Structure-Function Studies of β-lactam Biosynthetic Enzymes

Linda Öster
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2005

Acta Universitatis Agriculturae Sueciae
2005:1

ISSN 1652-6880
ISBN 91-576-7000-5

Tryck: SLU Service/Repro, Uppsala 2005

Abstract

Öster, L., 2005. *Structure-Function Studies of β-lactam Biosynthetic Enzymes*. Doctoral dissertation.
ISSN 1652-6880, ISBN 91-576-7000-5.

β-lactam compounds belong to the most important antibiotics in current use. The increasing occurrence of bacterial resistance against antibiotics, which threatens to move us back to pre-antibiotic era, calls for the development of new antibiotics. β-lactams are produced by fermentation in the microorganism, therefore knowledge about the biosynthetic enzymes involved is vital for the production of new antibiotics to escape resistance. This thesis presents structure-function studies on enzymes involved in the biosynthesis of the β-lactam antibiotics cephalosporins and cephamycins

DAOCS is a non-heme Fe(II) dioxygenase and catalyses the oxidative ring expansion of the penicillin nucleus into the nucleus of cephalosporins. Structures of DAOCS with substrates, product and cofactors obtained from two different crystal forms are presented here. The structural results suggest a mechanism for cephalosporin formation where 2-oxoglutarate and dioxygen need to react first to produce an oxidizing iron species, followed by reaction with the penicillin substrate. Structural differences in the two crystal forms indicate conformational flexibility in the C terminus and point to a role for the C terminus in catalysis.

Biosynthesis of cephamycins (C7-methoxylated cephalosporins) is catalysed by two enzymes, cmcI and cmcJ, but the details of catalysis are largely unknown. The crystal structure of cmcI, presented here, is a hexamer consisting of a C-terminal Rossman domain and a smaller N-terminal domain. The N-terminal domain is involved in oligomerisation and the Rossmann domain binds SAM and SAH in a fashion common for methyltransferases, with a bound magnesium ion in the vicinity of SAM. The expected cephalosporin binding site is occupied by PEG, which ligates the magnesium ion. From docking studies of a cephalosoporin molecule to the cmcI-Mg^{2+}-SAM structure, a model for substrate binding is proposed. Altogether, the results suggest cmcI is a methyltransferase that catalyses the second catalytic step in cephamycin biosynthesis.

Keywords: β-lactam antibiotics, 2-oxoglutarate dependent dioxygenases, cephamycin biosynthesis, protein crystallography, *Streptomyces clavuligerus*, deacetoxycephalosporin C Synthase, cmcI

Author's address: Linda Öster, Department of Molecular Biology, SLU, S-751 24 Uppsala, Sweden

Contents

Appendix

Appendix A

Table of data collection and refinement statistics for cmcI-NADH structures presented in chapter 5.

Papers I-IV

This thesis is based on the following papers, which will be referred to by their Roman numerals:

I. Valegård, K., Terwisscha van Scheltinga, A.C., Dubus, A., Ranghino, G., Öster, L.M., Hajdu, J. & Andersson, I. (2004) The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. *Nat. Struct. Mol. Biol.* 11, 95-101.

II. Öster, L.M., Terwisscha van Scheltinga, A.C., Valegård, K., MacKenzie Hose, A., Dubus, A., Hajdu, J. & Andersson, I. (2004) Conformational flexibility of the C Terminus with Implications for Substrate Binding and Catalysis Revealed in a New Crystal Form of Deacetoxycephalosporin C Synthase. *J. Mol. Biol.* 343, 157-171.

III. Lester, D.R., Öster, L.M., Svenda, M. & Andersson, I. (2004) Expression, purification, crystallization and preliminary X-ray diffraction studies of the cmcI component of *Streptomyces clavuligerus* 7α-cephem-methoxylase. *Acta Crystallogr. D* 60, 1618-1621.

IV. Öster, L.M., Lester, D.R., Terwisscha van Scheltinga, A.C., Svenda M., Généreux, C. & Andersson, I. (2004) Insights into Cephamycin Biosynthesis: the Crystal Structure of cmcI from *Streptomyces clavuligerus*. *In Manuscript*

Abbreviations

2ODD	2-oxoglutarate dependent dioxygenase
6-APA	6-aminopenicillianic acid
7-ACA	7-aminocephalosporanic acid
ACVS	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase
ANS	anthocyanidin synthase
AtsK	alkylsulfatase
CarC	carbapenem synthase
CAS	clavaminate synthase
COMT	cathechol O-methyltransferase
DAC	deacetylcephalosporin C
DACS	deacetylcephalosporin C synthase
DAOC	deacetoxycephalosporin C
DAOC/DACS	deacetoxy/deacetylcephalosporin synthase
DAOCS	deacetoxycephalosporin C synthase
FIH	factor inhibiting hypoxia-inducible factor
IPNS	isopenicillin N synthase
MAD	multiple-wavelength anomalous diffraction
NADH	nicotinamide adenine dinucleotide, reduced form
NCS	non-crystallographic symmetry
ORF	open reading frame
PBP	penicillin binding protein
PEG	polyethylene glycol
r.m.s	root-mean-square
SAD	single-wavelength anomalous diffraction
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
TauD	taurine/α-ketoglutarate dioxygenase
Å	Ångström

1. Introduction

At the beginning of the 1900's bacterial infectious diseases such as pneumonia and tuberculosis were among the leading causes of death worldwide. This is hard to imagine today with all available drugs, but without any antibiotics, a simple bacterial infection could be fatal. After the successful development of antibiotics such as penicillin in the 1940s, victory against bacterial infectious diseases was declared. However, new infectious diseases and in particular the increasing bacterial resistance against antibiotics, leading to the re-emergence of diseases once controlled, have forced us to revaluate that statement. Today bacterial infections are once again identified as one of the greatest threats to human health (Cohen, 2000).

Antibiotics (Greek *anti*, against, and *bios*, life) are naturally occurring microbial products or their derivatives that exhibit bacteriostatic or bactericidal effects, i.e. they inhibit or kill bacteria without harm to the eukaryotic host harbouring the infecting bacteria. Antibiotics belong to a group of compounds called secondary metabolites, generally characterised by being produced at low specific growth rates, and not being essential for the producing organism (bacteria or fungi) in pure culture. However, antibiotics are critical to the organism in their natural environment both for their survival and as a competitive advantage (Demain & Fang, 2000). The antibiotics are used as chemical weapons to kill other microbes in the surrounding microenvironment.

Over the years, most antibiotics have been discovered by screening of e.g. soil samples and marine environments for natural products that kill bacteria. The synthesis of these compounds poses a very complex and demanding task for the organic synthetic chemist and the yield, if any, is mostly extremely low. Large scale commercial synthesis is uneconomical and rarely a possibility. Instead, fermentation of the bacterial or fungal strains followed by extraction and purification of the compounds is the common method to obtain antibiotics for commercial use. Frequently, chemical modifications of the natural products, so called semi-synthetic modifications, are employed to obtain new variants that are more effective and/or better suited as drugs.

To prevent the same situation as in the beginning of the twentieth century to occur, it is vital to develop a more sensible use of antibiotics as drugs and ban the use of antibiotics as growth promoters in agriculture. Since most antibiotics in current use are derived from a rather small number of basic compounds and very few new antibiotics are presently approved for clinical use (Breithaupt, 1999), there is also a great need to expand the antibiotic repertoire and to produce new antibiotic variants not yet encountered by bacteria. In addition, cheaper and more environmentally friendly routes to already existing compounds are required for industrial production. To meet these needs, one way may be to study in detail the micro-organisms that produce antibiotics and to use their present machinery or modify it for novel antibiotic production.

1.1 β-lactam antibiotics

The discovery and development of the β-lactam antibiotics are among the most powerful and successful achievements of modern science and technology. In particular penicillins and cephalosporins represent the world's major biotechnology products with worldwide sales of ~ US$ 15 billion per year or ~ 65% of the total world market for antibiotics (Elander, 2003).

1.1.1 History - the discovery of penicillins, cephalosporins and cephamycins

Penicillins

In 1928 in London, Alexander Fleming by chance discovered that *Penicillium notatum* spores could lyse colonies of staphylococci and that the active factor, which he gave the name penicillin, could be found in the filtrate of the mould culture (Fleming, 1929). There were no further breakthroughs until in 1940, when Howard Florey and Ernst Chain undertook detailed studies of penicillin in Oxford. They demonstrated its chemotherapeutic activity in mice and confirmed its lack of toxicity (Florey *et al.*, 1949). Production of penicillin by the strain of *Penicillium notatum* was very slow and it took long before the first successful clinical tests on humans could be performed. The need for large-scale production was essential but due to World War II, this could not be done in England. Instead, collaborations with pharmaceutical companies in the U.S.A. led to strain improvements and medium modifications which increased the yield of penicillin 100-fold within two years. Eventually, in 1943 penicillin G (benzylpenicillin) was used to successfully treat those wounded in the war. Since then a number of different penicillins have been discovered and used in the clinic. In 1945 Fleming, Florey and Chain received the Nobel Prize for the discovery and production of penicillin.

Cephalosporins

Following the discovery of penicillin, a range of naturally occurring antibiotics were discovered. In Sardinia, Giuseppi Brotzu isolated a strain of *Cephalosporium acremonium* (now called *Acremonium chrysogenum*) from seawater near a sewage outfall in 1945. The strain secreted material with activity against a number of Gram-positive and Gram-negative bacteria (Brotzu, 1948). Brotzu had neither the facilities nor the expertise to carry his work further, so a culture of the fungus was sent to Howard Florey in Oxford. A number of different β-lactam antibiotics were isolated and analyzed (Burton & Abraham, 1951). Among the most important was penicillin N, a new penicillin with a hydrophilic side chain that showed greater activity against Gram-negative bacteria compared to the currently known penicillins with hydrophobic side chains. Another new discovery was cephalosporin C, which showed a rather low antibacterial activity compared to penicillins, but was stable in the presence of penicillin β-lactamases (β-lactamases are β-lactam hydrolysing enzymes which at that time had become a serious problem). The low potency of cephalosporin C was soon overcome by the introduction of semi-synthetic modifications, resulting in a great number of different cephalosporins in clinical use.

Cephamycins

Initially, it was believed that only fungi produced β-lactams. However, a report from Merck & Co. in the early 1960s, stated that a strain of streptomycetes produced penicillin N (Miller *et al.*, 1962). Further work at Merck and Eli Lilly & Co. showed that various species of *Streptomyces* and *Nocardia* produced cephalosporins modified at C7 (cephamycins) and/or at the side chain attached to C3 (for numbering see Figure 1.2) (Nagarajan *et al.*, 1971; Stapley *et al.*, 1972). The discovery of cephamycin C led to research on and development of prokaryotic cephalosporins. The presence of the methoxy group on the β-lactam ring made the molecule more active against Gram-negative and anaerobic pathogens and more resistant to Gram-negative β-lactamases (Stapley *et al.*, 1979) compared to the existing β-lactams. Like fungal cephalosporin C, cephamycin C was never used clinically but was employed for semi-synthesis of many medicinally useful compounds.

1.1.2 Chemistry

The three-dimensional structure of penicillin was determined in the mid-1940s by Dorothy Hodgkin in her pioneering X-ray crystallographic work (Crowfoot *et al.*, 1949). All penicillins are chemically characterised by the presence of a four-membered heterocyclic ring, the β-lactam ring, which is fused to a five-membered thiazolidine (penam) ring (Figure 1.1).

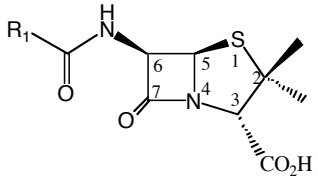

Figure 1.1 The structure of the penicillin core: 6-aminopenicillanic acid (6-APA).

Variants of penicillin contain different acyl side chains (R_1) attached by an amide linkage to the amino group of the penicillin nucleus, named 6-aminopenicillanic acid (6-APA).

In cephalosporins, the four-membered β-lactam ring is fused to a six-membered dihydrothiazine (cephem) ring (Figure 1.2). There are many different variations of cephalosporins with different side-chains at R_1 and R_2 of the cephalosporin nucleus 7-aminocephalosporanic acid (7-ACA).

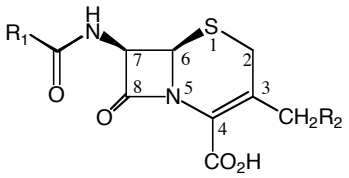

Figure 1.2 The structure of the cephalosporin core: 7-aminocephalosporanic acid (7-ACA).

β-Lactams are conventionally numbered using the ring-sulphur as position-1; for penicillins and cephalosporins the numbering is illustrated in figures 1.1 and 1.2, respectively. Substituents on the bicyclic systems are usually described by the prefix α- or β- rather than the *exo* and *endo* notation.

The cephamycins differ from the cephalosporins in the presence of a methoxyl group at the C7-position (Figure 1.3).

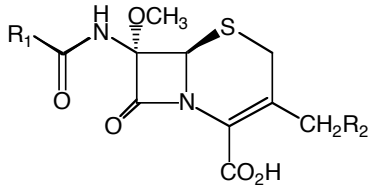

Figure 1.3 The structure of the cephamycin core.

Since the discovery of cephamycins, a number of other groups of β-lactam antibiotics have been found, many of clinical importance. Common to all of these is the β-lactam ring. Variations are achieved by alteration to the core of the five or six-membered rings, see Figure 1.4.

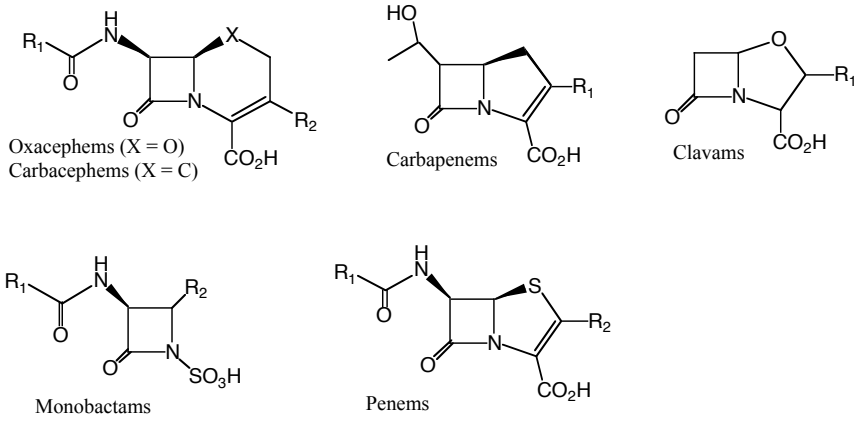

Figure 1.4 Core structures of major classes of β-lactam compounds.

Further variations of the β-lactams can be achieved by altering the different side chains at the R_1 and/or R_2 positions. By these variations, different compounds can be made that exhibit different antibiotic activity and spectrum and also influence the solubility and stability of the compound. Some of these variations are naturally synthesized but many are made by semi-synthetic modifications of fermented material, see Figure 1.5 for some examples.

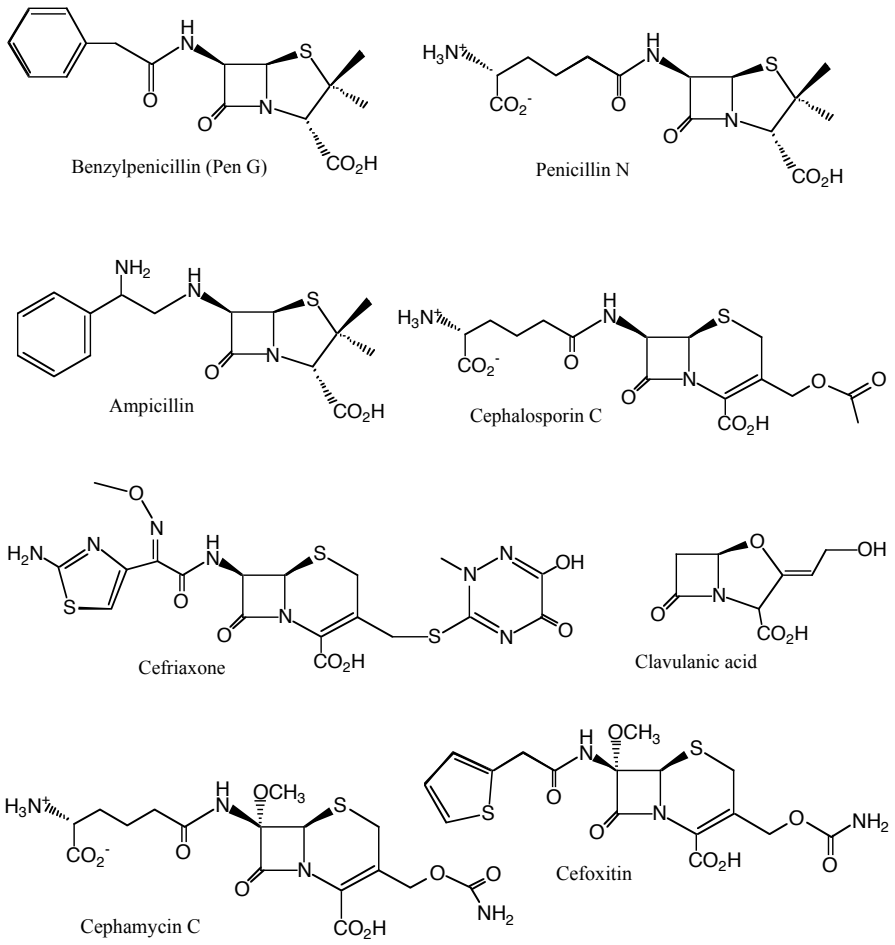

Figure 1.5 Examples of β-lactam antibiotics.

