Endnote.
Chapter 4.

Results:

Protein purification, characterisation and crystallisation.
4.1 Purification.

The protein purification procedure used was fast and relatively efficient enabling protein, suitable for crystallisation purposes, to be obtained on the tens of mg scale. The yield, which is less than 50%, is adequate for crystallisation purposes. The progress of the purification was assessed at each stage and the results are presented in Table 4.

Table 4  Purification of native L1.

<table>
<thead>
<tr>
<th>Purification step</th>
<th>Protein conc. (mg/ml)</th>
<th>Volume (mg)</th>
<th>Total protein (mg)</th>
<th>Activity (total EU)</th>
<th>% Recovery</th>
<th>Specific activity (EU/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>7.5</td>
<td>150</td>
<td>1125</td>
<td>4.8 x 10^7</td>
<td>100</td>
<td>4.3 x 10^6</td>
</tr>
<tr>
<td>(NH₄)₂SO₄ cut</td>
<td>5.4</td>
<td>130</td>
<td>702</td>
<td>3.5 x 10^7</td>
<td>73</td>
<td>5.1 x 10^6</td>
</tr>
<tr>
<td>Q-flow through</td>
<td>0.397</td>
<td>130</td>
<td>51.61</td>
<td>2.4 x 10^7</td>
<td>50</td>
<td>4.7 x 10^6</td>
</tr>
<tr>
<td>S-seph fractions</td>
<td>0.65</td>
<td>27</td>
<td>17.55</td>
<td>2.2 x 10^7</td>
<td>46</td>
<td>1.3 x 10^6</td>
</tr>
<tr>
<td>Gel filtration</td>
<td>1.6 x 10^10</td>
<td>10</td>
<td>16</td>
<td>2.0 x 10^7</td>
<td>42</td>
<td>1.25 x 10^6</td>
</tr>
</tbody>
</table>

EU (enzyme units) is defined as the amount enzyme required to catalyse 1μmole of substrate per minute.

Purification of Se-met incorporated L1 proceeded similarly to the native protein except that the yield was considerably lower (ca. 20% relative to the native prep). Figure 15 shows a SDS PAGE gel of the peak fractions from the gel filtration step.
Figure 15  SDS PAGE of the peak fractions from gel filtration.

L = Material prior to gel filtration
GF = Peak LI fractions after gel filtration
4.2 Dynamic light scattering.

The purified protein was shown to be monodisperse by dynamic light scattering (i.e., the protein species was uniform in solution with no aggregation). The procedure also provided an estimate of the molecular mass of L1; the data is shown in table 5.

Table 5 Mean values from dynamic light scattering analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>purified L1 (1mg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation diffusion coefficient; $D_T$</td>
<td>478</td>
</tr>
<tr>
<td>Hydrodynamic radius; $R_h$ (nm)</td>
<td>4.2</td>
</tr>
<tr>
<td>Polydispersity (%)</td>
<td>7.1</td>
</tr>
<tr>
<td>Estimated molecular weight (kd)</td>
<td>95</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>17.5</td>
</tr>
<tr>
<td>Base line</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

*D$_T$ is expressed in units of $10^{-11}$ m$^2$/seconds.

The translation diffusion constant describes the motion of particles in solution. This coefficient ($D_T$) is converted to the hydrodynamic radius $R_h$, of the particles in solution using the Stokes-Einstein equation $R_h = K_b T / 6 \pi \eta D_T$, where $K_b$ is the Boltzmann constant ($1.38 \times 10^{-23}$ KJ), $T$ is the absolute temperature in Kelvin and $\eta$ is the solvent viscosity (poise). The hydrodynamic radius given in the output represents the median size present within the sample. The polydispersity value is expressed as a percentage and indicates the standard deviation of the spread of particle sizes about a reported average radius. A value of less than 15 percent indicates that the standard deviation of particle size is negligible and hence, the solution is largely monodisperse. An estimate of the molecular
weight of the species in the sample is calculated using the hydrodynamic radius. This is
done using an empirically derived relationship between values of $R_h$ and molecular weight
(MW) values for a number of well characterised globular proteins in buffered aqueous
solutions. The determination of MW by this method is therefore dependent on the shape
of the molecule and hence not likely to be accurate.

4.3 Determination of Molecular Mass.

The molecular mass of L1 was determined using electrospray mass spectroscopy
(ESMS). The results showed that the subunit mass was 27803 Da. The ESMS trace
(figure 16) also showed that the prepared L1 was very pure as only one species was
detected in the analysis. The results from gel filtration and lights scattering indicated that
L1 is polymeric in solution. This was confirmed by sedimentation equilibrium studies
with L1, which showed conclusively that L1 is a tetramer with a molecular mass of
114,838 Da. (see figure 17).
The mass of L1 was determined to be $28703\pm1.28$ Da.
The figure shows the radial distribution of the sample at equilibrium in the ultracentrifugation cell. This distribution is a function of the relative molecular mass of the species present and therefore can be used to determine molecular mass. The continuous line is the best fit to the data with \( M_r \) 114, 838 Da (\( \nu =0.732, \rho =1.0084 \)). The top section shows how the residuals are distributed about the non-linear fit. This is random and is indicative of the uniformity of the species in the sample. The data at equilibrium were analysed using the Origin scientific data analysis software.
4.4 Crystals of L1.

The crystals of native and Se-incorporated L1 were similar in morphology, crystal lattice and unit cell dimension. (Figure 18 and table 6).

**Table 6** Crystal size, lattice and unit cell dimensions.

<table>
<thead>
<tr>
<th>Crystal dimension (microns; $10^{-6}$ m)</th>
<th>70 x 70 x 440 (Maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal lattice, space group.</td>
<td>Primitive hexagonal, P6$_3$22.</td>
</tr>
</tbody>
</table>
| Unit cell dimensions ($\AA$; $^\circ$)    | a=105.6; b=105.6; c=97.9; $\alpha$=90.0;  
|                                           | $\beta$=90.0; $\gamma$=120.0 |

**Figure 18** L1 crystals
The crystals of L1 were tested for biological activity using the synthetic cephalosporin substrate, nitrocefin (O’Callaghan et al., 1972). Nitrocefin has an absorbance maximum at 386nm (yellow in the visible spectrum). However, on hydrolysis of the β-lactam ring, there is a shift in the absorbance max to 482nm, (red in the visible spectrum) (figure 19). On the addition of nitrocefin, the L1 crystals became red. The red colour was located only in the crystal itself and not in the surrounding liquor, indicating therefore that the hydrolysis of nitrocefin is due to the action of L1 within, or at the surface of, the crystal.

**Figure 19**  Hydrolysis of nitrocefin.

β-lactamase facilitated hydrolysis of nitrocefin results in a shift in the peak absorbance from 386nm (yellow) to 482nm (red).
Chapter 5.

Results

Crystallographic analysis
5.1 Native data

A native dataset to 2.4Å were collected at Daresbury, station 7.2; the data statistics are shown in table 7. The data were collected over a resolution range of 20-2.4Å and was of good quality throughout, as indicated by greater than 65% of the data in the highest resolution bin having intensity values better than 3 sigma. A higher resolution native was later collected at Hamburg (station X11) as described in chapter 3 (table 8).

Table 7 Native dataset to 2.4 Å.

<table>
<thead>
<tr>
<th>Max. resolution (Å) of shells</th>
<th>No. of unique reflections</th>
<th>Reflections I/sigI &gt; 3 (%)b</th>
<th>Completeness s (%)</th>
<th>Rsym (%)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.78</td>
<td>1759</td>
<td>96.9</td>
<td>98.5</td>
<td>5.5</td>
</tr>
<tr>
<td>3.80</td>
<td>1667</td>
<td>95.5</td>
<td>99.8</td>
<td>6.4</td>
</tr>
<tr>
<td>3.32</td>
<td>1637</td>
<td>93.5</td>
<td>100.0</td>
<td>7.5</td>
</tr>
<tr>
<td>3.02</td>
<td>1625</td>
<td>87.4</td>
<td>99.8</td>
<td>9.0</td>
</tr>
<tr>
<td>2.81</td>
<td>1611</td>
<td>82.9</td>
<td>99.8</td>
<td>11.9</td>
</tr>
<tr>
<td>2.64</td>
<td>1591</td>
<td>78.1</td>
<td>99.4</td>
<td>14.7</td>
</tr>
<tr>
<td>2.51</td>
<td>1594</td>
<td>72.9</td>
<td>99.6</td>
<td>18.8</td>
</tr>
<tr>
<td>2.40</td>
<td>1575</td>
<td>67.4</td>
<td>99.1</td>
<td>22.7</td>
</tr>
<tr>
<td>Totals</td>
<td>13059</td>
<td>84.6</td>
<td>99.5</td>
<td>6.9</td>
</tr>
</tbody>
</table>

The values given in the columns are bin values corresponding to the resolution shells. The totals are given in the bottom row. I/sigI>3, shows the percentage of reflection whose observed intensities (I) are at least three times greater than the standard deviation of the average intensity. Rsym describes the quality of the data, defined as \( \Sigma |I - \langle I \rangle| / \Sigma |I| \), where \( I \) is the observed intensity and \( \langle I \rangle \) is the average intensity. For the native dataset, all the symmetry equivalents are included.
Table 8  Native dataset to 1.7 Å.

<table>
<thead>
<tr>
<th>Max. resolution (Å)</th>
<th>*No. of reflections</th>
<th>*Reflections I/sigI&gt;3 (%)</th>
<th>*Completeness (%)</th>
<th>*R_sym (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.89</td>
<td>2862</td>
<td>98.6</td>
<td>90.8</td>
<td>0.056</td>
</tr>
<tr>
<td>3.10</td>
<td>2914</td>
<td>97.8</td>
<td>97.5</td>
<td>0.049</td>
</tr>
<tr>
<td>2.71</td>
<td>2929</td>
<td>95.7</td>
<td>99.8</td>
<td>0.051</td>
</tr>
<tr>
<td>2.46</td>
<td>2913</td>
<td>92.9</td>
<td>99.9</td>
<td>0.058</td>
</tr>
<tr>
<td>2.29</td>
<td>2902</td>
<td>91.8</td>
<td>99.9</td>
<td>0.058</td>
</tr>
<tr>
<td>2.15</td>
<td>2890</td>
<td>88.5</td>
<td>99.9</td>
<td>0.063</td>
</tr>
<tr>
<td>2.05</td>
<td>2903</td>
<td>85.7</td>
<td>100.0</td>
<td>0.074</td>
</tr>
<tr>
<td>1.96</td>
<td>2863</td>
<td>81.0</td>
<td>100.0</td>
<td>0.094</td>
</tr>
<tr>
<td>1.88</td>
<td>2854</td>
<td>75.4</td>
<td>100.0</td>
<td>0.127</td>
</tr>
<tr>
<td>1.82</td>
<td>2864</td>
<td>69.2</td>
<td>100.0</td>
<td>0.135</td>
</tr>
<tr>
<td>1.76</td>
<td>2874</td>
<td>61.8</td>
<td>99.9</td>
<td>0.156</td>
</tr>
<tr>
<td>1.71</td>
<td>2808</td>
<td>58.3</td>
<td>99.8</td>
<td>0.196</td>
</tr>
<tr>
<td>Totals</td>
<td>34576</td>
<td>83.3</td>
<td>98.8</td>
<td>0.054</td>
</tr>
</tbody>
</table>

*The values given in the columns are bin values corresponding to the resolution shells. The totals are given in the bottom row. *I/sigI>3, shows the percentage of reflections whose observed intensities (I) are atleast three times greater than the standard deviation of the average intensity (sigI). *R_sym describes the quality of the data and is defined as Σij |Ij| - Ij | Σ[Ij], where Ij is the observed intensity and [I] is the average intensity.

The high resolution data were collected by making 2 passes, the first to collect the high resolution terms and the second to collect the intense low resolution terms. This strategy was necessary because the exposure times needed to collect good high resolution data resulted in some of the low resolution data being overloaded. Due to time constraints, the low resolution terms were not completely collected. The quality of the data is nevertheless very good to 1.82Å and reasonable beyond that. After analysing the native datasets, a space group defining all the symmetry operators present within one unit cell could be assigned. In our case, there were two enantiomorphic possibilities, i.e.; P6422 or P6222. This ambiguity is due to both of these space groups having the same pattern of
systematically absent reflections. The real difference between these two space groups is the handedness of the screw axis present, which cannot be distinguished from the raw data alone. This ambiguity was resolved as described later.

### 5.2 MAD data

The characteristic X-ray fluorescence of both the Zn and Se atoms, within the crystal, were measured (figure 20 and 21) to obtain the wavelengths at which the anomalous and dispersive signals are at their respective maxima. The experimentally determined wavelengths for the MAD datasets are shown in table 9 together with a summary of the data collection statistics.

#### Table 9 Data processing statistics for the MAD datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>( \lambda ) (Å)</th>
<th>Max. resolution (Å)</th>
<th>Data completeness (%)</th>
<th>Data redundancy</th>
<th>( R_{\text{sym}}(%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote</td>
<td>0.9200</td>
<td>2.85</td>
<td>98.6</td>
<td>4.15</td>
<td>4.9</td>
</tr>
<tr>
<td>Se ( f' )</td>
<td>0.9797</td>
<td>3.00</td>
<td>99.1</td>
<td>3.2 (5.7)</td>
<td>3.9 (6.0)</td>
</tr>
<tr>
<td>Se ( f'' )</td>
<td>0.9801</td>
<td>3.00</td>
<td>96.9</td>
<td>3.2 (5.7)</td>
<td>3.9 (4.8)</td>
</tr>
<tr>
<td>Zn ( f' )</td>
<td>1.2821</td>
<td>3.10</td>
<td>98.7</td>
<td>2.5 (4.5)</td>
<td>4.3 (5.3)</td>
</tr>
<tr>
<td>Zn ( f'' )</td>
<td>1.2835</td>
<td>3.10</td>
<td>99.4</td>
<td>3.3 (5.6)</td>
<td>4.3 (4.5)</td>
</tr>
</tbody>
</table>

\( R_{\text{sym}} \) describes the quality of the data and is defined as \( \Sigma \left| I - \langle I \rangle \right| / \Sigma |I| \), where \( I \) is the observed intensity and \( \langle I \rangle \) is the average intensity. For the native dataset, all the symmetry equivalents are included. The values shown in parenthesis, in the data redundancy and \( R_{\text{sym}} \) columns, represent the values obtained including the merged bijvoet pairs.
Figure 20  Fluorescence scan of Zn close to the absorption edge.

![Fluorescence scan of Zn close to the absorption edge.](image)

Figure 21  Fluorescence scan of Se close to the absorption edge.

![Fluorescence scan of Se close to the absorption edge.](image)
5.3 Phasing

Initial phases were calculated using CCP4 procedures as described in chapter 3; the results are presented below. Alternatively, the phases could have been calculated using the automated approach of SOLVE. Thus post priori to the structure determination, HASSP (CCP4 approach) and SOLVE were compared with respect to their ability to determine the positions of the anomalous scattering atoms. The results of this investigation are presented in section 5.4.

The zinc positions in L1 were determined by calculating an anomalous difference Patterson using the zinc \( f' \) data. The Harker section \( z=1/3 \) (there are two Harker sections in this space group, \( z=1/3 \) and \( x=0 \)) appeared to be interpretable (figure 22a). Two sets of peaks were observed that could have accounted for the zinc ions. The task of determining the anomalous scattering sites was carried out using the automated approach of HASSP. HASSP found two sites which, when put into MLPHARE, produced phases with a figure of merit of 0.59. The Harker section also contained cross vectors between zinc ions. These cross vectors appear on the Harker section because the sites are quite close together (figure 22b).