1.1.3 Mechanism of action

β-Lactams exert their lethal action on growing bacterial cells only. The addition of the antibiotic to a sensitive strain usually results in cell lysis, but the actual mechanism of killing is not entirely understood (Bayles, 2000). The β-lactam antibiotics interfere with the bacterial cell wall synthesis by inhibiting penicillin binding proteins (PBPs), also called DD-transpeptidases (Matagne *et al.*, 1999). PBP's are membrane-bound enzymes that catalyze the cross-linking of peptidoglycan polymers in the peptidoglycan synthesis. The peptidoglycan is an essential component of the bacterial cell wall with no equivalence in the eukaryotic world. The bacterial cell wall protects and determines the shape of the organism. The β-lactams acts as pseudosubstrates of the PBP's and the β-lactam ring acylates their active site serine, and forms a stable covalent acyl-enzyme complex (Figure

1.6). The inactivation of the PBP's eventually leads to a cell wall unable to withstand osmotic forces and bacteriolysis occurs.

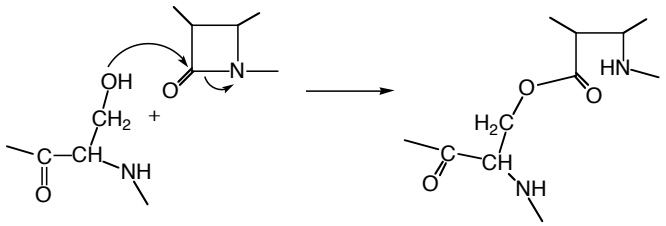

Figure 1.6 Interaction between the β-lactam antibiotics and the active site serine of the PBP, resulting in a stable acyl-enzyme complex.

1.1.4 Resistance

Already in 1940's, resistance to penicillin in some strains of *E. coli* was noted by Abraham and Chain (Abraham & Chain, 1940). During the 1950's, resistance became a serious problem as the staphylococcal population had built up resistance to penicillins via selection of penicillin β-lactamase-producing strains.

Bacteria have several strategies to escape the activity of antibiotics (Walsh, 2000): enzymatic destruction, target site (PBP) alterations, diminished permeability for the drug to enter the cell, and active efflux systems that pump the drug out of the cell. Bacteria utilise all to fight β-lactams, but the most common and well-known mechanism is the enzymatic destruction via synthesis of β-lactamases. This group of enzymes hydrolyse the essential β-lactam ring (Figure 1.7), thus precluding further reactions with PBPs.

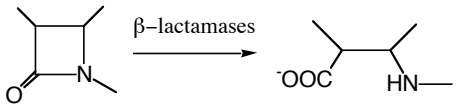

Figure 1.7 The hydrolysis of the β-lactam ring by β-lactamases.

The β-lactamases are secreted into the periplasm in Gram-negative bacteria and to the outer medium in Gram-positive species to destroy β-lactam antibiotics before they can reach the PBP targets in the cytoplasmic membrane. β-Lactamases are structurally related to PBPs and have probably evolved from them (Poole, 2004). Several hundred β-lactamases have now been identified and have been classified into four groups: A, B, C and D. Of these, the class B β-lactamases are zinc-dependent enzymes and the remaining (A, C and D) are serine hydrolases.

During the years, two main strategies have been employed to circumvent the resistance caused by β-lactamases. The first one involves alterations of the structure of the β-lactam core (Figure 1.5) to produce new variants not recognized by the β-lactamases. Unfortunately, it is usually only a question of time until a new mutation allows a modified β-lactamase to recognize and hydrolyse the new

compound. The other strategy involves inhibition of the β-lactamases. One example is clavulanic acid, a naturally occurring β-lactam (Figure 1.5) that has little antibacterial activity in itself but binds the β-lactamases irreversibly and is used together with β-lactam antibiotics (Walsh, 2000). However, clavulanic acid does not inhibit all β-lactamases.

1.2 Biosynthesis of β-lactam antibiotics

Filamentous fungi such as *Penicillium chrysogenum* and *Cephalosporium acremonium* only produce β-lactam antibiotics with penicillin or cephalosporin structure whereas bacteria such as *Streptomyces clavuligerus* and *Amycolatopsis lactamdurans* synthesize a wider variety of β-lactam structures including cephamycins, clavams, carbapenems and monobactams etc. The penicillins, cephalosporins and cephamycins share the same biosynthetic pathway starting from three amino acids. Their biosynthesis differs from the biosynthesis of clavams, carbapenems and monobactams, which are synthesised from different precursors via different biosynthetic pathways, although they all contain the β-lactam ring.

The enzymes studied in this thesis are from *Streptomyces clavuligerus*, a gram-positive bacterium that belongs to the family of actinomycetes. The actinomycetes are mycelially growing, filamentous fungal-like bacteria that grow in soil. Of these, species of Streptomyces are known as some of the best producers of secondary metabolites, including antibiotics. Two Streptomyces genomes, *S. coelicolor* (Bentley *et al.*, 2002) and *S. avermitilis* (Ikeda *et al.*, 2003), have been sequenced, revealing very large genomes for bacterial standards (8.8 Mbp and 9.0 Mbp) with many protein coding sequences and a high G + C content (>70%). A relatively large percentage of the protein coding sequences (4.5 and 6%) are predicted to encode biosynthetic enzymes involved in the production of secondary metabolites. These genes usually seem to cluster.

1.2.1 The cephamycin gene cluster in Streptomyces clavuligerus

In addition to the genes encoding the biosynthetic enzymes, the β-lactam producing clusters usually also contain accessory genes necessary for the antibiotic production. Such accessory genes are of three types: resistance genes that encode for the machinery by which the host protects itself from the antibiotic being produced, export genes that transport the antibiotic outside the cell, and regulatory genes that control the expression of genes in the cluster.

The genes responsible for cephamycin biosynthesis in *S. clavuligerus* are clustered (Figure 1.8) and have been estimated to comprise approximately 38 kb (Alexander & Jensen, 1998). The cluster is depicted in Figure 1.8, with the genes encoding the biosynthetic enzymes marked with stars, the other belong to the accessory genes.

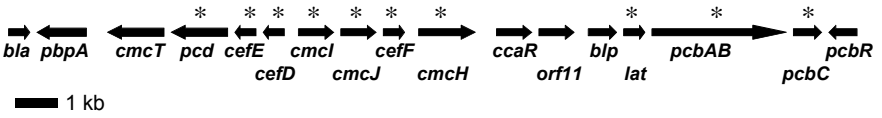

Figure 1.8 The Cephamycin gene cluster in *Streptomyces clavuligerus* with the genes encoding biosynthetic enzymes marked with stars.

1.2.2 The biosynthesis of penicillins, cephalosporins and cephamycins

The biosynthetic routes described here are outlined in Figure 1.9. The pathway starts with the condensation of three L-amino acids: L-α-aminoadipic acid, L-cysteine and L-valine with a configurational inversion of L-valine to form the linear tripeptide δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV). This is a non-ribosomal peptide biosynthesis catalysed by a single multifunctional enzyme, ACV synthetase (ACVS) (Byford *et al.*, 1997), which requires Mg^{2+} and ATP for activity.

In the next step, the ACV tripeptide is cyclized by isopenicillin N synthase (IPNS), to give isopenicillin N, a β-lactam compound with weak antibiotic activity which is the precursor of all penicillin and cephalosporin antibiotics (Baldwin & Schofield, 1992). IPNS is a 38 kDa enzyme that uses Fe(II) and molecular oxygen as cofactors. IPNS belongs to the family of non-heme Fe(II) oxygenases and cyclizes ACV by the removal of four hydrogens to form water. The crystal structure of IPNS from *Aspergillus nidulans*, the first structure of an enzyme from this family, was solved in 1995 (Roach *et al.*, 1995). Since then, there have been a number of high-resolution structures of substrate, product and intermediates published that have lead to detailed knowledge of the mechanism of IPNS (Roach *et al.*, 1997; Burzlaff *et al.*, 1999).

After the formation of isopenicillin N, there is a branch point between organisms producing cephalosporins and those producing penicillins only. The product of the next step in cephalosporin producers, penicillin N, which has not been found in any penicillin-only producers (Baldwin & Schofield, 1992), is formed from isopenicillin N in an epimerisation reaction catalysed by isopenicillin N/penicillin N epimerase (Jensen *et al.*, 1983).

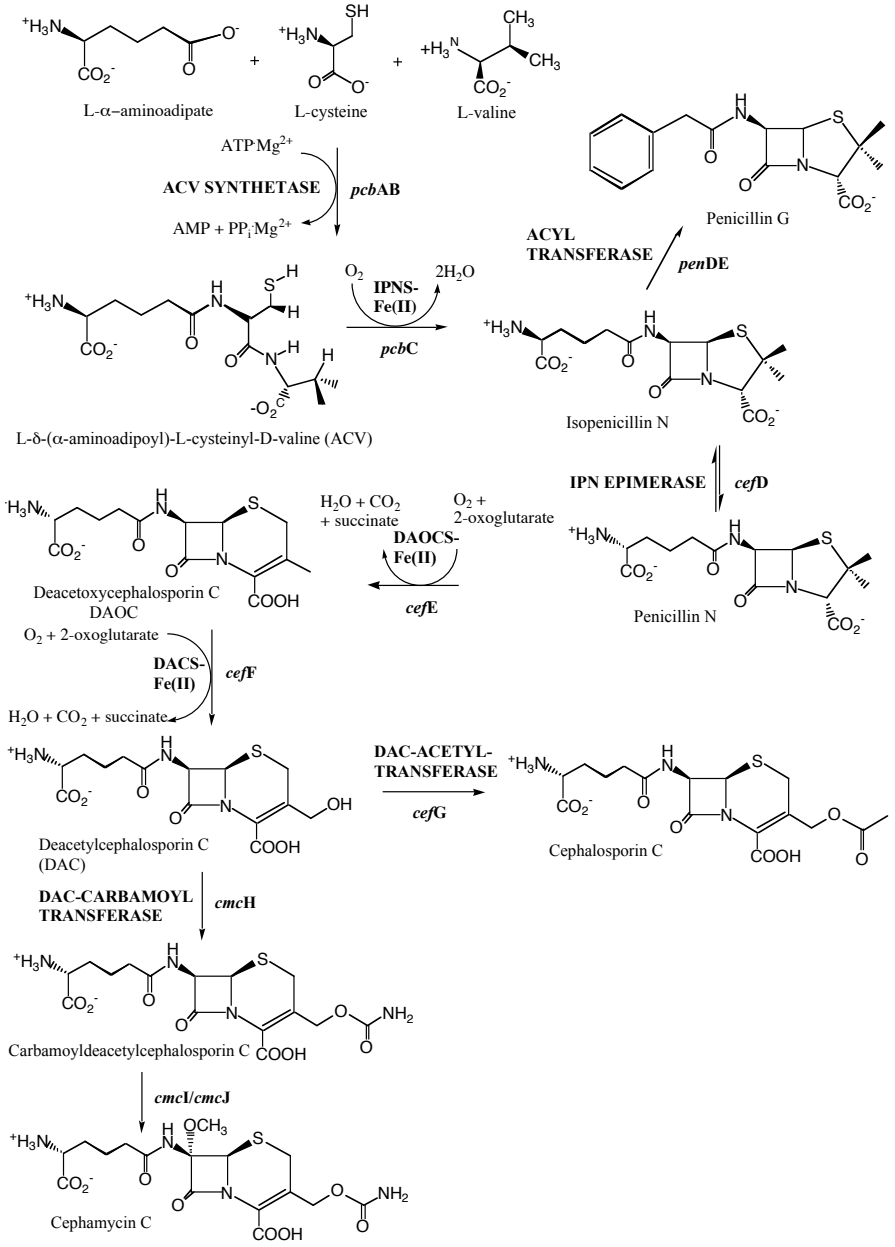

Figure 1.9 The biosynthetic pathway of penicillin, cephalosporin and cephamycin.

In penicillin-producing fungi that do not synthesize cephalosporins (i.e *Penicillium chrysogenum*), the final reaction is catalyzed by the acyl coenzyme A:isopenicillin N acyltransferase (IAT) (Brakhage, 1998). This enzyme has not been found in bacteria. In this step the hydrophilic L-α-AAA side-chain is exchanged for a hydrophobic acyl group, e.g. phenylacetyl in penicillin G. IAT shows broad substrate specificity (Luengo, 1995) and during fermentation the synthesis can be directed toward a specific penicillin by the addition of appropriate precursor molecules.

In cephalosporin producing organisms, the five-membered thiazolidine ring of penicillin N is expanded to the six-membered dihydrothiazine ring characteristic of all cephalosporin antibiotics. In eukaryotic cephalosporin producers (e.g. *Cephalosporium acremonium*), the ring expansion to deacetoxycephalosporin C (DAOC) and a subsequent hydroxylation at C-3' to form deacetylcephalosporin C (DAC) are catalysed by the bifunctional enzyme deacetoxy/deacetylcephalosporin synthase (DAOC/DACS). However, in prokaryotic cephalosporin producers (e.g. *Streptomyces clavuligerus*) the ring expansion and hydroxylation reactions are catalysed by two separate enzymes (Jensen *et al*., 1985), deacetoxycephalosporin C synthase (DAOCS) and deacetylcephalosporin C synthase (DACS). DAOCS, DACS and DAOC/DACS are all dioxygenases requiring Fe(II), 2-oxoglutarate and dioxygen to remove two electrons from the substrate, two from the co-substrate to form succinate, CO_2 and water. All three enzymes show a high sequence homology (DAOCS and DACS, 59%; DAOCS and DAOC/DACS, 57%; DACS and DAOC/DACS, 54%). The structure of DAOCS from *Streptomyces clavuligerus* (Valegård *et al.*, 1998) was solved in 1998 and will be further discussed later in this thesis.

In the final step in cephalosporin C formation an acetyl moiety from acetyl-CoA is transferred to the hydroxyl group of DAC. This step is catalyzed by acetyl-CoA:DAC acetyltransferase (Brakhage, 1998). DAC constitutes the second branch point in the biosynthesis, dividing the pathways for cephalosporin C and cephamycin C biosynthesis.

Cephamycin C biosynthesis has so far only been reported in bacteria such as *Streptomyces clavuligerus* and *Amacolatopsis lactamdurans*. In this pathway the C-3 acetoxy group of DAC is replaced by a carbamoyl group. This step is catalyzed by DAC-carbamoyl transferase and requires carbamoyl phophate, ATP and Mn^{2+}- and Mg^{2+}-ions (Brewer et al., 1980; Coque et al., 1995a). The biosynthetic routes leading to methoxylated cephalosporin (7-α-methoxy-cephalosporin) have not been fully elucidated and will be discussed further in chapter 5 of this thesis.

1.2.3 The biosynthesis of clavulanic acid

Clavulanic acid (Figure 1.5) is a β-lactam with weak antibacterial activity, but can act as a potent serine β-lactamase inhibitor (Baggaley *et al.*, 1997) and is administrated in combination with penicillins in the clinic. Chemical synthesis of clavulanic acid is even more complicated than the synthesis of penicillins and cephalosporins and the industrial production therefore involves fermentation of *Streptomyces clavuligerus*.

Biosynthesis of clavulanic acid and other clavams have only been reported in species of *Streptomyces* (Baggaley *et al.*, 1997). Even though clavulanic acid shares the β-lactam ring with penicillins and cephalosporins it is synthesised via a different pathway and from different precursors. The clavulanic acid gene cluster in *Streptomyces clavuligerus* contains at least 18 genes (Ward & Hodgson, 1993; Hodgson *et al.*, 1995; Li *et al.*, 2000; Mellado *et al.*, 2002), lies adjacent to the cephamycin gene cluster (Ward & Hodgson, 1993), and has not yet been fully elucidated.

The pathway (Figure 1.10) starts with a condensation of the precursors L-arginine and D-glyceraldehyde 3-phosphate to give N^2-(2-carboxyethyl)arginine catalysed by N^2-(2-carboxyethyl)arginine synthase (CEAS) (Khaleeli *et al.*, 1999; Caines *et al.*, 2004). The next enzyme, β-lactam synthetase (β-LS) (Bachmann *et al.*, 1998), catalyses the formation of the β-lactam ring via a mechanism completely distinct from the IPNS-catalysed formation of the β-lactam ring of penicillins and cephalosporins. β-LS shows the highest sequence and structural homology to class B asparagine synthetases (Miller *et al.*, 2001). The next enzyme in the pathway, clavaminate synthase (CAS), a Fe(II)-2-oxoglutarate dependent dioxygenase with homology to IPNS and DAOCS, catalyses 3 distinct oxygenase reactions in the biosynthesis of clavulanic acid. In the first CAS catalysed reaction, the monocyclic β-lactam is hydroxylated to give a secondary alcohol: guanidinoproclavaminic acid (Baldwin *et al.*, 1993) (that is neither a substrate nor an inhibitor of CAS). The following step is catalysed by proclavaminic acid amidino hydrolase (PAH), which hydrolyses the guanidino group to give proclavaminic acid (Elkins et al., 2002a). Next, CAS mediates the formation of the bicyclic nucleus of clavulanic acid in a two-step reaction involving oxidative cyclization, followed by desaturation to form clavaminate (Baldwin *et al.*, 1991; Salowe *et al.*, 1991). The structure of CAS has been elucidated (Zhang *et al.*, 2000) and is similar to that of DAOCS and IPNS. So far, CAS is the only enzyme identified in the clavulanic acid pathway that bears similarities to enzymes in the penicillin/cephalosporin pathway.

The pathway beyond clavaminate is not well characterized, and the only other known intermediate between clavaminate and clavulanic acid is clavaldehyde which is reduced in an NADPH-dependent reaction to clavulanic acid by the action of clavulanic acid dehydrogenase (CAD) (Nicholson *et al.*, 1994). The mechanism by which clavaminate undergoes stereochemical inversion and side chain modification to form clavaldehyde is unknown. There are yet many open reading frames (ORF) in the clavulanic acid biosynthetic gene cluster with unassigned

function. Biochemical and structural characterisation of these ORFs will be crucial in understanding the biosynthesis of clavulanic acid.