Subsequently, a phased dispersive difference Fourier (using the phases derived from the zinscs) revealed the positions of 5 of the 6 Se sites. These sites were then incorporated into the phase calculation with MLPHARE. The refined occupancies of the Zn and Se sites are presented in table 10. The resulting phases were of high quality, with a mean figure of merit (FOM) of 0.79.
(a) The Harker section $z=1/3$ from Pattersons calculated using the anomalous (above) and dispersive (acetate 1) differences from the Zn edge data. There are two sets of unique peaks. As the two zinc ions are close together, cross vectors between the two zincs ions are also present on the Harker section.
Figure 2.2  Harker section $z=1/3$.

(a) The Harker section $z=1/3$ from Pattersons calculated using the anomalous (above) and dispersive (acetate 1) differences from the Zn edge data. There are two sets of unique peaks. As the two zinc ions are close together, cross vectors between the two zincs ions are also present on the Harker section.
The Harker section $z=1/3$ from the dispersive Patterson from the Zn edge data. The Harker section ($z=1/3$) of a Patterson synthesis from just the calculated zinc sites is overlaid on top (Acetate 2). There are 2 sets of unique peaks. The peaks corresponding to Harker vectors are denoted H whereas, peaks corresponding to the cross peaks are denoted C.
The Harker section \( z = 1/3 \) from the dispersive Patterson from the Zn edge data. The Harker section \( (z = 1/3) \) of a Patterson synthesis from just the calculated zinc sites is overlaid on top (Acetate 2). There are 2 sets of unique peaks. The peaks corresponding to Harker vectors are demoted H whereas, peaks corresponding to the cross peaks are denoted C.
Table 10  MLPHARE refined anomalous scattering sites.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Site no.</th>
<th>Real occupancy</th>
<th>Anomalous occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se f'</td>
<td>1</td>
<td>0.163</td>
<td>0.755</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.176</td>
<td>1.319</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.162</td>
<td>1.347</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.154</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.118</td>
<td>0.966</td>
</tr>
<tr>
<td>Se f''</td>
<td>1</td>
<td>0.156</td>
<td>1.679</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.141</td>
<td>2.902</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.107</td>
<td>2.856</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.129</td>
<td>2.101</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.780</td>
<td>2.037</td>
</tr>
<tr>
<td>Zn f'</td>
<td>1</td>
<td>0.239</td>
<td>2.606</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.248</td>
<td>2.615</td>
</tr>
<tr>
<td>Zn f''</td>
<td>1</td>
<td>0.146</td>
<td>4.783</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.178</td>
<td>4.390</td>
</tr>
</tbody>
</table>

*The occupancy values are on an arbitrary scale.
Table 11 Phasing statistics.

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>14-10</th>
<th>10-7.5</th>
<th>7.5-6</th>
<th>6-5</th>
<th>5-4.3</th>
<th>4.3-3.8</th>
<th>3.8-3.3</th>
<th>3.3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOM</td>
<td>0.9208</td>
<td>0.9182</td>
<td>0.9213</td>
<td>0.8848</td>
<td>0.8460</td>
<td>0.8343</td>
<td>0.7853</td>
<td>0.6611</td>
</tr>
<tr>
<td>Mean FOM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.79</td>
</tr>
</tbody>
</table>

An initial electron density map was calculated to 2.85Å using these empirical phases. The space group ambiguity (P6$_2$22 versus P6$_4$22) was resolved by inspection of the two electron density maps, which showed unambiguously the space group to be P6$_4$22. As shown in Table 12, the final model, containing 267 residues (2-268) with 293 water molecules, was refined to a crystallographic R-factor of 18.3%; the free R-factor was 23.3%. A residue was missing at both the N and C terminus of the protein. The stereochemistry and correctness of the model was further checked by generating a Ramachandran plot (Ramachandran et al., 1963) (figure 23) of the final model. 99% of the residues fell within the allowed regions. Most of the outliers were glycines with the exception of Asn53. The portion of the final electron density map at 1.7Å is shown in figure 24.

Table 12 Refinement. (resolution range 2 0 -1.7Å)

<table>
<thead>
<tr>
<th>$R_{(cryst)}$ (%)$^a$</th>
<th>18.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{(free)}$ (%)$^b$</td>
<td>23.3</td>
</tr>
<tr>
<td>R.M.S bond lengths (Å)</td>
<td>0.011</td>
</tr>
<tr>
<td>R.M.S bond angles (degrees)</td>
<td>1.6</td>
</tr>
<tr>
<td>Average B</td>
<td>24.1Å$^2$</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>2003</td>
</tr>
<tr>
<td>No. water molecules</td>
<td>293</td>
</tr>
</tbody>
</table>

$^a$R(cryst) = $\Sigma_{hk l} | F_p-F_p(calc)|^2 / $\Sigma_{hk l} F_p$. $^b$As R(cryst) but calculated on the 5% of data excluded from the refinement.
Figure 23  Ramachandran plot.

Plot statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions [A,B,L]</td>
<td>202</td>
<td>90.6%</td>
</tr>
<tr>
<td>Residues in additional allowed regions [a,b,l,p]</td>
<td>20</td>
<td>9.0%</td>
</tr>
<tr>
<td>Residues in generously allowed regions [-a,-b,-l,-p]</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td>1</td>
<td>0.4%</td>
</tr>
<tr>
<td>Number of non-glycine and non-proline residues</td>
<td>223</td>
<td>100.0%</td>
</tr>
<tr>
<td>Number of end-residues (excl. Gly and Pro)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Number of glycine residues (shown as triangles)</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Number of proline residues</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Total number of residues</td>
<td>268</td>
<td></td>
</tr>
</tbody>
</table>
Figure 24  Electron density map at 1.7 Å.

The figure shows the refined atomic model of a portion of the protein together with the final $2F_o - F_c$ electron density map. The map is contoured at $1\sigma$. 
5.4 **Comparison of HASSP and SOLVE in phasing.**

The determination of zinc positions from the zinc edge data was easily accomplished by both HASSP and SOLVE. The figure of merit of the phases calculated (using MLPHARE) from these sites are similar because the sites are essentially the same (table 13). The Se sites were identified by calculation of a dispersive difference Fourier using the empirical zinc phases and by SOLVE. Both methods returned six sites, although, the sixth site from the dispersive Fourier was initially overlooked (table 13). It was also investigated whether SOLVE could find all eight anomalous scattering sites given both the Se and zinc edge data. Although eight sites were found (corresponding to 2 zincs and 6 seleniums), but these sites did not correspond to the correct sites and produced a lower FOM (0.6). These results are discussed in chapter 7.

**Table 13** Comparison of HASSP and SOLVE in the determination of anomalous scattering sites.

<table>
<thead>
<tr>
<th>Atom type(s)</th>
<th>No of sites</th>
<th>Mean figure of merit$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOLVE</td>
<td>HASSP</td>
</tr>
<tr>
<td>Zn</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Se</td>
<td>6</td>
<td>$^b$ 5</td>
</tr>
<tr>
<td>Zn and Se</td>
<td>8 ($2\text{zn, 6 se}$)</td>
<td>8$^b$ ($2\text{zn, 5se}$)</td>
</tr>
</tbody>
</table>

$^a$Figure of merit is from MLPHARE for sites found by both methods. $^b$5 selenium sites were located by calculating a phased difference Fourier using the phases derived from the zinc data. $^c$The figure of merit from MLPHARE using zinc and selenium sites determined, individually as opposed to together, by SOLVE is 0.82. The correct sites were not found when both sets of data (Zn and Se) were included together in SOLVE. $^d$The figure of merit here is calculated using original zinc sites and the selenium sites from the dispersive Fourier. $^e$The sixth selenium position was initially overlooked in the dispersive Fourier. Thus, Post priori to structure determination, the sixth selenium was added to calculate new phases. The figure of merit of these new phases was calculated to be 0.902.
The figure of merit of a set of phases is approximately related to the cosine of the error of those phases. A better indicator of the quality of calculated phases can be obtained by calculating the error between that of the starting phases and the phases of the refined model, which are taken to be accurate within the limits of the error of the data (at 1.7Å in this case), post priori to structure determination. The two sets of phases, that is the model phases and the MAD phases at 3.0Å, were compared using PHISTATS (CCP4, 1994) and the results of the analysis are shown in table 14.