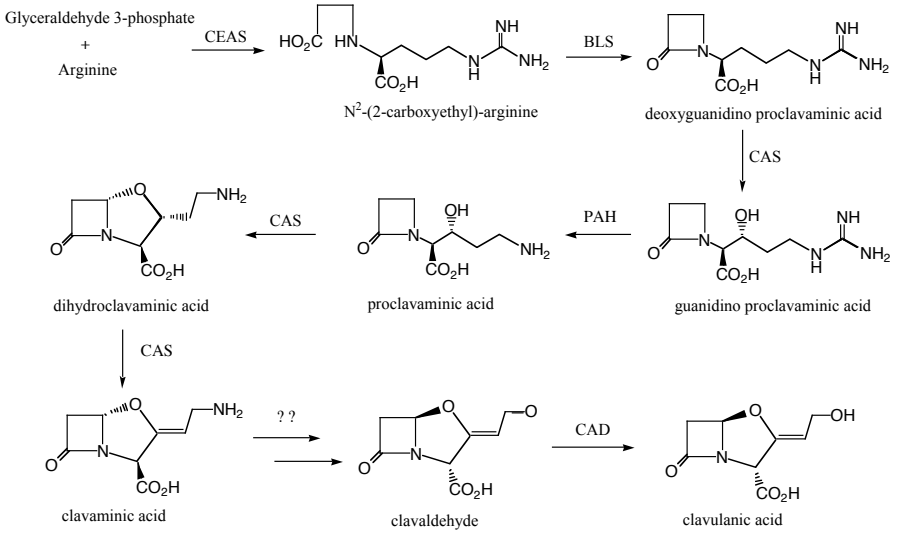

Figure 1.10 Biosynthetic pathway leading to clavulanic acid.

1.2.4 Non-heme Fe(II) oxygenases

An observant reader might have noticed the somewhat over-representative number of non-heme iron oxygenases involved in the biosynthesis of β-lactam antibiotics. From penicillin, cephalosporin and cephamycin biosynthesis, IPNS, DAOCS, DACS, DAOC/DACS all belong to this group together with possibly yet another enzyme in the late steps of cephamycin biosynthesis (discussed in chapter 5). For clavulanic acid biosynthesis, the enzyme CAS that catalyses three distinct steps in the synthesis also belongs to this family. All of these except IPNS require 2-oxoglutarate as a cofactor and are named 2-oxoglutarate dependent dioxygenases (2ODDs).

This group of enzymes catalyse a variety of important oxidative reactions in many organisms ranging from prokaryotes to eukaryotes. Many of the enzymes function in pathways that have medical, pharmaceutical or agricultural importance, and several of them have been implicated in disease states (Hausinger, 2004). Several structures of non-heme Fe(II) oxygenases have been solved and will be further discussed later in the thesis. The enzymes share a jelly roll fold with the iron ligated by a 2-His-1-Asp/Glu motif (the 2-His-1-carboxylate facial triad motif) (Schofield & Zhang, 1999). Even if the crystal structures together with spectroscopic studies have increased the knowledge of the mechanism of these enzymes significantly, there is still much left to learn, in particular regarding the

reaction with oxygen. The enzymes are thought to function through a highly oxidising iron(IV) species, the ferryl, but as yet the only evidence for the ferryl is through detection by spectroscopy of a ferryl intermediate in the reaction catalysed by taurine/α-ketoglutarate dioxygenase, TauD (Proshlyakov *et al.*, 2004). A more complete understanding of the catalysis performed by these versatile mononuclear ferrous enzymes would be significant in scientific, medicinal and industrial terms.

1.3 Industrial production of β-lactam antibiotics

Although most β-lactam antibiotics in clinical use can be synthesized chemically at low yields, this procedure is far too expensive and complicated for commercial use. Instead, fermentation and semi-synthetic modifications are the preferred ways to obtain β-lactams on a large scale. Today, fermentation is a very modern procedure giving high yields to a moderate cost (Elander, 2003), in particular with penicillins. However, there is still plenty of room for optimisations and developments, not at least in an environmental perspective since the procedure is very dirty and creates lots of waste products. With the increasing resistance against β-lactams there is also a great need to develop new antibiotics.

1.3.1 Penicillins

Penicillin V and penicillin G are the only natural penicillins in clinical use. They are obtained by varying the precursors added to the fermentation broth of *Penicillium chrysogenum*; phenylacetic acid for penicillin G or phenoxyacetic acid for penicillin V. However, the majority of penicillins used today are obtained by semi-synthetic modifications of the penicillin core, 6-aminopenicillanic acid (6-APA), Figure 1.1. 6-APA used to be obtained by a chemical splitting technique of fermented penicillin but due to environmental pollution and high costs the preferred method today is enzymatic conversion using penicillin G- or V acylase (Arroyo *et al.*, 2002). Recent structural and site-directed mutational studies on different β-lactam acylases have lead to better yields and also to the possibilities to alter the substrates and use the acylases to synthesise semi-synthetic β-lactam antibiotics (Sio & Quax, 2004).

1.3.2 Cephalosporins and cephamycins

Commercial cephalosporins, the most widely used antibiotics today, are all semi-synthetic derivatives. Unfortunately, fermentation to obtain cephalosporins gives comparably low yields and is expensive. High-yielding strains of *Acremonium chrysogenum* are used in large-scale fermentations to obtain cephalosporin C, in this case a big problem is the chemical instability of the molecule leading to degradation during the fermentation. Different purifications to obtain cephalosporin C are available, which is then converted to 7-aminocephalosporanic acid (7-ACA) (Figure 1.2), which is used as a starting material for two-thirds of

the commercial cephalosporins. 7-ACA can be obtained by either chemical or enzymatic deacylation, the latter is the common method today (Elander, 2003) since it is more environmentally friendly and cost-effective. Two genetically engineered enzymes are involved: D-amino acid oxidase (DAO) converts the α-aminoadipoyl side-chain to glutaryl-7-ACA, and glutaryl acylase (GLA) removes the glutaryl side-chain.

The remaining one-third of commercial cephalosporins are derived from 7-aminodeacetoxycephalosporanic acid (7-ADCA). Until recently, the preferred method to obtain 7-ADCA was chemical expansion of penicillin G to phenylacetyl-7-ADCA followed by enzymatic conversion to 7-ADCA catalysed by penicillin acylase. A more environmentally friendly procedure has been developed (Velasco *et al.*, 2000), in which the *cefEF* gene in *Acremonium chrysogenum* is replaced with the *cefE* gene of *S. clavuligerus*. This produces high titres of deacetoxycephalosporin C (DAOC), which is subsequently enzymatically converted to 7-ADCA by D-amino acid oxidase and glutaryl acylase.

Cephamycins are derived in a similar way as cephalosporins but give even lower fermentation yields. There is therefore great potential in further developing the production of cephamycins.

1.3.3 Developments and future possibilities

Traditionally, β-lactam production has been improved by classical strain improvement techniques, usually involving an iterative process where the production strains are exposed to random mutagenesis followed by analysis and selection of superior production strains. Developments in molecular biology enabled the introduction of directed genetic modifications through recombinant DNA technology, usually referred to as metabolic engineering (Thykaer & Nielsen, 2003). Different approaches are gene deletion, gene over-expression or heterologous expression of new genes and there are many successful examples in the production of β-lactams.

For example, *Penicillium chrysogenum* strains have a greater biosynthetic capacity than cephalosporin producing strains and by moving the *cefE* gene (deacetoxycephalosporin C synthase) from *Streptomyces clavuligerus* to *P. chrysogenum,* adipyl-7-aminodeacetoxycephalosporanic acid (adipyl-7-ADCA) is produced in the presence of adipic acid (Crawford *et al.*, 1995). From adipyl-7-ADCA different cephalosporin derivatives can be obtained. A lot of effort has also been invested in improving the rate-limiting enzymes in the production of β-lactam antibiotics. This is dependent on the strain used to synthesize the antibiotic, but generally ACVS has been identified as the limiting enzyme in fungi and actinomycetes. By cloning the ACVS gene (*pcbAB*) or replacing its weak promoter with a strong one, penicillin production in *A. nidulans* has been shown to increase dramatically (Kennedy & Turner, 1996). Also, cloning of the expandase/hydroxylase gene (*cefEF*) in an industrial strain of *C. acremonium* was shown to increase cephalosporin C production (Skatrud *et al.*, 1989). Further

examples where cloning of one or several genes have improved the β-lactam production significantly have been reported. To obtain higher yields in fermentations, further developments of this type will be required for industrial production in the future.

An even more challenging and rewarding task would be to alter the biosynthetic enzymes genetically to modify their substrate and product specificity in order to make new variants of β-lactam antibiotics not possible today. Using enzymatic methods would also constitute a more environmentally friendly and clean process compared with the chemical alterations that are common today. Different approaches could be envisaged. For example, IPNS could be altered to accept modified tripeptides as substrates, and this could lead to new biosynthetic routes to penems and other bicyclic β-lactams. Cephalosporins with hydrophobic side-chains would be desirable products. The natural substrate of DAOCS is penicillin N, that has a polar side-chain and therefore yields cephalosporins with a polar side-chain. A cephalosporin with a hydrophobic side-chain could be readily purified compared with the corresponding compounds with polar side-chains. By engineering DAOCS to produce cephalosporins of choice without the need for synthetic modifications, production costs and toxic by-products could be reduced. DAOCS has been shown to have a broad substrate specificity (Lloyd *et al.*, 1999; Dubus *et al.*, 2001) and enzymatic expansion of substrates with hydrophobic side-chains has been demonstrated with the wild-type enzyme under certain conditions (Cho *et al.*, 1998; Fernandez *et al.*, 1999; Demain & Baez-Vasquez, 2000). However, a genetic redesign would allow a more efficient conversion and could be used in industry.

These are just a few examples of enzymes that could potentially be modified, there are many others possible and different approaches could be taken to obtain new β-lactam antibiotics. With increasing knowledge about the enzymes involved in the biosynthesis obtained from a combination of structural, mutagenesis and activity studies this could become a very powerful tool (Barends *et al.*, 2004). Different approaches are (i) targeted random mutagenesis (ii) site-directed mutagenesis (iii) directed evolution, e.g. domain shuffling involving several genes (Crameri *et al.*, 1998; Arnold, 2001).

2. The aim of the thesis

The aim of this thesis is to elucidate the mechanism of action of enzymes in the biosynthetic pathways leading to β-lactam antibiotics. The knowledge of the biosynthesis of the β-lactams has increased a great deal during the last years due to e.g. sequencing of the gene clusters and functional assignment of almost all genes involved. However, this has led to further important questions that require careful investigations of the function and mechanism of the biosynthetic enzymes.

We employ a structural approach to study these enzymes using X-ray crystallography as the main method. A three-dimensional structure of a protein is a crucial ingredient to obtain detailed knowledge at the molecular level of the protein in question. In addition, to be able to propose mechanisms and elucidate protein interactions with cofactors and substrates/products, protein structures with bound ligands are invaluable.

Apart from being important basic research, the knowledge from this study could be used to produce new β-lactam antibiotics not available today and also improve the existing production with higher yields and more environmentally friendly procedures. Furthermore, an increased chemical knowledge of the versatile group of non-heme iron dioxygenases is significant in a broad scientific context and could be beneficial in both medicine and industry.

3. Deacetoxycephalosporin C Synthase, DAOCS (Paper I and II)

3.1 Background

Deacetoxycephalosporin C Synthase (DAOCS) from *Streptomyces clavuligerus* catalyzes the oxidative ring-expansion of the five-membered thiazolidine ring of the penicillin nucleus into the six-membered dihydrothiazine ring of cephalosporins (Figure 3.1). The 311 amino acid long enzyme shows broad substrate specificity, and can expand the thiazolidine ring in various penicillins but at a significantly lower efficiency compared to the physiological substrate penicillin N. DAOCS belongs to the family of 2-oxoglutarate dependent dioxygenases (2ODDs) which perform a wide range of reactions and require Fe(II), 2-oxoglutarate, dioxygen and usually a reducing agent for activity *in vitro* (Prescott, 1993).

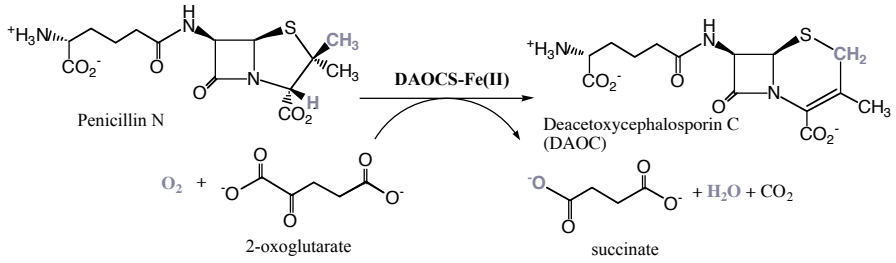

Figure 3.1 The reaction catalyzed by DAOCS.

The structure of DAOCS was the first from this family to be determined (Valegård *et al.*, 1998) and showed a similar fold as the closely related IPNS (Roach *et al.*, 1995). The structure is a distorted jelly roll fold with flanking helices, with the active site buried within the central core of the β-strands (Figure 3.2). Several structures of 2ODDs have since been determined and although the sequence similarity in the family is low, the structures share a common fold. These enzymes all utilize a conserved 2-His-1-carboxylate facial triad that binds the divalent iron (Hegg & Que, 1997).

DAOCS was determined to high resolution in complex with Fe(II) and the co-substrate 2-oxoglutarate and a mechanism for the initial steps leading to a ferryl intermediate was suggested (Valegård *et al.*, 1998). However, to reveal details of the later steps of the reaction leading to cephalosporin formation and to determine in detail penicillin/cephalosporin interactions with DAOCS in order to pinpoint which residues are important for catalysis and binding, further complexes were crucial. Such knowledge would be invaluable for future genetic redesign of DAOCS to accept and expand a wider range of penicillins not possible today, e.g. hydrophobic penicillins. Furthermore, at the start of this project, few details were known about the important group of 2ODDs and DAOCS was the only structure available at that time. To understand how these versatile enzymes use the highly

reactive iron in catalysis is of general importance and in particular we wanted to understand how the high oxidation-state ferryl intermediate is formed.

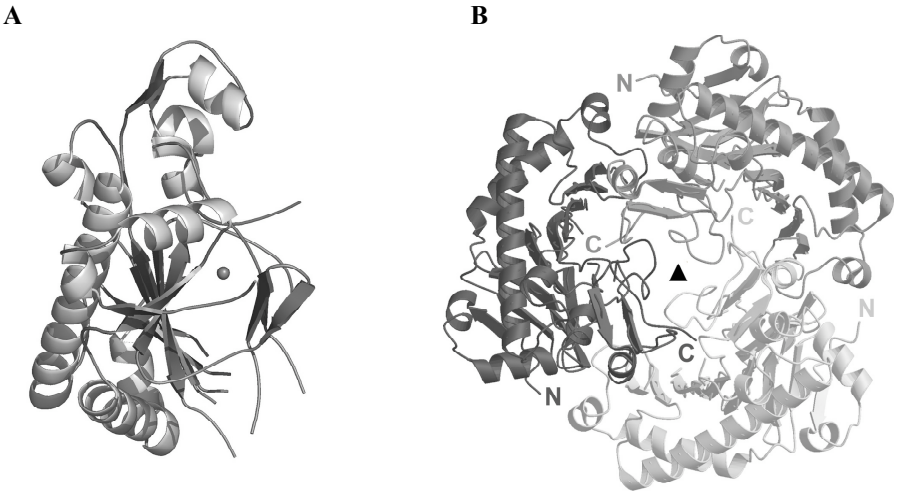

Figure 3.2 The crystal structure of DAOCS. (A) The monomer in complex with Fe(II), shown as a sphere. (B) The crystallographic trimer of DAOCS in the R3 crystal form viewed down the 3-fold axis. The positions of the N and C termini are indicated.

3.1.1 The need for a new crystal form of DAOCS

The structure of DAOCS was solved using crystals belonging to space group R3 (Valegård *et al.*, 1998). The packing of the individual molecules in this crystal form was identified as a possible complicating factor for obtaining complexes of DAOCS. In these crystals, residues 308-311 (the carboxy-terminal arm) of the apo enzyme participate in an intermolecular contact. The arm binds in the active site of a neighbouring molecule in a cyclic fashion, forming tight trimers (Figure 3.2). Dynamic laser light scattering on the apo enzyme in solution revealed an equilibrium between monomers and trimers, however in the presence of iron and 2-oxoglutarate the equilibrium was shifted towards the monomeric form (Lloyd *et al.*, 1999). This implies that it is the monomeric form of DAOCS that is catalytically active. It was therefore believed that the C-terminal arm could interfere with penicillin/cephalosporin binding.

A second complication is merohedral twinning in the R3 crystal form (Terwisscha van Scheltinga *et al.*, 2001). Merohedral twinning is a packing anomaly in which the distinct domains in the crystal superimpose completely in three dimensions. The twinning complicated the structural determination of DAOCS. It is also a complicating factor in the study of ligand complexes of DAOCS since detwinning the data amplifies experimental errors. The errors increase with increasing twin fraction and become infinite in a perfect twin. The twin fraction of the DAOCS R3 crystals varies from almost zero to almost perfect twinning (0.5). Merohedral twinning can only be detected by collecting and scaling a few degrees of data (10-

15°), therefore many crystals need to be prepared and screened to find crystals with acceptable twin fractions. In practice this means that to obtain reliable atomic data, only crystals with a twin fraction less than 0.3 are acceptable.