**Table 14** The phase error of the initial empirical phases calculated using HASSP/MLPHARE relative to the final refined model.

<table>
<thead>
<tr>
<th>Heavy atoms</th>
<th>Error of initial phases (degrees)</th>
<th>Error of the DM modified phases (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HASSP/DF</td>
<td>HASSP/DF</td>
</tr>
<tr>
<td>2 Zn</td>
<td>47</td>
<td>42</td>
</tr>
<tr>
<td>6 Se</td>
<td>41</td>
<td>36</td>
</tr>
<tr>
<td>2 Zn + 6 se</td>
<td>34</td>
<td>27</td>
</tr>
</tbody>
</table>

The table shows the errors of the starting phases calculated from the MAD data using just the zinc, selenium and both sets of heavy atoms relative to the final model refined against the 1.7Å native data. The initial phases were calculated from sites found by HASSP and phased difference Fourier (DF) using MLPHARE. The model structure factors were calculated using SFALL (CCP4, 1994).
Chapter 6.

Results

The Structure of *Stenotrophomonas maltophilia* L1 and Mechanistic studies
6.1 The structure of L1

The open reading frame of the metallo β-lactamase, L1, codes for a polypeptide chain of 290 residues, of which the first 21 residues constitute a signal peptide. The remaining 269 residues agree with the molecular mass determined from the ESMS data. The atomic model of the structure was hence numbered from the end of the signal sequence to the end of the polypeptide chain (i.e. 1-269). One residue at each end of the polypeptide could not be built as there was no corresponding density in the electron density maps.

L1 is consists of four identical subunits (figure 25) which interact to form a 222 tetramer. The subunits exhibit a αβαβα fold that is divided into 2 roughly equivalent αβ domains (N and C terminal domains). Each domain is comprised of an antiparallel β sheet and two alpha helices. The two helices are parallel in the N and antiparallel in the C-terminal domain. There is an additional helix, named α3, which is positioned in between the two domains and is adjacent to the active site. In addition to the interdomain helix, a strand, named β7, is also involved in connecting the two domains. There are two cysteine residues present within the molecule (Cys 218 and Cys 246) which form an intramolecular disulphide bridge.

The topology of L1 is β1β2β3β4α1β5α2β6 (α3β7) β8β9β10β11α4β12α5 (figure 26). Three short helical turns (310 helix) are also observed prior to β1, α1 and α5. The loop regions between α3 and β7 and β12 and α5 are elongated (with respect to similar regions in the other known structures) and accommodate an additional 37 residues (compared with ccrA) or 42 residues (compared with BCII) in the mature polypeptide.
The electrostatic potential surface of L1 was calculated using GRASP. This showed positively and negatively charged regions around the active site. The positive charge at the active site is primarily due to the two Zn\(^{2+}\) ions (figure 27).

**Figure 25**  Subunit structure of L1

The active site zicons are depicted as green spheres. Surrounding the active site are two loop regions called Loop A and Loop B. The intramolecular disulphide bridge present in L1 is drawn in yellow.
The secondary structure of L1 is shown above. The helices are coloured in yellow and the β-strands in pink.
Figure 27  Surface electrostatic potential of L1.

Surface electrostatic potential of L1. Positive potential is shown in blue and negative potential is shown in red.
6.2 The structure of the active site.

The active site of L1 lies within a groove formed between the β sheets of the two domains and has associated with it the loop regions that connect β₃-α₂, β₉-β₁₀, and α₄-β₁₂. The active site contains two zincs, which are approximately 3.5 Å apart. The first zinc (named Zn1) is co-ordinated by three histidines (His 84, His 86, and His 160) and a water molecule, named Wat1, in an approximately tetrahedral arrangement. The second zinc (named Zn2) is co-ordinated by two histidines (His 89, His 225), an aspartate (Asp 88) and two water molecules, Wat1 and Wat2. The geometry is approximately trigonal bipyramidal. Wat1 is thus shared by the two zincs (figure 28).

As well as the aforementioned loops, the active site is flanked on two sides by two further extended loops. The first of these is situated above the N-terminal interface of the active site region and connects the interdomain helix (α₃) to the interdomain strand (β₁₀) and will be referred to as loop A (see figure 25). The second flanking loop lies above the C-terminal interface of the active site region and connects β₁₁-α₄ which will be referred to as loop B (see figure 25). Loop B contains a conserved glycine (Gly195) and a largely conserved asparagine (Asn196) residue that has until recently, been regarded as being invariant amongst the metallo β-lactamases (see chapter 1); the role of these residues remains unclear.

The geometry of the active site residues is maintained by an extensive hydrogen-bonding network (figure 28, table 15). This is necessary because for a histidine to act as a zinc ligand, it must be in the correct tautomeric form; that is, the nitrogen on the imidazole ring which interacts with the zinc ion must be unprotonated. In L1, each histidine is maintained in the correct tautomeric form via a hydrogen bond between its protonated nitrogen with an aspartate or a serine residue. In some cases, this bond is mediated by a water bridge.
Table 15  Zinc co-ordination.

<table>
<thead>
<tr>
<th>Cation</th>
<th>Distance (Å)</th>
<th>Ligand</th>
<th>Hydrogen bonding network. Ligand - Orienting group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zn1</td>
<td>2.03</td>
<td>His84 NE2</td>
<td>His84 ND1 - Asp184 OD1</td>
</tr>
<tr>
<td>Zn 1</td>
<td>2.11</td>
<td>His86 ND1</td>
<td>His86 NE2 - Asp121 OD2</td>
</tr>
<tr>
<td>Zn 1</td>
<td>2.05</td>
<td>His160 NE2</td>
<td>His160 ND1 - Ser187 O</td>
</tr>
<tr>
<td>Zn 1</td>
<td>1.86</td>
<td>Wat1</td>
<td></td>
</tr>
<tr>
<td>Zn 2</td>
<td>2.07</td>
<td>Asp88 OD2</td>
<td></td>
</tr>
<tr>
<td>Zn 2</td>
<td>2.02</td>
<td>His89 NE2</td>
<td>His89 ND1 - Wat23-Asp53 OD2</td>
</tr>
<tr>
<td>Zn 2</td>
<td>2.07</td>
<td>His225 NE2</td>
<td>His225 ND1 - Wat22-Asp37 O</td>
</tr>
<tr>
<td>Zn 2</td>
<td>2.07</td>
<td>Wat1</td>
<td></td>
</tr>
<tr>
<td>Zn 2</td>
<td>2.40</td>
<td>Wat2</td>
<td>Wat2 - Ser185 OG</td>
</tr>
</tbody>
</table>

108
The above figure shows the configuration of the active site of L1. The grey spheres represent the zinc ions. Zn1 is co-ordinated to His84, His86, His160 and Wat1, and Zn2 is co-ordinated to Asp88, His89, His225, Wat1 and Wat2. The active site is embedded in an extensive network of hydrogen bonds.
6.3 The L1 tetramer

Inspection of the tetramer reveals that each subunit is in contact with each of its three counterparts (figure 29). Tetramerisation therefore occurs through three distinct interfaces. These interfaces were characterised by solvent accessibility calculations and the nature of the interactions are described below.

Interface 1
The interface between molecules A and B is largely hydrophobic. An interaction between Met140 from molecule A with Pro162 in molecule B creates a hydrophobic pocket that accommodates the side chains of Tyr199 and Pro200 of molecule B. Furthermore, Val112 from each molecule packs against a hydrophobic patch composed of Val112, Ile138 and Ala108 on the opposing subunit (see figure 30).

Interface 2
The interaction between subunits A–C is extensive and occurs via the extended N terminal coil region that crosses over with the equivalent structure, on the opposing subunit, to form a W shape (figure 31). The interactions are mainly hydrophobic but also involves some hydrogen bonding (table 16). With respect to the hydrophobic interactions, two leucines are particularly important; Leu5 is situated within a hydrophobic pocket created by Ala10, Tyr11 and Ala15 at the C terminal end of this loop in the opposing subunit; Leu 8 packs into a hydrophobic cavity generated by the sidechains of Met56, Pro57, Met59, Gln52 and His62 at the N-terminus of the first helix (α1) in the opposing subunit. As can be seen in figure 31, the remaining residues of the N-terminal extended loop pack against the equivalent structure on the opposing subunit to bury additional hydrophobic surfaces.
**Interface 3**

The third interface involving Molecules A and D, completes the necessary contacts for tetramerisation (figure 32). This interface appears to be the least of the three and primarily involves a bifurcated hydrogen bond between a lysine (Lys103) and the main chain carbonyls of two alanines (Ala97 and 102) on the opposing subunit (table 16).

The largest contribution in terms of buried surface area is made by interface 2 which accounts for approximately 1245 Å² of surface area per molecule. Interface 1 and 2, which are less extensive, account for 688 Å² and 332 Å² of surface area per molecule respectively.

**Table 16**   Hydrogen bonds involved in the formation of the tetramer.

<table>
<thead>
<tr>
<th>Molecule A</th>
<th>Distance (Å)</th>
<th>Molecule C</th>
<th>Molecule D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu 18 O</td>
<td>2.98</td>
<td>Leu 5 NH</td>
<td>Lys 103 NZH</td>
</tr>
<tr>
<td>Val 13 NH</td>
<td>2.84</td>
<td>Gln 7 OE1</td>
<td></td>
</tr>
<tr>
<td>Val 13 O</td>
<td>2.83</td>
<td>Gln 7 NEH2</td>
<td></td>
</tr>
<tr>
<td>Ala 97 O</td>
<td>2.99</td>
<td></td>
<td>Lys 103 NZH</td>
</tr>
<tr>
<td>Ala 102 O</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The zinc ions are shown as grey spheres. Each subunit, depicted in one of four different colours, is involved in contacts with every other subunit in the tetramer. The 222 symmetry of the molecule can be clearly seen.
Molecule A is shown as an electrostatic potential surface. The blue colouring indicates regions of positive potential whereas the red indicates regions of negative potential. Molecule B is depicted as a liquorice model. The figure shows Met140 docked into a hydrophobic pocket on the surface of molecule A. There are additional hydrophobic contacts between Gly109, Ala108, Val112 and Ile138 of molecule B and an equivalent region in molecule A.
Molecule A is shown as an electrostatic potential surface. The blue colouring indicates regions of positive potential whereas the red indicates regions of negative potential. Molecule C is depicted as a licorice model. The docking of the two leucines from molecule C into apolar cavities can be clearly seen. The other hydrophobic residues on the N-terminal loops of both molecules are buried by these regions packing against each other.
Figure 32  Intermolecular contacts - Interface 3.

Molecule A is shown as an electrostatic potential surface. The blue colouring indicates regions of positive potential whereas the red indicates regions of negative potential. Molecule D is depicted as a licorice model. Lys103 is docked into a hydrophobic cavity and its positive charge is stabilised by the formation of the bifurcated hydrogen bond with two alanine residues on the molecular surface (Not shown).
6.4 Inhibition of L1.

Analysis of the crystallographic data collected from pre-soaked or co-crystallised L1-inhibitor crystals showed that these were actually native with no inhibitor bound. There are several reasons why the inhibitor may not have been observed at the active site.

- The molecule is not a good inhibitor.
- The inhibitor molecule is unable to gain access to the active site of the enzyme in the crystal.
- The enzyme in the crystal is restrained such that it cannot bind the inhibitor properly.

The first of these was investigated by studying the inhibition kinetics of L1 in the presence of the TFK inhibitor. The results are presented in table 17.

Table 17 The effect of the TFK inhibitor on $K_m$ and $V_{max}$.

<table>
<thead>
<tr>
<th>[Inhibitor] (nanomolar)</th>
<th>$K_m$ (micromolar)</th>
<th>$V_{max}$ (M s$^{-1}$) x 10$^{-7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15.7±2.2</td>
<td>1.29±0.22</td>
</tr>
<tr>
<td>10</td>
<td>15.4±2.4</td>
<td>1.31±0.17</td>
</tr>
<tr>
<td>20</td>
<td>15.1±3.1</td>
<td>1.28±0.30</td>
</tr>
<tr>
<td>100</td>
<td>16.1±1.7</td>
<td>1.31±0.26</td>
</tr>
<tr>
<td>1000</td>
<td>16.3±3.3</td>
<td>1.32±0.21</td>
</tr>
</tbody>
</table>

The reported $K_i$ of the inhibitors used in this study is around 5 nM (Walter et al., 1997). The concentration of protein used to initiate the reaction was approximately 3nM.

The results of the inhibition studies with L1 show that neither $K_m$ or $V_{max}$ are significantly changed. The conclusion is therefore that the TFK molecules used in this
work are either not inhibitors of the metallo β-lactamase L1 or are less potent than reported in the literature (i.e. $K_i$ value of greater than 1μM). These results are discussed in detail in the next chapter. The values of $K_m$ and $V_{max}$ for hydrolysis of meropenem determined in this work, in the absence of inhibitor, are consistent with those reported by Felici et al., (1995). A summary of the data is shown in table 18.

Table 18    Kinetics of meropenem hydrolysis.

<table>
<thead>
<tr>
<th>[substrate] (micromolar)</th>
<th>$V_{max}$ (M s$^{-1}$) x10$^{-10}$</th>
<th>$K_m$ (micromolar)</th>
<th>$K_{cat}$ (seconds$^{-1}$)</th>
<th>$K_{cat}/K_m$ (M$^{-1}$ s$^{-1}$) x10$^7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-250</td>
<td>1.30±0.22</td>
<td>15.7±2.2</td>
<td>36</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Chapter 7.

Discussion.
7.1 Discussion of phasing procedures.

Despite the similarity in the overall folds of the metallo β-lactamases solved to date, the structure of L1 could not be determined by molecular replacement (MR). This is probably because of the low sequence homology between L1 and the other class B β-lactamases (see chapter 1, table 1), which would manifest itself as a significant r.m.s deviation between the model and the novel structure. Minor topological difference exist between the known structures. In BCII there are 5 adjacent strands in the N-terminal domain and 4 adjacent strands in the C-terminal domain, which is the opposite to than observed in L1. In CcrA, there is an additional β strand that is not observed in either BCII or L1. Apart from this strand, the remaining secondary structural elements occur at similar positions along the sequence for all three enzymes. However, a structure based alignment of L1 with ccrA and BCII (using just c-alpha models) showed an average r.m.s deviation of 7.1 Å (see figure 33). The N-terminal domain of the structures appear to be more similar (average r.m.s deviation of 4.1 Å) than the C-terminal domains (average r.m.s deviation of 10.24 Å). In addition to the large r.m.s, the low solvent content of the L1 crystals (ca. 40 %) made it difficult to specify a sphere of integration for the Patterson search which would include most of the self vectors whilst minimising the contributions from cross vectors. In contrast, the r.m.s difference between ccrA and BCII is only 1.65Å (all main chain atoms) and indeed, Carfi et al., (1998b) were able to determine the structure of ccrA using MR with BCII as the search model.