For these reasons a different crystal form of DAOCS was needed. Extensive crystallisation screening of native DAOCS was performed without any success. Therefore, modifications of DAOCS to break the trimer formation and abolish twinning were considered. An obvious approach is truncation of the C terminus. However, the C-terminus has been implicated to play an important role in the function of DAOCS and mutations and truncations of the C-terminus had been shown to yield kinetically altered or even inactive DAOCS (Lee *et al.*, 2001; Chin & Sim, 2002; Lee *et al.*, 2002). Alterations in this region were therefore avoided. Instead, a N-terminal 6x-histidine tag was added to DAOCS and a new crystal form was obtained. This crystal form will now be presented in some detail followed by a presentation of the complexes obtained by working with both crystal forms in parallel.

3.2 His-tagged DAOCS (Paper II)

3.2.1 Expression, purification and oligomeric state

The gene for DAOCS was re-cloned with an N-terminal 6x-histidine tag into vector pET-15b and over-expressed in *E. coli*. The His-tagged protein was separated on a nickel affinity column and further purified by gel-filtration. The tag greatly simplified the purification of DAOCS compared to previously described procedures (Lloyd *et al.*, 1999). Mononuclear ferrous enzymes are oxygen-sensitive and degrade under aerobic conditions (Barlow *et al.*, 1997). With the new construct, the time for purification was considerably shortened, which minimised exposure to oxygen and resulted in less degradation of the protein.

Using gel-filtration analysis and dynamic laser light-scattering measurements, His-tagged apo-DAOCS was shown to be monomeric in solution. In sharp contrast, the native apo-DAOCS exists as an equilibrium mixture of monomeric and trimeric forms in solution (Lloyd *et al.*, 1999).

3.2.2 Activity meassurements

A spectrophotometric assay has been developed by Dubus *et al.* (2001) to measure the activity of DAOCS. The assay is based on the continuous monitoring of absorbance at 260 nm due to the production of the deacetoxycephem ring. To measure absorbance at 260 nm involves a high background absorbance, caused by the assay ingredients. For instance, a high concentration of ascorbate (100 μM) that absorbs strongly at this wavelength is necessary. In the measurements on His-tagged DAOCS presented next, the concentration of penicillin had to be relatively

high due to a high K_m. This caused an even higher absorbance background and a 0.2 cm cuvette had to be used to measure the absorbance satisfactorily.

The assay was used to investigate if His-tagged DAOCS was active. The substrate used was penicillin G since the natural substrate penicillin N, is very labile and not commercially available. The result shows that DAOCS is active and can expand penicillin G to the corresponding cephem with an apparent K_m of 20.6 mM and a k_{cat} of 0.056 s^{-1} (Paper II). The K_m value is significantly increased in comparison to native DAOCS for which a value of 0.89 mM has been obtained (Dubus *et al.*, 2001). However, the k_{cat} value is in a similar range as the k_{cat} for native DAOCS: 0.079 s^{-1}. These results indicate that the His-tag affects substrate binding but not the rate of the turnover. Judged by inspection of the structure, where the N-terminus is located on the surface far away from the active site in both crystal forms, the effect of the His-tag on the kinetic properties is likely to be indirect.

3.2.3 Crystallisation

Crystals of His-tagged DAOCS were obtained after extensive screening in conditions modified to those used to obtain native DAOCS crystals. The new crystals were obtained by lowering the ammonium sulphate concentration and by reducing the crystallisation temperature from 20° C to 4° C. Further, no addition of 2-oxoglutarate was necessary to obtain crystals as was the case with the native protein. In addition, the crystals of His-tagged protein grew spontaneously without streak seeding, which was crucial to obtain native crystals. Another important factor was to slow down the rate of vapour diffusion by the use of Al's oil (Hampton Research) layered over the reservoir solution, to avoid the formation of microcrystals. The crystals are shaped like hexagonal rods instead of the rhombohedrally shaped native crystals (Figure 3.3).

A **B**

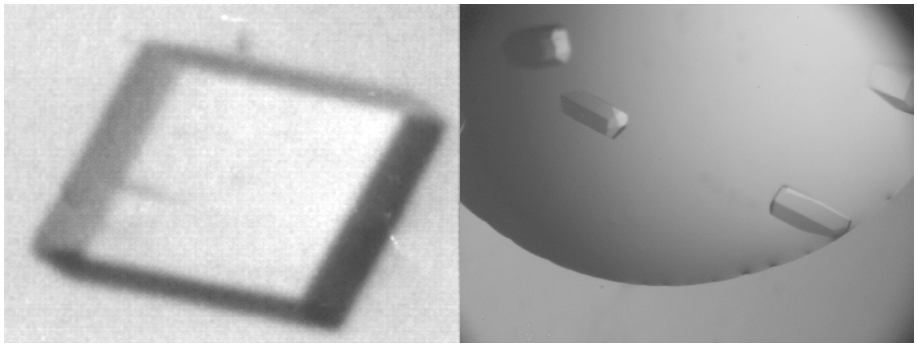

Figure 3.3 Different crystal forms of apo-DAOCS. (A) Native, space group R3 (crystal form I). (A) His-tagged, space group $P3_121$ (crystal form II).

3.2.4 The apo-structure of His-tagged DAOCS

The crystals of His-tagged DAOCS belong to space group $P3_121$ and are not twinned judged by their standard probability distribution of intensities (CCP4, 1994). The structure was solved by molecular replacement with the native apo-DAOCS as a search model. From now on, for simplicity, the R3 crystals will be named crystal form I, and the $P3_121$ crystals, crystal form II.

The packing of molecules in the two crystal forms differ considerably. In crystal form II, the trimer formation is broken and the C-terminus is not engaged in inter-molecular contacts (Figure 3.4), instead the last 12 residues are disordered. By inspection of the crystal packing it is easy to envisage why a modification at the N terminus would affect the interaction with the C terminus. In form I crystals (Figure 3.2), the N terminus is locked on the surface of the molecule and packs against two adjacent molecules of DAOCS (one within the same trimer, the other with a neighbouring trimer). Additional residues attached to the N terminus would disrupt these interactions and interrupt the trimer formation. This is what we observe with the His-tagged protein; in crystal form II both the N- and C-terminal tails are at the surface of the molecule, and not involved in any intermolecular contacts.

The packing of molecules in form II crystals is loose with big channels of solvent (Figure 3.4). Hence, the solvent content of crystal form II is much higher than in crystal form I: ~70% compared to ~40%. This allows substantial flexibility in the lattice and results in high Wilson B factor of the protein (62 $Å^2$). The high solvent content and loose packing result in fragile crystals that are difficult to handle and could also explain the difference in diffraction quality compared to crystal form I (2.3 Å compared to 1.3 Å).

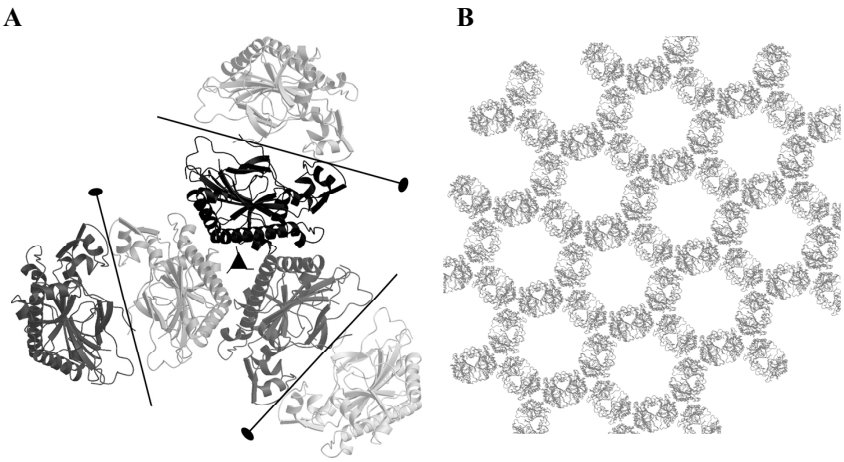

Figure 3.4 Packing of DAOCS in form II crystals. (A) Crystal form II viewed down the 3-fold screw axis. (B) Zoom-out to show the large solvent channels.

As expected, the overall folds of His-tagged and native DAOCS are very similar, (Figure 3.5). A superposition of the Cα atoms with a cut-off of 3.8 Å gave an r.m.s deviation of 1.1 Å for the two structures, with 20 residues not included in the superposition as they lie outside the cut-off limit. Most of these residues are found in the C-terminal region where the most significant differences between the two structures are observed. Analysis of the two structures with the program DYNDOM (Hayward & Berendsen, 1998) indicated the presence of two domains, comprising residue 1-267 and 268-299, respectively, where the C-terminal domain in crystal form II is rotated 16° towards the active site on a hinge consisting of Pro267 and Asn268. As mentioned above, the C-terminal arm (residue 308-311) is no longer fixed in the active site of the neighbouring molecule and there is no electron density for the last 12 residues of crystal form II, which have not been modelled.

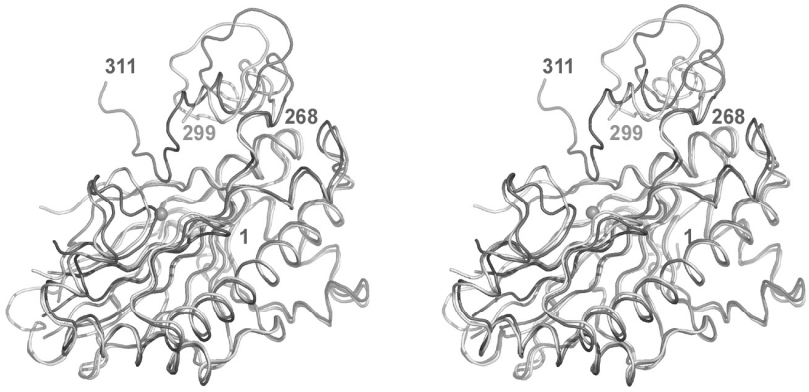

Figure 3.5 Stereographic superposition of the Cα traces of crystal form I (dark grey) and II (light grey) of DAOCS with the active site iron indicated as a sphere. The hinge residue (268) and the N- and C termini are indicated.

3.3 Ligand complexes with DAOCS

To obtain penicillin and cephalosporin complexes in crystals of DAOCS was recognised not to be a straightforward task due to a number of reasons. First oxygen is a substrate, therefore all complexes had to be prepared and frozen in an anaerobic box to avoid reaction and/or oxidation of the divalent iron. Furthermore, both DAOCS and the penicillin/cephalosporin molecules are sensitive and easily degrade or hydrolyse. The natural substrate, penicillin N, is very labile and not commercially available. DAOCS has been shown to expand a number of other penicillins but with a significantly higher K_m (Lloyd *et al.*, 1999), leading to a requirement for high soaking concentrations of these compounds. Like penicillin N, these penicillins are labile and easily hydrolysed. To dissolve and keep the penicillins/cephalosporins soluble at the required concentrations in the extreme conditions of high concentrations of iron and ammonium sulphate was not straightforward and was not possible for all substrates.

3.3.1 Complex with DAOCS-Fe(II)

Electron density maps obtained from $P3_121$ crystals of His-tagged DAOCS into which Fe(II) had been soaked featured strong density at the iron-binding site (Paper II). Moreover, some unexpected density was found at the iron *trans* to His243. This density was successfully modelled as ethylene glycol, which was used as cryo-protectant. One further ethylene glycol molecule was also found and modelled in the active site (Figure 3.6). The ordered binding of cryo-protectants in proteins is commonly observed (Garman & Schneider, 1997) and has also been observed for glycerol in crystal form I of DAOCS. For this reason, no cryo-protectant was used in further work with ligand complexes. The penicillin and cephalosporin molecules seem to partly function as cryo-protectants, since freezing of apo DAOCS without any cryo-protectant results in non-diffracting crystals. The binding of ethylene glycol and glycerol in the active site points to the capacity of DAOCS to bind small organic molecules containing carboxyl and carbonyl groups in the active site.

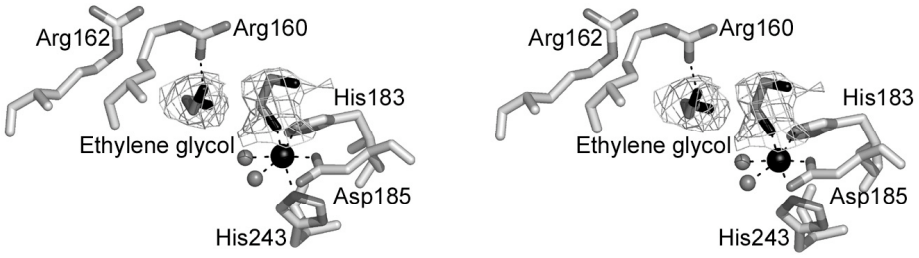

Figure 3.6 Stereoview of the active site of His-tagged DAOCS-Fe(II) at 2.5 Å resolution. The structure was obtained by soaking crystals in 20 mM $Fe(NH_4)_2(SO_4)_2$ followed by a quick soak in 30% ethylene glycol before flash freezing in liquid nitrogen. The $2mF_{obs}$-DF_{calc} electron density map is contoured at 1.0 σ where σ is the root mean square electron density for the unit cell.

3.3.2 Complex with DAOCS-Fe(II)-2-oxoglutarate

The structure solution of other 2ODDs has revealed two different binding modes of the co-substrate 2-oxoglutarate to the iron. One is represented by DAOCS and is also observed in anthocyanidin synthase, ANS (Wilmouth *et al.*, 2002). The other binding mode is represented by CAS (Zhang *et al.*, 2000) and is also found in TauD (Elkins et al., 2002b), alkylsulfatase, AtsK (Muller *et al.*, 2004) and factor inhibiting hypoxia-inducible factor, FIH (Elkins *et al.*, 2003). 2-Oxoglutarate binding has also been reported in carbapenem synthase, CarC, (Clifton *et al.*, 2003) but here 2-oxoglutarate binds in an intermediate position between the binding sites in DAOCS and CAS, but closer to the DAOCS binding. The relative positions of the 1-carboxylate of 2-oxoglutarate and a ligating water are transposed in the structures of CAS and DAOCS (Figure 3.7), and hence the proposed position of the ferryl will vary in the structures.

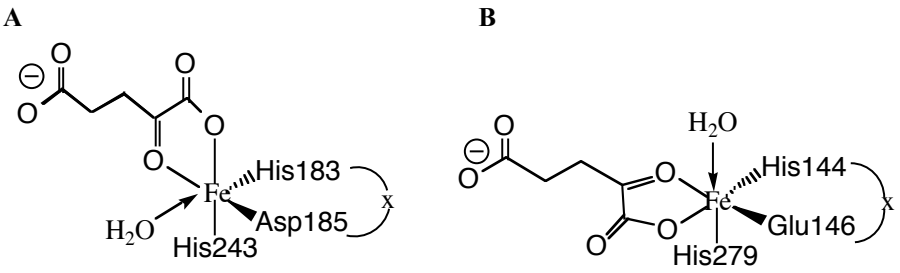

Figure 3.7 Comparison of the iron coordination of (A) DAOCS-Fe(II)-2-oxoglutarate and (B) CAS-Fe(II)-2-oxoglutarate.

For this reason we wanted to investigate the 2-oxoglutarate binding to DAOCS also in the $P3_121$ crystal form. This turned out not to be easy. Due to the presence of ethylene glycol in the active site in the DAOCS-Fe(II) structure the use of cryo-protectants that could bind in the active site was avoided. Data from crystals treated in different ways was collected: (i) without cryo-protectant, (ii) with 2-oxoglutarate as a 'cryo-protectant' at very high concentrations (1 M) and (iii) with oils as cryo-protectant. The data are not good in any of the cases. There are ice-rings that result in low completeness in some resolution shells, and the crystals do not diffract well enough to reveal details (~3.1-3.2 Å). However, inspection of the active site shows weak density at the corresponding 2-oxoglutarate binding site in crystal form I of DAOCS, but the structures were not possible to refine since the observation/parameter ratio was too low.

3.3.3 Complex with DAOCS-Fe(II)-succinate

A complex of DAOCS ligated by iron and the co-product succinate was obtained using the R3 crystal form (Paper I). Succinate ligates the iron in a monodentate manner at the position expected if it were produced by oxidative decarboxylation of the bound 2-oxoglutarate, following the previously published mechanism (Valegård *et al.*, 1998). Like 2-oxoglutarate, succinate forms a salt bridge to Arg258. The observed binding of succinate further supports the 2-oxoglutarate binding and the mechanistic proposal for the first half of the reaction (Valegård *et al.*, 1998).

3.3.4 Complex with DAOCS-Fe(II)-penicillin

Soaking form II crystals of DAOCS with iron, 2-oxoglutarate and ampicillin produced weak density at the iron that could be modelled and refined as ampicillin (Figure 3.8) (Paper II). There was no density for 2-oxoglutarate. The resolution, 2.7 Å, limits the details that can be interpreted from the structure. Penicillin complexes with crystal form I (Paper I) have been obtained under similar conditions, but to a significantly higher resolution: 1.5 and 1.7 Å. With crystal form I, a mixture of penicillin and 2-oxoglutarate were found to bind. The binding sites of the penicillins clearly overlap in the two crystal forms (Figure 3.8). To

obtain productive penicillin complexes of DAOCS it was crucial to soak the crystals in a combination of penicillin and 2-oxoglutarate or penicillin and succinate. Without the presence of the co-substrate or co-product a different and presumably unproductive binding mode of the penicillin molecule was observed (Paper I) (Figure 3.8).

In the productive binding in both crystal forms, penicillin directly ligates the iron with the sulphur atom of the thiazolidine ring *trans* to His243. This positions the β-methyl group to be inserted into the penicillin ring during catalysis in close proximity of the proposed ferryl position. This is in contrast to the unproductive penicillin binding where the C3 carboxylate on the thiazolidine ring ligates the iron and hence the β-methyl group is far away from the ferryl position (Figure 3.8), too far away for this binding to be productive. In the productive complexes, the penicillin side chains are in close vicinity to Arg160 and Arg162 that from mutational studies have been shown to be crucial for activity and proposed to bind the substrate (Chin *et al.*, 2001; Lipscomb *et al.*, 2002). In addition, the productive penicillin is bound in the same orientation as the penicillin substrate in the closely related IPNS (Burzlaff *et al.*, 1999) (Figure 3.8).

In both crystal forms of DAOCS, the binding site for penicillin in the productive complexes was shown to overlap with 2-oxoglutarate or succinate binding. This has mechanistic implications for catalysis by DAOCS, precluding simultaneous binding of penicillin and 2-oxoglutarate or succinate, which will be discussed below.

Even if the topological organisation of the penicillin ligands around the iron is found to be similar in the two crystal forms, there are also some clear differences observed (Figure 3.8). These are mainly caused by differences in crystal packing. The position of the penicillin side chain in crystal form II clashes with the position of residues 304-306 of the C-terminal arm in the form I crystals. The penicillin in crystal form I has to assume a different orientation and move towards Arg160, which results in the disorder of this residue. The penicillin in crystal form II interacts with both Arg160 and Arg162. The difference is also seen at the orientation of the penicillin nucleus, which in crystal form II is rotated by ~70° relative to its position in the form I crystals.

Figure 3.8 (A) Stereoview of the active site of His-tagged DAOCS-Fe(II)-ampicillin at 2.7 Å resolution. The structure was obtained by soaking crystals in 20 mM $Fe(NH_4)_2(SO_4)_2$, 33 mM 2-oxoglutarate and 88 mM ampicillin. The $2mF_{obs}$-DF_{calc} electron density map is contoured at 1.0 σ where σ is the root mean square electron density for the unit cell. (B) Productive penicillin binding in crystal form II. (C) Penicillin binding in IPNS, pdb-id 1QJE. (D) Unproductive penicillin binding in crystal form I. (E) Superposition of the active sites in Form II DAOCS-Fe(II)-ampicillin (coloured orange, ampicillin coloured yellow) and Form I DAOCS-Fe(II)-ampicillin (coloured light blue, ampicillin coloured dark blue).

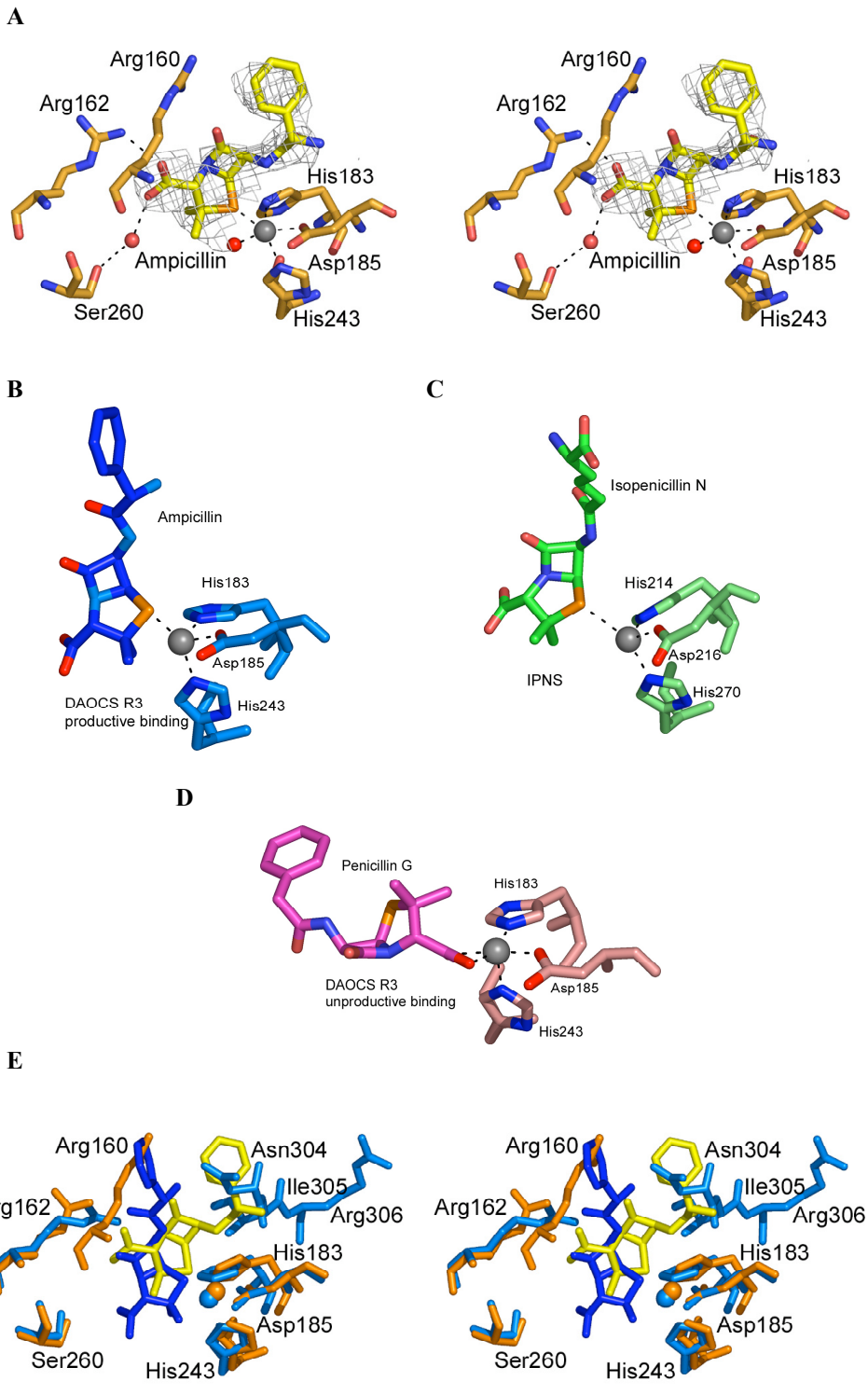

Figure 3.8 (A)-(E) Legend on previous page.

3.3.5 Complex of DAOCS-Fe(II)-cephalosporin

Soaking of crystal form II with iron and the natural product deacetoxycephalosporin C produced weakly diffracting crystals and diffraction images with ice-rings. This was probably due to the absence of cryo-protectant. Extensive screening was performed and eventually a suitable crystal was found that yielded a data set to 3.0 Å. Weak density around the iron was modelled and refined as DAOC (Figure 3.9) (Paper II). Soaking of crystal form I with iron, succinate and deacetoxycephalosporin C produed a DAOCS-Fe(II)-DAOC complex to 1.7 Å (Paper I). There is no density for succinate.

In both crystal forms, the cephalosporin molecules bind in the same orientation as the productively bound penicillins and partly overlap with each other (Figure 3.9). DAOC in both structures is shown to overlap with 2-oxoglutarate and succinate. In the form I structure, the sulphur atom of the dihydrothiazine ring directly ligates to the iron whereas in the form II structure the dihydrothiazine ring is further away from the iron and the methyl group of the dihydrothiazine ring points toward the iron.

Like the penicillin molecules, the cephalosporin molecules in the two crystal forms take up slightly different orientations (Figure 3.9), with the cephalosporin molecule of crystal form II positioned further away from the iron. The cephalosporin binding site of crystal form II does not overlap with the position of the C-terminus of the same molecule from crystal form I, but comes very close to the position of the C-terminus of yet another molecule within the trimer of crystal form I. The differences in cephalosporin binding observed between the two crystal forms are similar to the differences in penicillin binding and are probably an effect of the crystal packing.

Figure 3.9 (A) Stereoview of the active site of his-tagged DAOCS-Fe(II)-DAOC at 3.0 Å resolution. The structure was obtained by soaking crystals in 20 mM $Fe(NH_4)_2(SO_4)_2$, and 150 mM deacetoxycephalosporin C. The $2mF_{obs}$-DF_{calc} electron density map is contoured at 0.9σ where σ is the root mean square electron density for the unit cell. (B) Superposition of the active sites in form II DAOCS-Fe(II)-DAOC (coloured orange, DAOC coloured yellow) and form I DAOCS-Fe(II)-DAOC (coloured light blue, DAOC coloured dark blue, C-terminus from symmetry-related molecule coloured cyan).

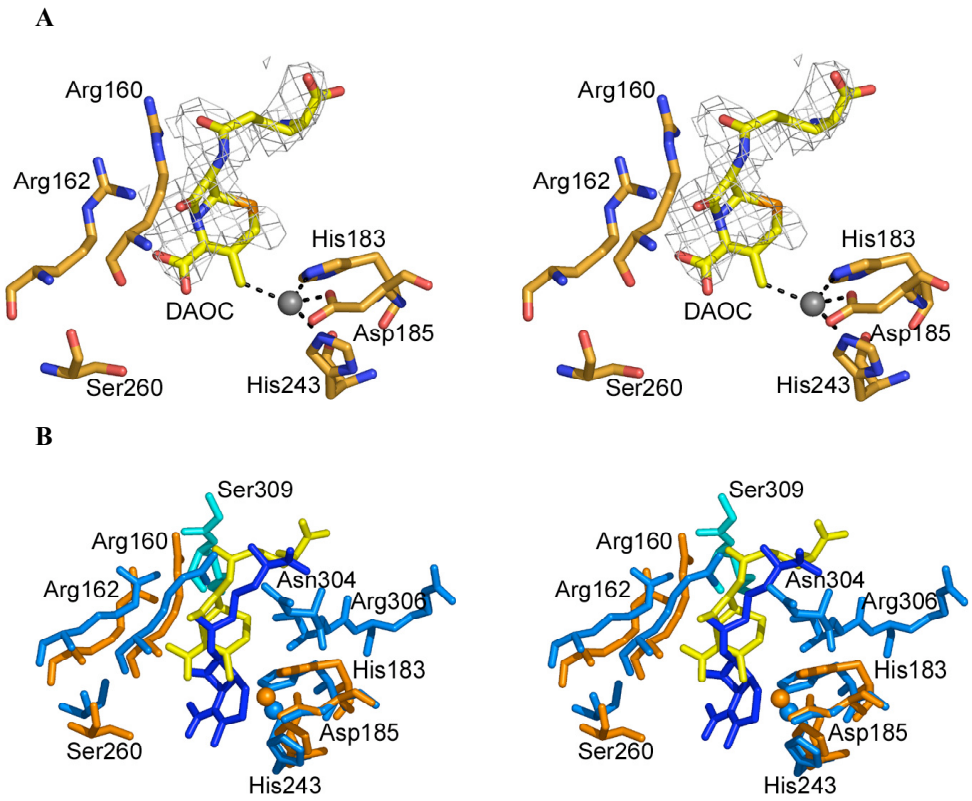

Figure 3.9 (A)-(B) Legend on previous page.

3.4 The active site and the role of the C terminus

Substantial differences in the C terminal domain of DAOCS were observed between the two crystal forms, starting at the hinge between residues 267 and 268. The differences suggest that the C-terminal domain is flexible and can take up different conformations.

Further differences between the two crystal forms were observed upon binding penicillin and cephalosporin molecules in the active site. The differences are caused by the crystal packing where the C terminus in crystal form I interferes with penicillin and cephalosporin binding. Therefore one can speculate that crystal form II probably more reflects the active enzyme in solution, since no adjustments due to the presence of the C terminus are necessary when binding substrate/product.

Inspection of the binding of penicillin and cephalosporin reveal surprisingly few interactions of DAOCS to the side-chains of the penicillin/cephalosporin molecules, except from Arg160 and Arg162. Site-directed mutagenesis studies combined with activity assays have demonstrated the importance of the C-terminal

arm for activity, where some mutations even were shown to increase the activity for hydrophobic penicillins (Chin *et al.*, 2001; Chin & Sim, 2002). Residues can be added without loss of activity but the successive shortening of the C-terminal arm diminishes catalytic activity (Lee *et al.*, 2001). The C terminus has also been shown to be important in the closely related enzymes IPNS (Sami *et al.*, 1997) and DAOC/DACS (Lloyd *et al.*, 2004).

Based on the lack of direct interaction of the penicillin/cephalosporin molecule with amino acids of the enzyme and the well-established importance of the C terminus, it seems reasonable that the C terminus of DAOCS plays an important role in binding and orienting the substrate. In crystal form I the C terminus is locked in the tight trimer interactions but why any interaction of the C terminus is not seen in crystal form II one can only speculate. It is possible that for this to occur substantial rearrangements are necessary of the whole C terminal domain, starting at the hinge at residue 268. Due to crystal packing this might not be possible in crystal form II, even if the packing is relatively loose in the crystal. Unsuccessful attempts have been made to co-crystallise DAOCS in the presence of penicillin.

Regarding the function of the trimers, one can speculate that DAOCS in the unactive apo-form protects the flexible C terminus from proteolytic degradation by forming tight trimers. In the presence of iron and 2-oxoglutarate the trimer dissociates and the C terminus guides and binds the penicillin substrate.

It is not directly obvious from the structures which residues should be altered or the nature of the change to obtain a more efficient DAOCS and/or a DAOCS with altered substrate specificity. The interactions from Arg160 and Arg162 to the penicillin/cephalosporin involve the ring-core structure or the carbonyl group on the side chain of the penicillin/cephalosporin. These interactions are also important for hydrophobic penicillins and therefore should not be modified in order to modify the side-chain preference of DAOCS. There are no direct interactions from DAOCS with the tail of the side-chain of the penicillins, it is likely that the C terminus will interact with this part. Therefore the C terminus of DAOCS may be a prime target for mutagenesis, which has already been shown in several examples (Lee *et al.*, 2001; Wei *et al.*, 2003; Chin *et al.*, 2004). Furthermore, the substrate binding pocket of DAOCS seems relatively wide and flexible which might be advantageous since it could be fairly easy to alter the substrate specificity. In fact, in Paper II speculations that DAOCS might be capable of other catalytic activities are made.

3.5 The mechanism of DAOCS

All the crystal complexes obtained with DAOCS indicate that the binding of the co-substrate/substrate and co-product/product in the active site overlap and that these molecules therefore cannot bind at the same time. This has mechanistic implications for catalysis by DAOCS.

Substrate interference of this type should affect the kinetic behaviour of DAOCS in solution. Indeed this was shown using steady-state rate measurements of native DAOCS where one substrate concentration was varied while the other was fixed (Paper I). The results show a major decrease in the steady-state rates at fixed penicillin concentrations (using both penicillin G and ampicillin) at increasing concentrations of 2-oxogutarate. The same behaviour was also observed at fixed 2-oxoglutarate concentrations with increasing concentrations of penicillin, albeit to a less extent. This behaviour is expected when two ligands compete for the same binding site on the enzyme and is in support of our structural results. A commonly observed feature in many 2-oxoglutarate dioxygenases including DAOCS is uncoupled turnover, where 2-oxoglutarate is converted to succinate and carbon dioxide without concomitant oxidation of 'prime' substrate (Prescott & Lloyd, 2000). The rate of uncoupled turnover increases with increasing concentration of 2-oxoglutarate, consistent with our steady-state results.

The difference observed in penicillin binding in the absence or presence of 2-oxoglutarate or succinate suggests that the active site is incomplete for productive binding without the co-substrate or co-product. The incoming penicillin must encounter an active site with altered geometry and charge state when the co-substrate or co-product is present and may bind differently than in the unliganded enzyme.

From our structures of DAOCS a mechanism for cephalosporin formation was proposed (Figure 3.10). The reaction sequence starts with binding of 2-oxoglutarate, which activates the iron for oxygen binding. With bound oxygen, oxidative decarboxylation of 2-oxoglutarate occurs and results in the formation of an oxidizing intermediate plus succinate and carbon dioxide as by-products. Carbon dioxide is a weak ligand to the iron and is likely to leave, whereas succinate remains bound and stabilizes the oxidizing iron species. Quantum molecular dynamic calculations of different iron-oxygen adducts by Graziella Ranghino identified the planar peroxo as more stable than the ferryl species (Paper I). The oxidizing iron species may thus be stored transiently in the peroxo form within the protein to await a reaction with the penicillin substrate. When penicillin expels succinate, it triggers oxidative attack on itself. The binding of penicillin brings the β-methyl group of the penicillin core within reacting distance of the expected ferryl oxygen. Ring expansion, involving radical formation and transfer of two electrons and two protons to the oxygen, results in the formation of the cephalosporin product and water.