Attempts to obtain isomorphous derivatives of L1 were also unsuccessful. Initial reports suggested that L1 binds two zins at the active site, a fact now confirmed by the structure presented here, which made it a candidate for use in MAD experiments. However due to the relatively small anomalous signal of zinc, and hence the uncertainty of the success of the MAD experiment using just the intrinsic zinc centres, crystals were grown from protein that had Se-met incorporated. Fluorescence scans showed that there was a
Figure 33. Alignment of secondary structural features of L1 and CcrA.

L1 is shown in red and CcrA in blue. The overall r.m.s. deviation is 7.1 Å.
significant signal at the absorption edges of both metals. This phenomenon was exploited by collecting \( f' \) and \( f'' \) for both types of centres, leading ultimately to the high quality of the phases obtained from the experiment (table 9 page 89). The absorption edge of Se is at a lower wavelength than that of Zn and therefore, the anomalous scattering from the zinc contributes to the anomalous signal present in data collected at the Se edge. This could have made the task of finding the selenium sites more difficult because of additional cross peaks between the Se and Zn positions. However, once correct sites are found the quality of the phases are not diminished.

The use of the 'zinc' phases in the calculation of the Se dispersive Fourier simplified the problem of locating the selenium sites considerably. By this approach 5 of the 6 Se positions were identified and used in phasing. A peak corresponding to the sixth selenium atom was also present in the dispersive Fourier but was initially overlooked. Post priori to structure determination, MLPHARE with all six selemiums and two zincs gave phases with figure of merit of approximately 0.9.

7.2 CCP4 and SOLVE approaches to phase determination.

A comparison of CCP4 (HASSP and MLPHARE) and SOLVE methods to determine the positions of anomalous scattering atoms was performed. The advantage of SOLVE is that it is a highly automated procedure for finding sites in a MAD or a MIR experiment. SOLVE is able to do this because it uses each solution as a seed to find other solutions; that is it finds other sites by calculating difference Fouriers using the site it already has. The CCP4 approach relies on the sites generated by HASSP from difference Pattersons (from MIR/MIRas or MAD data). If there are many sites the difference Patterson can be difficult to interpret.
In this work, both approaches (CCP4 (HASSP) and SOLVE) were able to determine the zinc positions and gave similar FOMs and occupancies (table 12, page 93). The error of the phases calculated using the zinc sites is approximately 47° (table 13, page 94) and secondary structural features can be easily identified in the resulting electron density map. Therefore in this experiment, the zinc atoms alone would have been sufficient for structure determination by MAD.

SOLVE appeared to determine all six Se sites directly from the Se edge data (table 12, page 93). However, close examination of these sites showed that only four of the six sites correspond to actual selenium positions whilst the other two corresponded to the two zinzs. In retrospect, this could have occurred because of the contribution from the anomalous signal of zinc to the anomalous signal of selenium. This highlights a problem with the SOLVE procedure in that the sites determined may be incorrect in the presence of more than one type of anomalous scattering atom. The procedure was also unsucceful in determining all eight sites (two zinzs and six seleniums) when given both sets of MAD (Zn and Se) data.

The advantages of MAD phasing are clearly illustrated in this work. We were able to obtain initial phase estimates that gave a good electron density map and therefore a good initial model could be built. Overall, the quality of the phases derived from the MAD data is very good as could be seen by the relatively small phase errors (Table 14).

7.3 Structural features of L1.

The structure of the monomeric unit of L1 (αβα fold) is similar to that of BCII and ccrA (figure 34). The αβα fold has been previously observed in the N-terminal nucleophile (NtN) aminohydrolase family (Brannigan et al., 1995), DNASE I (Oefner et al., 1986) and
exonuclease III (Mol et al., 1995). These structures have very little sequence homology with the metallo β-lactamases and the topologies of the folds are different. It is likely then that the observed structural similarity is due to convergent evolution. There are minor topological differences between the known metallo β-lactamase structures themselves. BCII has five adjacent β strands at the N terminus of the polypeptide chain and four β strands after the interdomain helix (α₃); the opposite is observed in L1. CcrA has an additional β strand between the helices α₃ and α₄ which is not observed in BCII or L1.

Figure 34  A comparison of the structures of BCII, ccrA and L1.

The crystal structures of the metallo β-lactamases ccrA, L1 and BCII are shown above after they have been approximately aligned. The zinc ions are shown as grey spheres. The sequence homology between BCII and ccrA is significant (ca. 30%) but both share little homology with L1 (ca. 12%) (see table 1.0).
L1 displays the least sequence similarity to the other known metallo β-lactamases (table 1) and an alignment of these sequences (figure 5) shows that if the L1 were excluded, a further 16 residues would be conserved between the remaining sequences (a total of 23). If the *A. hydrophilla* enzyme is also excluded, the total number of conserved residues would increase to 46. These statistics demonstrate the extent of the sequence dissimilarity of L1 to the other class B enzymes.

L1 is organised as a tetramer in the crystal lattice but this, of itself, cannot be taken as evidence of the enzyme existing this way in solution. Prior to this work there was some suggestion that L1 might exist as a tetramer from gel filtration data (Saino *et al.*, 1982; Dufrense *et al.*, 1988; Paton *et al.*, 1994; Crowder *et al.*, 1998). This suggestion was definitively verified by the sedimentation equilibrium studies described here. In the spacegroup P6422 there are two different types of centre possessing 222 point group symmetry. The choice of which of these two ‘tetramers’ represented the structure of the molecule in solution was based on the extent of the intermolecular contacts.

The geometry of the active site of L1 is superficially reminiscent of that observed in ccrA (Concha *et al.*, 1996). However a closer comparison of the active sites of L1, ccrA and BCII (the latter two are virtually identical) shows that the structure reported here is substantially different from those reported previously (figure 35). The arrangement of the tetrahedral zinc ion, Zn1, and associated ligands is similar in all three structures and the position of the bridging water or shared hydroxide is little changed between ccrA and L1. It is at the second zinc ion, Zn2, and more particularly at the positions of the co-ordinating ligands that the most significant differences occur. In the active site of ccrA the equatorial positions in the trigonal bipyramidal co-ordination of Zn2 are occupied by the conserved residues His206 and Cys164 (His210 and Cys168 in BCII) and the shared hydroxide. The axial positions are filled by the conserved Asp86 (Asp90 in BCII) and an additional apical water Wat2. In L1, the conserved protein ligands are His 225 and Asp88. However the structure comparison shows that Asp88 cannot be superimposed.
with the equivalent acidic residue in ccrA and BCII. That is Asp88 adopts a position that is unique to L1. The reasons for this is likely to be the substitution of Cys 164/168 (ccrA/BCII) for a serine (Ser185) in L1 which is only able to act as a ligand indirectly through its hydrogen bonded water, Wat2.

**Figure 35**  Comparison of the active sites of ccrA and L1.

The active site of ccrA and L1 are shown in blue and red respectively (top). The superimposition of the two active sites is also shown (bottom).
In L1, the coordination sphere around Zn2 is completed by His89 acting as the third equatorial ligand. This fifth ligand (His89) is therefore provided from a completely different region of sequence to the fifth ligand in either BCII (Cys168) or ccrA (Cys164). The effect is, whilst Asp88 and Wat2 are retained as apical ligands, the positions of the three planar ligands (Wat1, His89, and His225) are moved such that the entire plane has undergone a rotation, with respect to the equivalent plane in ccrA, of approximately 76 degrees about the Wat1-Zn2 bond. An additional feature of metal coordination in L1 is the distorted trigonal bipyramidal geometry about Zn2. The Wat1-Zn2-His225 angle is 146° whilst the angle His225-Zn2-His89 is 102°. The apical ligands are also 14° away from colinearity. In addition, the bond between Zn2 and Wat2 is considerably longer than the other bonding distances. The mechanistic relevance of these observations will be discussed later.

A single sterically strained aspartate (Asp53 in L1) is conserved amongst the known metallo β-lactamases. In the structure of L1 the side chain of Asp53 makes hydrogen bonds with the main chain amide and side chain of Ser83 and with a water molecule Wat23. In L1 Wat23 is hydrogen bonded to His89ND1 and thus it appears that Asp53 acts indirectly to stabilise the appropriate tautomeric form of this histidine. In ccrA, Wat23 is replaced by a Na⁺ ion. This Na⁺ ion is penta-coordinated and is proposed to be important in the orientation of three zinc liganding groups.

### 7.4 Inhibition studies

Substrate binding to enzymes can be indirectly investigated using inhibitors that bind to the active site. The inhibitory effects of trifluoromethyl ketones (TFK’s) have been demonstrated previously for carboxypeptidase A, serine proteases, and in the metallo β-lactamases (see chapter 1). In the latter case, inhibition has been shown to be competitive...
(Walter et al., 1997). However, these results could not be reproduced in this work. That is the activity of L1 was not inhibited by increasing concentrations of the inhibitor. These results are puzzling given the reported potency of the TFK molecules as inhibitors of L1 (see chapter 1). There are several possible explanations. The first was that the inhibitor is hydrolysed by the enzyme. This hypothesis could not be tested by a spectroscopic assay in the visible spectrum because the inhibitor does not contain a chromophoric group. Also the phenyl group of the inhibitor has only a small signal in the UV spectrum and in any case is not likely to be affected during the hydrolysis reaction. It is also possible that the TFK molecule will have degraded in solution to a form that was no longer an inhibitor of the metallo β-lactamases.

The TFK molecule used in this work was only soluble in organic solvents such as dimethyl sulfoxide (DMSO) and methanol. DMSO could not be used for crystallographic analysis because its presence caused severe cracking of the L1 crystals. On the other hand, TFK dissolved in methanol did not damage the crystals. However, it was later discovered that methanol could, and does, react with the keto group of the TFK to form a hemiketal. This is believed to have disrupted the binding interactions between L1 and the TFK.

TFK molecules are highly reactive and readily undergo addition reactions to form either hydrates or hemiketals (see figure 36) in the presence of water and alcohols (J P Guthrie, 1977). The equilibrium constant of the hydrate to carbonyl reaction for an aralkyl trifluoro ketone (C₆H₅COCF₃) was found to be about 78. This showed that under normal conditions (i.e., at pH 7.0 in aqueous solutions at 25°C) the compound is largely present in the hydrated form. Similarly, in the presence of alcohols, such as methanol, the equilibrium constant for the transition of aralkyl ketone to the hemiketal is 270 so that almost all of the compound will be present as the hemiketal. The formation of the ketal from the hemiketal was found to be extremely slow and an equilibrium constant was not calculated. The equilibrium constants could be measured with the fluorinated compound

127
because $^{19}$F shifts are very sensitive to substituents and the signals from hydrate and hemiketal forms are well separated in the nuclear magnetic resonance (nmr) spectra.

**Figure 36**  Reactions of the TFKs with water and organic solvents.

![Diagram](image)

It is therefore likely that TFK inhibitors are present, under standard conditions, as hydrates or hemiketals in solution prior to binding to the protein. This would make sense because the inhibitors, as hemiketals or hydrates, would then resemble the proposed transition state of the β-lactam and thus bind tightly. Indeed, the binding of the TFK compound used in this work was reported to be around 3 nanomolar, much greater than the binding affinities of the β-lactam substrates (Walter et al, 1998; Felici et al, 1993). This mode of binding is supported by the structure of the complex of a TFK with carboxypeptidase A (CPA) (Christianson et al, 1986). Here the keto group has reacted with water and is bound to the active site in its hydrated form; the interactions are outlined below (figure 37). The tetrahedral geometry of the hydrated ketone is similar to the peptide carbonyls in their transition state in the reaction with CPA (Christianson et al, 1986).
Based on the above observations and substrate modelling, which is discussed in the next section, the binding of a TFK inhibitor to L1 could occur as follows. Modelling work suggests that substrate binding is facilitated by the co-ordination of the substrate via the zinc ions. Therefore, the TFK molecule would presumably co-ordinate to Zn ions via one of the hydroxyl group of the hydrated ketone moiety and the phenol group could form stabilising hydrophobic interactions with loop A; a model of the binding is shown in figure 38. However, in the hemiketal form, as result of the reaction with methanol, the addition of the sterically larger and more hydrophobic methoxy group may disrupt important binding interactions with L1.

**Figure 37** The complex of carboxypeptidase A with a TFK inhibitor.

One of the hydrated oxygens forms a bifurcated hydrogen bond with Glu270 and Ser197. The first hydrated oxygen forms an electrostatic interaction with the zinc and participates in forming a hydrogen bond with Arg127. The carboxyl group of the molecule is stabilised by two hydrogen bonds between Tyr248 and Arg145. The benzyl group is buried in a hydrophobic pocket called the specificity pocket. The trifluoro substituent does not appear to be participating in binding.
Figure 38  A scheme of the complex of L1 and a TFK inhibitor.

A simple model of the TFK hydrate bound to L1

The hemiketal may not be able to make the same interactions as the hydrate because of steric blocking by the methoxy group.
7.5 Substrate binding.

As there was no crystallographic data regarding inhibitor binding molecular modelling studies were carried out. These calculations were carried out by C. S. Verma at the University of York. The modelling procedure involved manual docking of substrate into the active site followed by Monte-Carlo energy minimisation calculations. The docking procedure was initially guided by three assumptions. The first assumption was based on the need for a good nucleophilic group; Wat1 seemed the best candidate to fulfil this role as it is likely to exist as a hydroxide ion and appears suitably positioned to make a nucleophilic attack on the carbon of the β-lactam carbonyl. Secondly it was assumed that Zn1 would co-ordinate to the β-lactam carbonyl oxygen and consequently increase the polarisation of this group. Finally, it was assumed that Ser187 would form a hydrogen bond with the conserved carboxyl moiety of the substrate. It has been proposed that a lysine at an equivalent position in ccrA (Lys167) makes a charged hydrogen bond with the same carboxyl group. From the initial modelling, it became apparent that a consequence of these interactions is that the β-lactam nitrogen is brought into close proximity to Zn2 (ca. 2.8 Å). The entry of this nitrogen into the co-ordination sphere of Zn2 may be facilitated by the distorted trigonal byprimidal geometry observed in the apo-enzyme. In the model of the enzyme-substrate complex Zn2 is therefore hexa-co-ordinated with octahedral geometry.

Three substrates, ampicillin (a penicillin), ceftazidime (a cephalosporin), and imipenem (a carbapenem) were modelled in to the active site. These three were chosen so that the mode of binding could be compared with the suggestions of Concha et al. (1996). The modelling studies indicate that the bulky hydrophobic side chains at position 2 of ampicillin and ceftazidime pack against hydrophobic residues located around the active site. These include Phe124 and Ile128 from the loop region between α3 - β7 (loop A). An equivalent loop in ccrA has been implicated in the binding of both substrate and
inhibitor (see chapter 1). Loop B is also close to the active site (see figure 24) and contains a conserved glycine (Gly195) and a largely conserved asparagine (Asn196) but the modelling studies do not suggest a role for this loop in substrate binding. However, Concha et al., (1996) propose that the equivalent asparagine in ccrA (Asn193) participates with Zn1 in forming an oxyanion hole. The equivalent Asn in L1 is approximately 14Å away from the carbonyl oxygen of the substrate and therefore, a substantial rearrangement would be necessary for it to participate in catalysis.