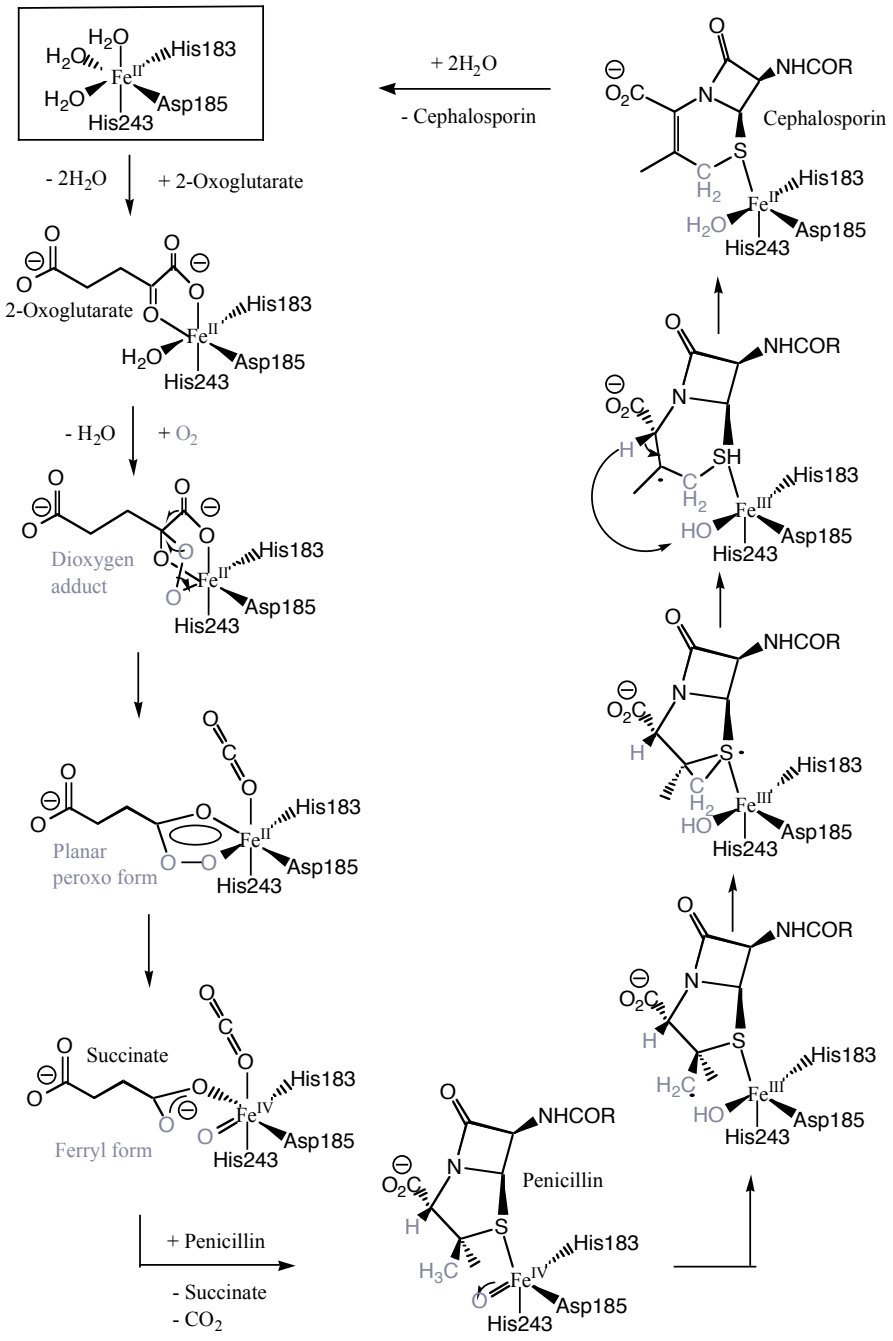

Figure 3.10 Proposed mechanism for the ring expansion catalysed by DAOCS based on the mode of penicillin and cephalosporin binding in the crystal structures. The presumed oxidation states of the iron are marked.

3.6 Comparison of DAOCS to other 2ODDs

The overlapping binding of substrate and co-substrate in DAOCS proposed here differs to other 2-oxoglutarate dependent dioxygenases where structural information on substrate and co-substrate binding is available. In CAS (Zhang *et al.*, 2000), TauD (Elkins et al., 2002b), AtsK (Muller *et al.*, 2004), ANS (Wilmouth *et al.*, 2002), FIH (Elkins *et al.*, 2003) and CarC (Clifton *et al.*, 2003), 2-oxoglutarate and substrate bind at different positions simultaneously.

Furthermore, the proposed mechanism for DAOCS differs from the generally proposed mechanism of 2ODDs, reviewed in Schofield & Zhang, 1999. It is based on kinetic data (Holme, 1975; Kivirikko & Pihlajaniemi, 1998) that indicate an ordered sequential mechanism, with the binding of 2-oxoglutarate followed by that of the prime substrate and last dioxygen. Product release occurs in order: carbon dioxide followed by succinate or prime product, depending on their relative concentrations. This mechanism is supported by the structures of 2ODDs with non-overlapping substrate and co-substrate binding. Spectroscopic studies of CAS (Zhou *et al.*, 2001) and TauD (Price *et al.*, 2003) indicate that only when both substrate and co-substrate are present is the ferrous site converted into a five-coordinated species with enhanced reactivity toward dioxygen. Based on this, a mechanism was proposed where the binding of the main substrate triggers a rearrangement to a five-coordinated species that will bind oxygen, leading to oxidation of the substrate.

The position of the substrate in DAOCS also differs compared to other 2ODDs (CAS, CarC, ANS, AtsK, HIF and TauD). In DAOCS, the penicillin substrate is found to directly ligate the iron whereas in other 2ODDs, the substrate is positioned at a distance of more than 4 Å from the iron, albeit at the same side of the iron as in DAOCS. Since 2-oxoglutarate binding in CAS and others are different compared to DAOCS (Figure 3.7), the ferryl was proposed to bind *trans* to the position corresponding to His243 in DAOCS. In this case, the ferryl would be perfectly positioned to react with the bound substrates in these enzymes. However, an intriguing observation with CAS indicates that oxygen may be bound in several modes and limits the certainty the ferryl and 2-oxoglutarate positions in these enzymes can be determined. Upon exposure of CAS-Fe(II)-2-oxoglutarate-deoxyguanidinoproclavaminate crystals to nitric oxide (NO, dioxygen analogue), NO was found not to bind at the expected position but *trans* to the position corresponding to His183 in DAOCS. This could be interpreted that NO binding causes a rearrangement around the iron resulting in 2-oxoglutarate iron-ligation similar to DAOCS (Zhang *et al.*, 2002). If NO correctly mimics bound dioxygen in CAS, the ferryl would be incorrectly oriented to react with the bound substrate and the catalysis would have to involve a rearrangement to flip the ferryl to the correct position after CO_2 release. It is of course possible that the observed NO binding does not correspond to the correct dioxygen binding in CAS, but without further evidence this observation has to be taken into account to show that the mechanism of other 2ODDs is not completely elucidated.

As described, DAOCS seems to differ from other 2ODDs investigated so far in a number of ways. In fact, DAOCS seems to resemble IPNS more closely. The amino acid sequences of DAOCS and IPNS are significantly more homologous compared to DAOCS and other 2ODDs, excluding the closely related DACS and DAOCS/DACS. The high homology is also revealed in structural terms. A DALI search (Holm & Sander, 1993) of the DAOCS structure revealed IPNS as the clearly most similar structure, Table 3.1. Furthermore, the penicillin binding in IPNS and DAOCS is very similar, with the sulphur atoms ligating the iron at the same position and the same orientation of the penicillin side-chains (Figure 3.8). However, DAOCS and IPNS share their fold with other 2ODDs, which implies a divergent evolutionary relationship. Since the genes for DAOCS, DACS and IPNS are all positioned in the same gene cluster one can imagine that they have evolved from the same ancestor and, following gene duplication have taken up different functions.

Table 3.1 DALI search results for structures similar to DAOCS

PDB ID	Z-Score	r.m.s.d	RA	%SI	Description
1BK0	22.7	2.6	245	21	IPNS
1GP4	19.4	3.2	241	20	ANS
1E5R	7.1	3.9	143	13	Proline 3-hydroxylase
1DRT	6.3	3.9	148	9	CAS
1GQW	6.0	4.0	137	8	TauD
1NX4	5.4	3.2	123	10	CarC

PDB, Protein Data Bank; r.m.s.d, root-mean-square deviation; RA, number of residues aligned; &SI, percent sequence identity over the aligned fragments.

From this reasoning it may appear that DAOCS is an exception to the other 2ODDs. However, still very few enzymes in the diverse family of 2ODDs have been thoroughly investigated and it is very likely that also other 2ODDs function in a similar way as DAOCS. The differences point to the diversity of the family. Further studies on members in the family of 2ODDs using a combination of methods are necessary to yield more knowledge about these enzymes. In particular, knowledge about the high oxidation-state intermediates is crucial to understand their mechanism more completely.

4. ORF7 in the clavulanic acid biosynthesis

4.1 Background

The biosynthetic route of clavulanic acid in *S. clavuligerus*, presented in section 1.2.3, is not fully characterised. In particular, some of the late steps are yet uncharacterised. The identified gene cluster contains at least 18 genes and many of the identified open reading frames (ORFs) have not had a function assigned to them yet.

One of these genes is *orf7*, which was shown to be vital for clavulanic acid biosynthesis using a gene disruption mutation approach (Jensen *et al.*, 2000). The protein encoded by the *orf7* gene, ORF7, has an estimated molecular weight of 61 kDa. Sequence comparison using BLAST identified it to be most similar to bacterial oligopeptide-binding proteins. ORF7 has therefore been suggested to bind and/or transport clavulanic acid and/or precursors. Another gene, *orf15* encoding a putative oligopeptide-binding protein, ORF15, was identified further down in the clavulanic acid gene cluster, with 48% amino acid sequence similarity to ORF7. Both genes are required for clavulanic acid production and cannot replace each other (Lorenzana *et al.*, 2004). However, clavam intermediates are formed in the *orf15*-disrupted mutant cells.

4.2 Expression and purification of ORF7

ORF7 from *S. clavuligerus* was cloned to the pET-24a vector by collaborators in Oxford and sent to Uppsala. The plasmid was transformed into *E. coli* BL21 DE3 cells and the expression gave high yield of soluble protein.

The initial purification of ORF7 outlined in Oxford was further developed in Uppsala. ORF7 is expressed in native form and the purification involves many steps, but the whole procedure can be performed within 30 hours. The first step involves an ion-exchange column with Q-Sepharose fast flow media followed by a hydrophobic interaction purification step. This is followed by gelfiltration using a Superdex200 column and another ion-exchange column, MonoQ, as the last step.

The protein appears reasonably pure already after the hydrophobic interaction step and very little improvement in the purity is achieved in the following steps of the purification. ORF7 separates from gel filtration and MonoQ columns in sharp peaks without apparent overlap from other proteins. Nevertheless, SDS PAGE (Figure 4.1) and native gel analysis shows the protein is not pure. There are several major low molecular weight contaminant bands present, in addition to the 61 kDa band of ORF7. Speculations to the identity of these bands led to two main theories. One possibility is break-down products of ORF7. Proteolytic degradation and break-down of sensitive proteins during purification is a common problem in protein purification. In the purification of ORF7, normal precautions to avoid this have been taken: a tablet of a cocktail of protease inhibitors is added to the cells during the cell disruption and EDTA is present during the whole

purification. All purification steps are performed at 4°C and without delays in between. The other possible explanation for the impurities relates to the putative function of ORF7, as a homologue to oligopeptide-binding proteins. It is possible that ORF7 could bind different peptides/proteins present in the *E. coli* cells during the expression. Attempts were made to improve the purification of ORF7 by e.g. the use of other columns without any significant differences observed.

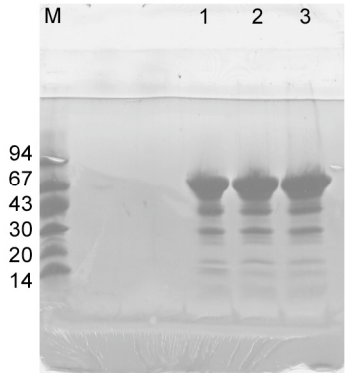

Figure 4.1 A 20% SDS PAGE gel of the purified ORF7: lanes 1, 2 and 3 show the same sample. Lane M is the low-molecular weight marker with the molecular weights indicated to the left in kDa.

4.3 Crystallisation of ORF7

Despite impurities, extensive crystallisation screening of ORF7 was performed using different methods, crystallisation screens and at different temperatures. The most promising conditions were found with the hanging drop method at 20°C with approximately 2 M ammonium sulphate as precipitating agent and in HEPES-buffer pH 7.5-8. These conditions gave needle-shaped crystals (Figure 4.2). Further screening to improve the needles was performed, many additives were tested, and also seeding was tried. These trials did not result in dramatic improvement – the needle-like crystals prevailed (Figure 4.2). The needles were tested for diffraction at the ESRF but no diffraction was observed. No salt diffraction was observed indicating that the crystals were indeed protein.

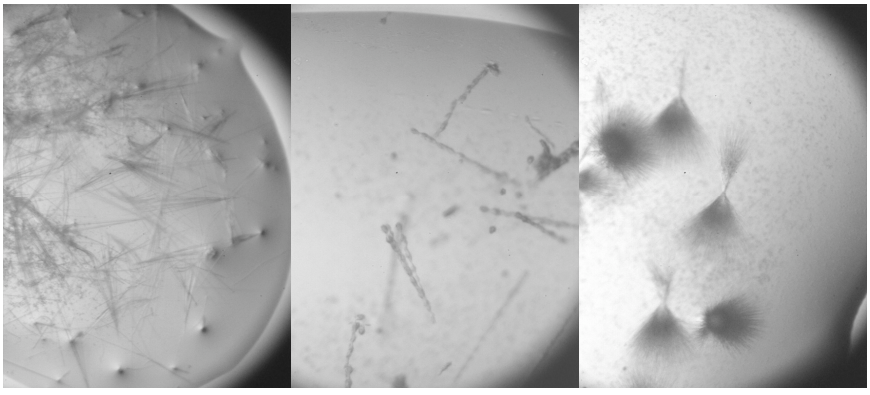

Figure 4.2 Different needle-shaped crystals of ORF7.

5. Cephamycin C biosynthesis: the structure and function of cmcI (Paper III and IV)

5.1 Background

Cephalosporins and cephamycins differ by a methoxyl group at the C-7α position of the β-lactam ring (Figure 5.1). This methoxyl group influences the antibiotic activity spectrum of the molecule and increases its resistance against β-lactamases. The main focus of this chapter is how this methoxyl group is incorporated into the cephalosporin molecule in *Streptomyces clavuligerus* and the role of the enzyme cmcI in this process.

Work on cell free extracts of *S. clavuligerus* has demonstrated that the methoxylation occurs in two steps, a hydroxylation followed by a methylation of the hydroxylated intermediate (O'Sullivan & Abraham, 1980; Hood *et al.*, 1983) (Figure 5.1). The conversion of cephalosporin to cephamycin requires the addition of Fe(II), 2-oxoglutarate, S-adenosyl-L-methionine (SAM) and a reducing agent (O'Sullivan & Abraham, 1980; Hood *et al.*, 1983). In the absence of SAM, the hydroxylated intermediate was formed (Hood *et al.*, 1983). Based on these results it was assumed that the hydroxylation reaction was catalysed by a 2ODD and the methylation by a SAM-dependent methyltransferase.

Except for the methoxylation, the biosynthesis of cephamycin C from deacetylcephalosporin C involves a carbamoylation at the C-3 position, catalysed by DAC-carbamoyl transferase (Coque *et al.,* 1995a). The order of carbamoylation and methoxylation has not been fully elucidated but it has been suggested that carbamoylation preceeds methoxylation (Baldwin & Schofield, 1992). The methoxylation has been shown to occur using both cephalosporin C and carbamoyldeacetylcephalosporin C as substrates (O'Sullivan & Abraham, 1980).

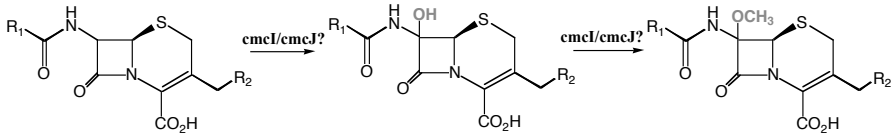

Figure 5.1 Biosynthesis of cephamycins from cephalosporins occurs via hydroxylation followed by methylation at the C-7α position.

Two genes, *cmcI* and *cmcJ*, in the cephamycin gene cluster of *Streptomyces clavuligerus* have been implicated in the methoxylation reaction (Alexander & Jensen, 1998). An initial search of these enzymes in protein databases and review articles identifies cmcI as the cephalosporin hydroxylase and cmcJ as the methyltransferase (Brakhage, 1998; Elander, 2003). However, more detailed analysis questions these assignments.

Sequence analysis shows that the 28 kDa gene product of c*mcI* has weak sequence characteristic of methyltransferases and monooxygenases, whereas the 32 kDa product of *cmcJ* shows weak sequence homology to a number of hypothetical proteins and a gibberellin desaturase with hydroxylase activity (Tudzynski *et al.*, 2003).

Contradictive results have been published on the enzymes involved in the methoxylation. The first report described the purification of a cephalosporin 7α-hydroxylase with an estimated molecular mass of 32 kDa from cells of *S. clavuligerus* (Xiao *et al.*, 1991). The enzyme catalysed the hydroxylase reaction with the cofactor requirements of a 2-oxoglutarate dependent dioxygenase. The first investigation on the genes *cmcI* and *cmcJ* came from Coque *et al.*, (1995). Cell-free extracts of *S. lividans* (not a cephamycin producer) transformed with the DNA fragments of *cmcI* and *cmcJ* from *A. lactamdurans* were shown to catalyse the methoxylation on addition of 2-oxoglutarate and SAM. *CmcI* alone showed weak hydroxylase activity in the presence of 2-oxoglutarate or NADH, whereas *cmcJ* showed no activity by itself. Further studies involving immunoaffinity, protein cross-linking and fluorescence spectroscopy (Enguita *et al.*, 1996), suggested that cmcI and cmcJ could form a complex able to perform both the hydroxylation and the methyl transfer reaction with the active site on cmcI, and cmcJ working as a helper protein.

Taking all of these facts together yield different theories on the function and mechanism of cmcI and cmcJ in the cephamycin biosynthesis. (i) CmcI is the hydroxylase, working as a 2ODD or using NADH and cmcJ is the methyltransferase. (ii) CmcI is the methyltransferase and cmcJ the hydroxylase. (iii) CmcI and cmcJ form a complex and perform both reactions together. Using a structural approach to answer these questions, the structure of cmcI from *S. clavuligerus* was determined and the results will be presented here. The work on cmcI can in some aspects be compared to structural genomics projects where structures of proteins with unknown functions are determined.

5.2 Expression, purification and crystallisation of cmcI (Paper III)

5.2.1 Expression and purification

The *cmcI* gene of *S. clavuligerus* was cloned into the pET32 vector. In this construct, cmcI is fused to thioredoxin, which in turn is labelled with a His-tag. The plasmid was transformed into *E. coli* and expressed at a high yield. The purification involved separation on a nickel affinity column, followed by removal of the thioredoxin tag by cleavage with factor Xa. The His-tagged thioredoxin was separated from cmcI using Ni-NTA beads, and the factor Xa was removed using benzamidine Sepharose beads. The remaining cmcI was further purified using a

gelfiltration column. This procedure yields very pure cmcI without any tags or extra residues attached.

The initial clone of cmcI (paper III) was discovered to have three differences at the amino acid level, L10Q, D160N and L200F, compared to the sequence in the Genbank (Alexander & Jensen, 1998). It was not immediately obvious whether these differences arose from PCR errors, sequencing errors, the presence of cmcI isoforms or different strain origins because at least one of the differences was encountered when cmcI was independently recloned. Further recloning indicated that the alterations at position 10 and 160 were mutations caused by PCR errors whereas no such indication was found at position 200. Multiple sequence alignments using Clustal-W (Paper IV) further indicated that residue 160 should be an aspartic acid, since it is highly conserved and the structure of cmcI revealed it to be located in the active site.

Therefore, a new expression plasmid was generated by site-directed mutagenesis for production of 'native' cmcI which differes from the Genbank sequence by a phenylalanine instead of leucine at position 200. This plasmid was used to produce protein for the majority of the structural studies and the product is referred to as **cmcI-D160**. The product of the original construct is referred to as **cmcI-N160**.

5.2.2 Crystallisation

The first crystals of cmcI-N160 appeared in PEG 4000, sodium acetate and Tris buffer, were box-shaped and did not diffract. Further screening in the presence of 2-oxoglutarate, a cofactor for 2ODDs, yielded needle-shaped crystals alongside the box-shaped ones in the same conditions (Figure 5.2). The needle-shaped crystals diffracted to around 2.5 Å resolution.

The crystallisation conditions used for cmcI-N160, only produced very thin needles of cmcI-D160 protein that were not suitable for data collection. Screening for new conditions yielded crystals in the presence of magnesium chloride instead of sodium acetate and either PEG 4000, or a mixture of PEG 8000/PEG 1000 or PEG 20000/PEG 550. 2-oxoglutarate was not necessary to obtain the needle-shaped crystals of cmcI-D160.

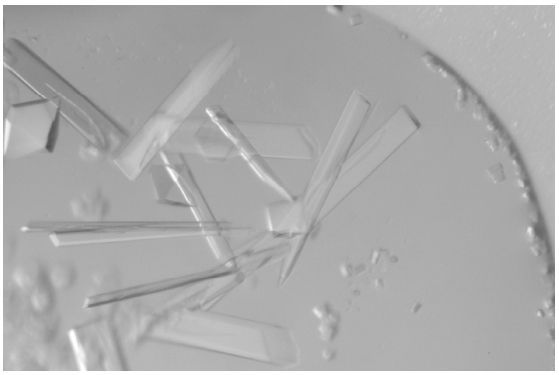

Figure 5.2 Crystallisation of cmcI-N160 in the presence of 2-oxoglutarate gave two crystal types: box- and needle-shaped crystals, of which only the needle-shaped crystals diffract.

5.2.3 Different crystal forms of cmcI

Morphologically similar crystals (needle-shaped) of cmcI have been shown to belong to two different space groups: $P2_1$ and $P2_12_12_1$. The orthorhombic crystals gave rise to two different unit cells (Table 5.1). There is no difference in diffraction limit (~2.5-3 Å) and the solvent content is very similar in all three crystal types.

Table 5.1 The different crystal forms observed for cmcI.

Space group	Cell dimensions a,b,c, β	Solvent content	Nr of molecules in a.u
$P2_1$	~ 93, 183, 103 Å, ~91°	55 %	12
$P2_12_12_1$	~ 93, 103, 183 Å	53 %	6
$P2_12_12_1$	~ 78, 145, 162 Å	58 %	6

The crystal structure revealed the same hexameric arrangement of cmcI in all three crystal forms. The differences between the crystal forms are caused by the altered packing of the hexamers relative each other, whereas the hexamer itself seems to form a rigid unit.

5.3 Crystal packing and oligomeric state of cmcI in solution

Judging by the first native data set collected on cmcI-N160, a large number of molecules of cmcI were indicated to be present in the asymmetric unit. This conclusion was based on the relatively large cell dimensions for a protein with a molecular weight of 28 kDa (space group $P2_1$, cell dimensions ~93, 183 and 103 Å and a β-angle of 91°). Unambiguous determination of the packing and the number of molecules in the asymmetric unit was not possible by calculations of the Matthews coefficient, V_M (Matthews, 1968), and solvent content. The differences between the values were very small and it appeared that there could be anything between 8 and 18 molecules present in the asymmetric unit (Table 5.2).

Table 5.2 V_M and solvent content as a function of the number of molecules in the asymmetric unit for the monoclinic space group of cmcI. For most protein crystals V_M is usually between 1.7-3.5 $Å^3$/Da with a mean value around 2.15 $Å^3$/Da and the solvent content is normally between 30-60% for a soluble protein (Drenth, 1994).

Nr of molecules in the a.u	V_M ($Å^3$/Da)	Solvent content (%)
8	4.1	69.6
9	3.6	65.8
10	3.3	62.0
11	3.0	58.3
12	2.7	54.5
13	2.5	50.7
14	2.3	46.9
15	2.2	43.1
16	2.0	39.3
17	1.9	35.5
18	1.8	31.7

To determine the oligomeric state of cmcI in solution, a number of different methods were used. Gel filtration analysis indicated that cmcI formed at least trimers, and similar results were obtained from native gel analysis. Dynamic laser light scattering showed a monomodal distribution with an estimated weight of around 180-200 kDa, depending on the temperature. Evidence for a hexamer was obtained from "native" (i.e. soft ionization) electrospray ionization mass spectrometry, which also showed monomers and dimers (Paper III). However, the appearance of the hexamer seemed to be dependent on the salt concentration of the sample. At 10 mM ammonium acetate pH 7 no hexamer was detected, whereas at 50 mM salt concentration a mixture of monomers, dimers and hexamers was observed. At even higher buffer concentrations no signal at all could be detected.

The self-rotation function (Paper III) showed two peaks in the $\chi = 60°$ section. These could either correspond to two sixfold axes or two threefold axes parallel to two twofold axes. Both options corresponds to 12 molecules in the asymmetric unit.

Taking all these facts together presented convincing arguments for the presence of twelve molecules in the asymmetric unit in the monoclinic crystals, probably derived from two hexamers, with one hexamer in the asymmetric unit of the orthorhombic crystals. This was corroborated from the structure of cmcI showing a hexameric arrangement.

5.4 Structure determination

At an early stage, significant non-isomorphism between the crystals of cmcI was detected, this was also the case for crystals of the same crystal form. Therefore, phasing with anomalous scatterers was recognised to be required. Extensive soaking and co-crystallisation with a number of different heavy atoms were performed and several MAD and SAD data sets were collected at the suitable wavelengths. However, none of these datasets produced high enough anomalous or dispersive signals to locate any heavy atom sites.

Selenomethionine substituted cmcI was produced (Budisa *et al.*, 1995) and was purified in a similar fashion as the unsubstituted protein, except including 10 mM β-mercaptoethanol in all buffers and adding a last desalting step. The 100 % incorporation of selenium was confirmed by mass-spectrometry. Crystals of selenomethionine-labelled cmcI-N160 were grown in similar conditions as native cmcI-N160, including 5 mM β-mercaptoethanol. Since cmcI contains ten methionines out of 236 amino acids, a high anomalous contribution, sufficient to solve the structure, was expected.

Despite collecting a number of MAD- and (highly redundant) SAD data sets on selenomethionine-labelled cmcI, the structure could not be solved. The X-ray fluorescence absorption spectrum of the frozen crystals produced high signals at the selenium K-edge, but the anomalous and dispersive signals from the data sets

were weak and never extended beyond 4-5 Å resolution. Due to the weak signals, the selenium sites were difficult to locate. Considering the many molecules of cmcI in the asymmetric unit and the presence of ten methionines in each monomer, a large number of selenium sites were expected, 60 in the orthorhombic space group and 120 in the monoclinic case. To locate the sites, a number of different programs were tried in particular Solve (Terwilliger & Berendzen, 1999), SHELXD (Schneider & Sheldrick, 2002) and SnB (Smith *et al.*, 1998). These are powerful programs and have successfully been used to locate high numbers of anomalous scatterers (Deacon & Ealick, 1999).

The absorption edge from the fluorescence scans had a 'smeared' appearance, which could be due to a mixture of oxidised and reduced seleniums in the selenomethionine substituted crystals. A sharp absorption edge is preferred to accurately determine the wavelengths corresponding to the peak and inflection points. To assure that only one oxidation species was present in the crystals at one time, attempts were made to fully oxidise the seleniums using hydrogen peroxide (Sharff *et al.*, 2000) or fully reduce them with dithiothreitol (DTT) (Thomazeau *et al.*, 2001). However, this did not result in significant differences either in the shape of the absorption edges from the scans or in the strength of the anomalous signals of the collected data.

A data set collected at beam line ID29 at ESRF finally yielded the structure solution from a selenomethionine crystal that was not exposed to either hydrogen peroxide or DTT. The crystal belonged to space group $P2_12_12_1$ with cell dimensions 92.8, 103.0 and 182.0 Å (Paper IV). Data were collected at three wavelengths, but due to radiation damage, only 400° of the peak data could be used. These data were scaled to 3.6 Å resolution and treated as a SAD data set. Compared to the previously collected data sets, this data set showed stronger anomalous signal and the signal was present to higher resolution. To locate the large number of selenium sites, a new program, Hybrid Substructure Search Procedure (HySS) was used (Grosse-Kunstleve & Adams, 2003). The present data set with its many anomalous scatterers provided a suitable test case for program development (Grosse-Kunstleve, personal communication). 52 Sites out of 60 possible were found with a correlation coefficient of 41%. A crucial ingredient for the substructure determination was to cut the data at 3.8 Å, including the data to 3.6 Å did not work. The reason for this is not entirely understood, the data is 100% complete at 3.6 Å and the anomalous signal is only slightly reduced between 3.8 and 3.6 Å.

The 52 initial sites were fed into SHARP (La Fortelle & Bricogne, 1997) for initial phasing. After removing and adding a few sites, 55 sites were used for final phase calculations, and the resolution was extended to 3.6 Å. Out of the 55 sites, one was later recognised as a false site, this meant that all seleniums except the N-terminal selenomethionines were located. The phases were subsequently improved by density modification procedures to give an excellent and interpretable electron density map. The map was further improved by exploiting the 6-fold non-crystallographic symmetry (NCS) of the asymmetric unit. An initial model was built into the 3.6 Å resolution map (Figure 5.3), comprising residues 9-230 (out

of 236 in total). This model was refined using NCS Phased Refinement in CCP4i, a program that cycles between Refmac5 (Murshudov *et al.*, 1997) and DM (Cowtan, 1994), which in turn uses the 6-fold averaging for map improvement.

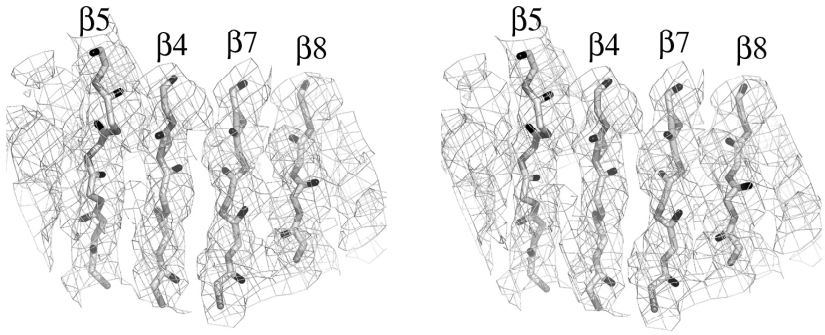

Figure 5.3 Stereoview of part of the initial electron density map at 3.6 Å with parts of the final Cα model overlaid. The electron density map was calculated with the phases derived from SAD phasing after density modification and 6-fold local averaging. The $2mF_{obs}$-DF_{calc} is contoured at 1.0 σ where σ is the root mean square electron density for the unit cell.

By using data from a cmcI-N160 crystal, the resolution was extended to 2.5 Å (Paper IV). This crystal belonged to the monoclinic space group $P2_1$, therefore molecular replacement was used with the hexamer as the search molecule. This model was initially refined using the NCS phased refinement with 12-fold averaging. Model building was cycled with maximum likelihood refinement in Refmac5 (Murshudov *et al.*, 1997) using strong NCS restraints, which were gradually relaxed. The NCS restraints were abandoned in the last stages of refinement and TLS parameter refinement with each monomer as a TLS group was implemented. The final model has an R_{cryst} of 18.9% and an R_{free} of 24.6%.

5.5 The structure of cmcI (Paper IV)

5.5.1 The monomer

The structure of cmcI revealed a cmcI monomer (Figure 5.4a) built up from an N-terminal domain involved in oligomerisation, and a C-terminal domain containing the proposed active site. The C-terminal domain, comprising amino acid residues 67-236, consists mainly of a modified Rossmann-like fold (Rossmann *et al.*, 1974) built up from a central parallel six-stranded β-sheet (β4-β9) extended by one anti-parallel strand (β10). A total of seven helices pack against the central sheet. The N-terminal domain, comprising amino acid residues 1-67, contains mainly loop structure but also two short α helices (α1 and α2), two small anti-parallel β

strands (β2-β3), and one β strand (β1) that forms an extension of the central β-sheet of the C-terminal domain of an adjacent molecule.

5.5.2 Quaternary structure

The N-terminal domain of cmcI is involved in extensive oligomerisation, forming two types of interactions with neighbouring molecules. Type one interaction involve residues 24-59 from one molecule that pack against the α3-, α8- and α9-helices and the loop comprising residues 215 to 226 of an adjacent molecule. An extension of the central β-sheet is formed by the interaction of the β1-strand of one molecule with the β9-strand of the adjacent molecule and vice versa, to form a dimer (Figure 5.4c). Except for the interactions from the N-terminal domain, there are also further tight interactions between the two monomers involving helices α3 and α4 from both monomers.

In the second type of interaction, residues 1-20 interact with a third molecule by packing against the helices α7, α8 and α9 and vice versa (Figure 5.4d). The type two interactions are weaker than the type one interactions, because the interface contains a large cavity that harbours the putative active site.

The oligomerisation is repeated cyclically and results in a hexameric arrangement of the molecules, where each molecule interacts with two neighbouring molecules (Figure 5.4e). The hexamer can be described as a dimer of trimers or a trimer of dimers and has the shape of a puckered doughnut, with a diameter of 100 Å in the widest dimension and 50 Å in the narrowest. In the middle of the hexamer is a solvent channel with a diameter of around 30 Å.

5.5.3 Structural alignment comparisons

When a structure of a protein with unknown function has been solved, structural comparison to the PDB database using a protein fold comparison server is an obvious approach to obtain information about the protein. Using the DALI server (Holm & Sander, 1993), a high structural homology of cmcI with small-molecule SAM dependent methyltransferases was found (Paper IV), with the most similar structure identified as cathechol O-methyltransferase (COMT) (Vidgren *et al.*, 1994). Dividing cmcI into N- and C-terminal domains and searching with each domain separately gave the same result for the C-terminal domain as for the whole monomer, whereas no similar structures were found with the N-terminal domain. Structural similarity of cmcI to NAD-dependent dehydrogenases was also observed. For both methyltransferases and dehydrogenases, the Rossmann fold domains overlap with cmcI and the nucleotide binding is positioned on the top of the Rossmann fold in these enzymes (Figure 5.5 and 5.7).

The structural similarity of cmcI to methyltransferases and dehydrogenases is interesting from a functional aspect. A methyltransferase has been proposed to be part of the methoxylation reaction and NADH stimulated hydroxylase activity has been reported (Coque *et al.,* 1995b).