For all three of the substrates modelled into the active site of L1, there was no single solution which was clearly better than the others. Instead, a family of related solutions was found in each case (see figures 39, 40 and 41). These solutions are suggested to constitute an ensemble of interactions of approximately the same energy. This observation, if true, would be consistent with the rather broad range of substrate specificity observed for L1.
Eight possible binding conformers were found for ampicillin within a range of 88 kcal/mol. The energy terms are somewhat arbitrary and do not correspond in real terms to kcal/mol. The aromatic side chain of ampicillin could pack against Ile124 of loop A or with the Tyr11, of the N-terminal loop region.
Six binding conformers were found for ceftazidime within a range of 24 kcal/mol. The energy terms are somewhat arbitrary and do not correspond in real terms to kcal/mol. The aromatic side chain extending from position 2 of the substrate appears to pack against Tyr 11 and Trp 17 from the N-terminal loop. The side chain of the substrate at position 6 is packing against loop A.
For imipenem, 10 binding conformers were found within an energy range of 7 kcal/mol. The energy terms are somewhat arbitrary and do not correspond in real terms to kcal/mol. These conformers again appear to interact with either Loop A (Ile128) or the N-terminal loop (Trp17).
7.6 Mechanism of catalysis.

Models for the catalytic mechanism of the metallo β-lactamases proposed by Concha et al (1996) and Bounaga et al (1998) have been reviewed in chapter 1. I present here a mechanistic model based on the structure of L1 and related modelling studies. A schematic of the proposed reaction pathway for L1 is shown in figure 41. On the approach of substrate, Zn1 and Zn2 co-ordinate to the carbonyl oxygen and the β-lactam nitrogen adopting trigonal bipyramidal and octahedral geometry respectively. An additional interaction occurs between Ser187 and the carboxyl group at position 3 of the substrate. Non specific hydrophobic interactions between Loop A and the large hydrophobic substituents that are commonly found at position 2 of many β-lactams also contribute to the overall binding.

The binding interactions outlined serve to approximately orient the substrate carbonyl carbon for nucleophilic attack by Wat1, which is believed to exist as a hydroxide ion. This produces a tetrahedral transition state in which the carbonyl oxygen exists as an oxyanion. The models of Concha et al., (1996), and Carfi et al., (1995), postulate that this chemical species makes additional interactions with a largely conserved asparagine at position 180 (BCII) or 193 (ccrA), on loop B, which together with Zn1 forms an oxyanion hole. The equivalent residue (Asn196) is conserved in L1, but as discussed in section 7.4 it is approximately 14Å away from the substrate carbonyl oxygen in the apo-enzyme structure. The modelling procedure cannot account for large scale conformational changes that may take place upon substrate binding and therefore one cannot rule out the involvement of Asn196. However, the modelling studies did suggest an alternative in which the oxyanion species is stabilised by the oxyanion hole generated by Zn1, the side chain of Tyr191 (either directly or through a water molecule) and a positive dipole at the N-terminus of a short helical turn made by residues 86-91.
Product formation occurs upon the breakage of the carbon-nitrogen bond of the β-lactam ring. This requires the donation of a proton to the nitrogen. From the ensemble of ampicillin structures, the nearest potential donor is Wat2 which is approximately 3.5Å away. Concha et al. (1996, 1997) make a similar proposal despite the different arrangement of ligands about Zn2 in ccrA compared with L1. In L1 the proposed role of Wat2 may be favoured by its interaction with Ser185 which may increase the acidity of the water. In addition, the formation of the oxyanion transition state is likely to be accompanied by a loss of co-ordination at Zn1 by the carbonyl oxygen as the geometry alters from planar to tetrahedral. A consequence of this is a rearrangement of ligands...
about Zn1. Wat2 is positioned to interact with Zn1 in the transition state complex, forming a contact which would make Wat2 more acidic and therefore more likely to protonate the amide nitrogen. Protonation of the β-lactam nitrogen leaves Wat2 in position as the bridging hydroxide between the two zinc ions. On dissociation of product, Wat2 becomes equivalent to the ground state bridging ligand Wat1 and an incoming ligand at the apical position of Zn2 is recruited from bulk solvent.

Alternative mechanisms proposed by Carfi et al., and Bounaga et al., suggest that the conserved aspartate (Asp86 in BCII, Asp99 in ccrA and Asp88 in L1) is a general acid in the reaction and donates a proton to the substrate nitrogen to facilitate hydrolysis of the lactam ring. In Carfi’s model, this Asp is thought to first deprotonate Wat1 to produce the more nucleophilic hydroxide species and subsequently donate the extracted proton to the substrate nitrogen, thereby acting as a proton shuttle. In Bounaga’s model, in contrast, Wat1 is thought to exist predominantly as the hydroxide ion and the conserved aspartate is instead proposed to deprotonate the oxyanion to form a dianionic species (see chapter 1, figure 9).

The breakage of the C-N bond in the β-lactam ring is not easily achieved and it is suggested that the collapse of the tetrahedral intermediate to give intact β-lactam would occur faster than C-N bond fission (Page et al., 1998). Bounaga et al. (1998) and Page et al. (1998) suggest that the formation of the dianionic species will make C-N bond fission more favourable than the reverse reaction. The argument is that the presence of the Zn bound hydroxide may increase the basicity of the oxyanion, by allowing a partial delocalisation of the negative charge, and thereby decrease the reactivity of this species. However, if the tetrahedral oxyanion is deprotonated to form the dianion, the negative charge will remain localised and the species will in general be more reactive. This negative charge can then participate in forming a double bond with the carbonyl carbon. A
consequence of this will be increased electron density around the substrate nitrogen, which can then make a nucleophilic attack on the nearby conserved aspartate.

Although one cannot rule out the mechanisms proposed by Carfi et al. and Bounaga et al., it is unlikely that the conserved aspartate in L1 can act as a proton shuttle. This is because the positively charged active site is likely to depress the pK of the aspartate below 3.9 making the protonation of its side chain unfavourable. Furthermore, from modelling studies with L1, the side chain of this aspartate is 4.8Å from the lactam nitrogen.

The mechanistic models that have been postulated so far all suggest that the reaction proceeds as a concerted process and until recently, there was little evidence to suggest otherwise. However a recent study of catalysis by ccrA by Wang et al. (1998), using stop-flow methods, suggests that reaction proceeds via the formation of a reaction intermediate that is broken down to product and enzyme. The rate constants for the consumption of substrate and the formation of products were found to be 900 s\(^{-1}\) and 150 s\(^{-1}\) respectively. The differences between the two rates suggest that an intermediate accumulates between substrate consumption and product formation. Wang et al. also provide evidence that the breakdown of the intermediate is the rate limiting step; it was found that formation of the intermediate is pH and D\(_2\)O insensitive whereas the rate of breakdown is reduced by approximately 50% in D\(_2\)O.

In summary, the catalytic mechanisms proposed so far agree that the reaction is likely to be initiated by nucleophilic attack from the zinc bound hydroxide ion. The most significant difference in the mechanisms proposed resides in the breakdown of the transition state complex, although all agree that a proton transfer event is likely to be involved. There is some evidence to suggest that the reaction could proceed through the accumulation of a reaction intermediate in ccrA. As discussed in the introduction, the inhibition of BCII, ccrA and L1 is different and perhaps reflects real differences in
substrate binding and the mechanism of catalysis. Therefore, it is not wise to assume a global mechanism at this time and more work is required to unravel the events that take place in the active sites of these enzymes.
References.


Harker D (1936). The application of three dimensional Patterson method and the crystal structure of Proustite, Ag$_3$AgS$_3$ and Pyrargyrite, Ag$_3$SbS$_3$. *J. Chemical Physics* 4, 381-390.


In _Isomorphous Replacement and Anomalous Scattering_, Proceedings of the CCP4 study weekend (Wolf W, Evans P R and Leslie A G W, eds), 80-86, Daresbury Laboratory, Warrington, UK.


Thompson J D, Higgins D G and Gibson T J (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignments through sequence weighting,