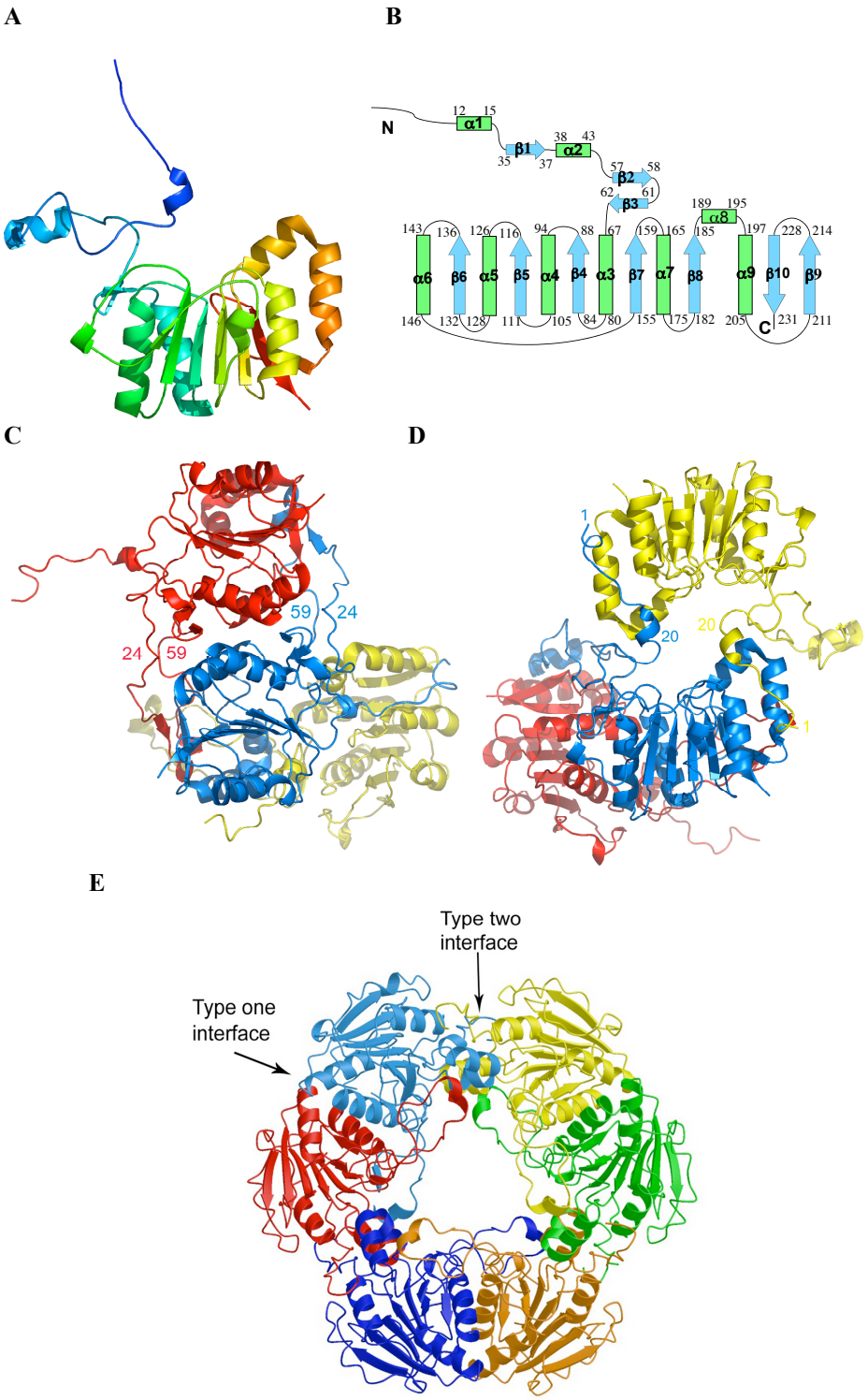

Figure 5.4 (A)-(E) Legend opposite.

Figure 5.4 The structure of cmcI. (A) Ribbon diagram with secondary structure elements colour ramped from blue to red by sequence from the N to the C terminus. (B) Topology diagram with the α-helices and β-strands labelled and coloured green and blue respectively. (C) Type one interactions between subunits. The two monomers in the dimer are in red and blue, and a third monomer (yellow) in the background. (D) Type two interactions between subunits (in blue and yellow) with a third monomer (red) in the background. (E) The hexamer of cmcI with the same colour coding as in C and D.

5.6 Complexes of cmcI

To further investigate the possible function of cmcI, crystals of complexes with a number of different cofactors were obtained by co-crystallisation. The cofactors were chosen by consideration to the different available theories on cephamycin biosynthesis.

5.6.1 Iron and 2-oxoglutarate

With respect to the proposal of a 2-oxoglutarate dependent dioxygenase in the cephamycin biosynthesis and the assignment in protein databases of cmcI for this function, thorough investigation of this possibility was made. Comparison of the structure of cmcI to structures of 2ODDs, did not detect similarity to the common jelly roll fold of 2ODDs. 2ODDs have been shown to use a conserved 2-His-1-carboxylate triad motif to bind the iron (Hegg & Que, 1997). Inspection of the cmcI structure, did not reveal a corresponding iron binding motif. Attempts to obtain activity of cmcI in the presence of Fe(II) and 2-oxoglutarate with or without SAM or SAM/Mg^{2+} was unsuccessful.

Attempts to co-crystallise and soak cmcI-N160 crystals with iron in an anaerobic box was performed, but did not result in specific iron binding. Co-crystallisation of cmcI-D160 in the presence of iron instead of magnesium in the anaerobic box produced thin needle-shaped crystals that were not suitable for data collection. In this case, one can suspect the iron probably takes the position of the missing magnesium ion.

From the structure of cmcI it is likely that cmcI is not the reported 2ODD involved in cephamycin biosynthesis and therefore the assignment of cmcI as the cephalosporin hydroxylase in protein databases is misleading!

5.6.2 SAM/SAH binding

Co-crystallisation of cmcI-N160 and D160 (paper IV) with SAM resulted in SAM binding to cmcI at a position corresponding to the binding site for SAM in methyltransferases, overlapping with the SAM binding site in COMT (Figure 5.5). The SAM binding site is located in a similar position in all known methyltransferases although the residues that bind SAM are poorly conserved

(Martin & McMillan, 2002). In cmcI, SAM is bound by a extensive network of hydrogen bonds and hydrophobic interactions.

A structure of cmcI-N160 in complex with S-adenosyl-L-homocysteine (SAH), the demethylated product of SAM, was also obtained (paper IV). The position of the SAH molecule and all interactions to the protein residues are identical to those in the SAM complex.

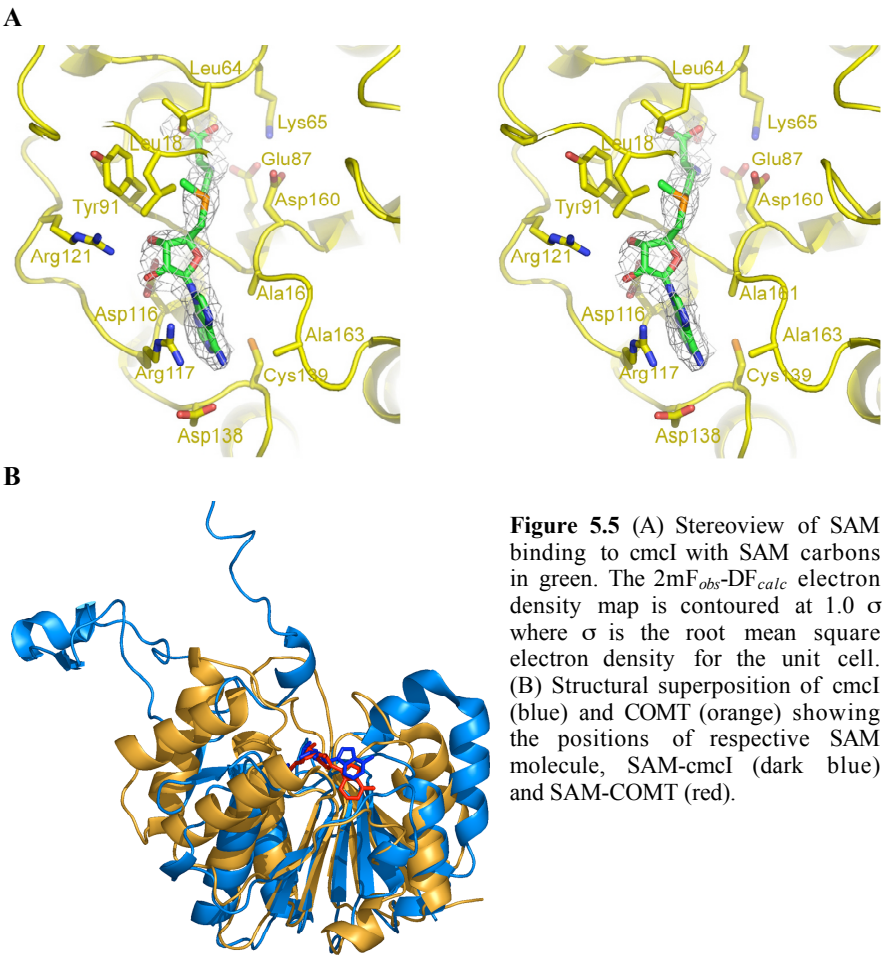

Figure 5.5 (A) Stereoview of SAM binding to cmcI with SAM carbons in green. The $2mF_{obs}$-DF_{calc} electron density map is contoured at 1.0 σ where σ is the root mean square electron density for the unit cell. (B) Structural superposition of cmcI (blue) and COMT (orange) showing the positions of respective SAM molecule, SAM-cmcI (dark blue) and SAM-COMT (red).

5.6.3 Magnesium binding and the importance of Asp160

A dramatic difference in the crystallisation conditions of cmcI-D160 compared to cmcI-N160 was observed, see section 5.2.2. It was found that cmcI-D160 crystals were only formed in the presence of Mg^{2+}.

In the structures of cmcI-D160 and the SAM-complex of cmcI-D160 (paper IV), a new density, which was modelled as a magnesium ion, appeared in the active site, close to the SAM binding position (Figure 5.6). The magnesium ion is ligated by

Asp160, Glu186 and Asp187. The position of the ion corresponds to the magnesium binding site in COMT, the methyltransferase identified as most structurally similar to cmcI. The magnesium ion is crucial for substrate binding and catalysis in COMT (Vidgren *et al.*, 1994), and it seems plausible that it plays a similar role in cmcI. Attempts have been made to co-crystallise and soak cmcI-N160 crystals with magnesium but no magnesium binding was obtained, which points to the importance of Asp160 in binding the ion. The corresponding ligand in COMT is Asp141.

Additional density, shaped as two elongated cylinders was observed close to the magnesium ion in the structures of cmcI-D160. Two PEG molecules with a length of 4 and 12 carbons, respectively, were modelled into the density. The PEG molecules are positioned in the cavity between two monomers on a two-fold symmetry axis. This is shown in Figure 5.6. The longer PEG extends all the way from the active site of one molecule to the active site of the other molecule. In the cmcI-D160 SAM complex, the density for both the magnesium ion and the PEG molecules is stronger compared to the structures lacking SAM, and the longer PEG molecule is extended by two carbon-units and ligates the magnesium ions directly (Figure 5.6). The PEG binding site overlaps with the substrate binding site in COMT and is likely to correspond to the substrate binding site also in cmcI. In the cmcI-N160 structures that do not bind magnesium ions, no density for PEG is present.

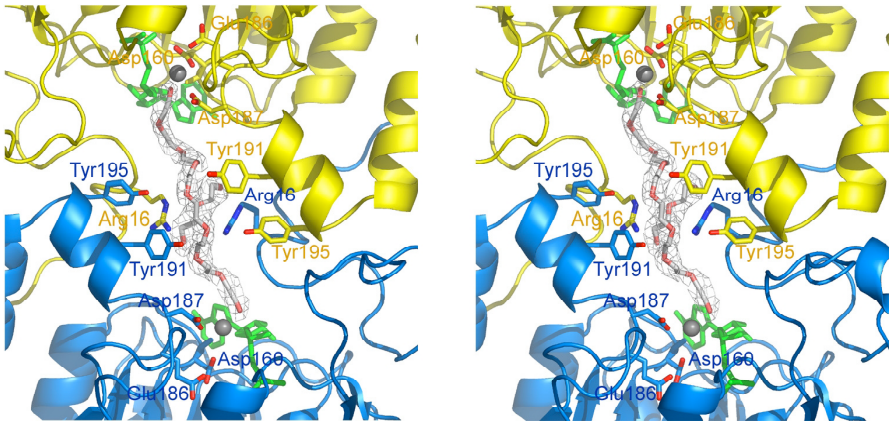

Figure 5.6 Stereoview showing the binding of magnesium and PEG in the structure of cmcI-D160-Mg^{2+}-SAM. One Mg^{2+} binds per subunit, and PEG (modelled as one 4-C unit and one 14-C unit) binds in the cavity between the two monomers coloured blue and yellow. PEG carbons are grey, the SAM molecules are green, and the Mg-ions are grey spheres. The $2mF_{obs}$-DF_{calc} electron density map at the PEG molecules is contoured at 1 σ where σ is the root-mean square electron density for the unit cell.

5.6.4 NADH binding

Co-crystallisation of cmcI-N160 and -D160 with 20 mM NADH (Appendix A) produced significantly different structures. In the cmcI-N160 structure, density for NADH was observed in the cavities between the molecules, resulting in only three molecules of NADH bound in the hexamer of cmcI (Figure 5.7). The binding does not correspond to the expected binding mode of NADH to a nucleotide-binding motif (Figure 5.7) and does not coincide with the SAM binding but overlaps with the observed PEG binding. The density corresponds to the central ribose-phosphate moiety of the NADH molecule, whereas there is only weak density for the nicotinamide and adenine parts.

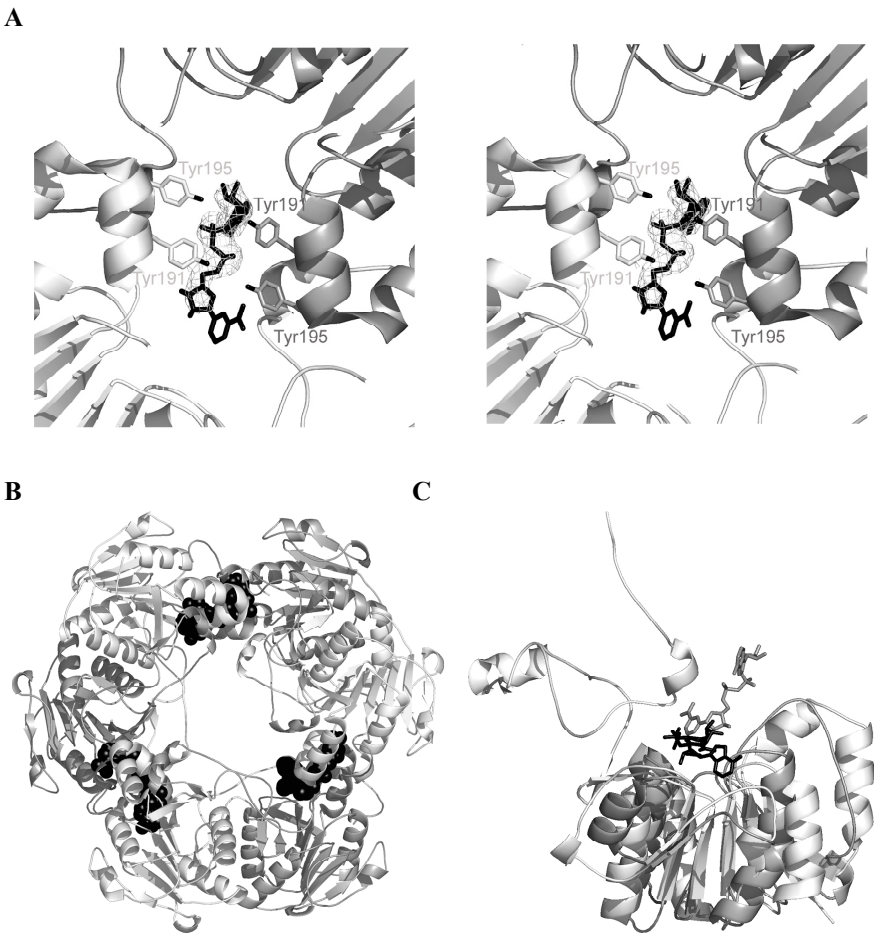

Figure 5.7 The binding of NADH to cmcI-N160. (A) Stereoview showing NADH binding with one cmcI molecule in dark grey and the other molecule in light grey. The $2mF_{obs}$-DF_{calc} electron density around the NADH molecule is contoured at 1 σ, where σ is the root mean square electron density for the unit cell. (B) Overview showing one hexamer with three bound NADH molecules as black spheres. (C) Superposition of cmcI (light grey) with human alcohol dehydrogenase (dark grey) (pdb-id 1agn) showing the positions of the respective NADH molecules, NADH-cmcI (grey) and NADH-alcohol dehydrogenase (black).

In sharp contrast, the structure of cmcI-D160 co-crystallised with NADH does not show NADH bound. Instead, continuous density for PEG is observed at the corresponding NADH binding site. Because of the unusual binding of NADH in the cmcI-N160 structure and the lack of NADH binding to the cmcI-D160 enzyme, it seems unlikely that the NADH binding is of functional significance, but rather reflects a structural role of NADH and the capacity of cmcI to bind large ligands in this cavity.

5.6.5 Docking of substrate

Despite persistent trials to co-crystallise cmcI with different cephalosporins, cephalosporin C, O-carbamoyl deacetoxycephalosporin C and 7-methoxy cephalosporin C, density for the cephalosporins were not detected in electron density maps of the corresponding structures. This is probably due to the bound PEG molecules, which interfere with the substrate binding.

Docking of O-carbamoyl-deacetylcephalosporin C to the refined structure of cmcI-D160 in complex with magnesium and SAM was performed using AutoDock3 (Morris *et al.*, 1998) (paper IV). The docking procedure consists of the generation of a grid of affinity potentials that represent the shape and electrostatic properties of the active site. As a starting point, a very large grid covering the whole cavity between two molecules was chosen. The docked conformations showed one minor cluster of conformations. The position of these conformations was used to create a more size-restricted grid. Docking of the molecule into this grid resulted in only one cluster of molecules (cluster cut-off r.m.s. deviation <1.5 Å).

The position and conformation of the docked molecule (Figure 5.8) makes chemical sense if cmcI is a methyltransferase. The methyl group on the SAM molecule is in a position where it could be transferred to the cephalosporin molecule, and the modelled cephalosporin overlaps with the bound competitive inhibitor in the structure of COMT (Vidgren *et al.*, 1994). The carbonyl oxygen of the β-lactam ring is at a distance of 3 Å to the magnesium ion. This positions C7 at a distance of over 5 Å to the methyl group of the SAM molecule, but with a hydroxyl group at the C7α, the distance is shortened to ~3.5 Å and the orientation could allow a methylation reaction to take place. The adipyl side chain of the cephalosporin molecule occupies the binding site of the longer PEG molecule in the crystal structure (Figure 5.8b). Several protein residues (Ser54, Tyr56, Asp66 and Gln223) are within van der Waals/hydrogen bonding distance to the carbamoyl side chain.

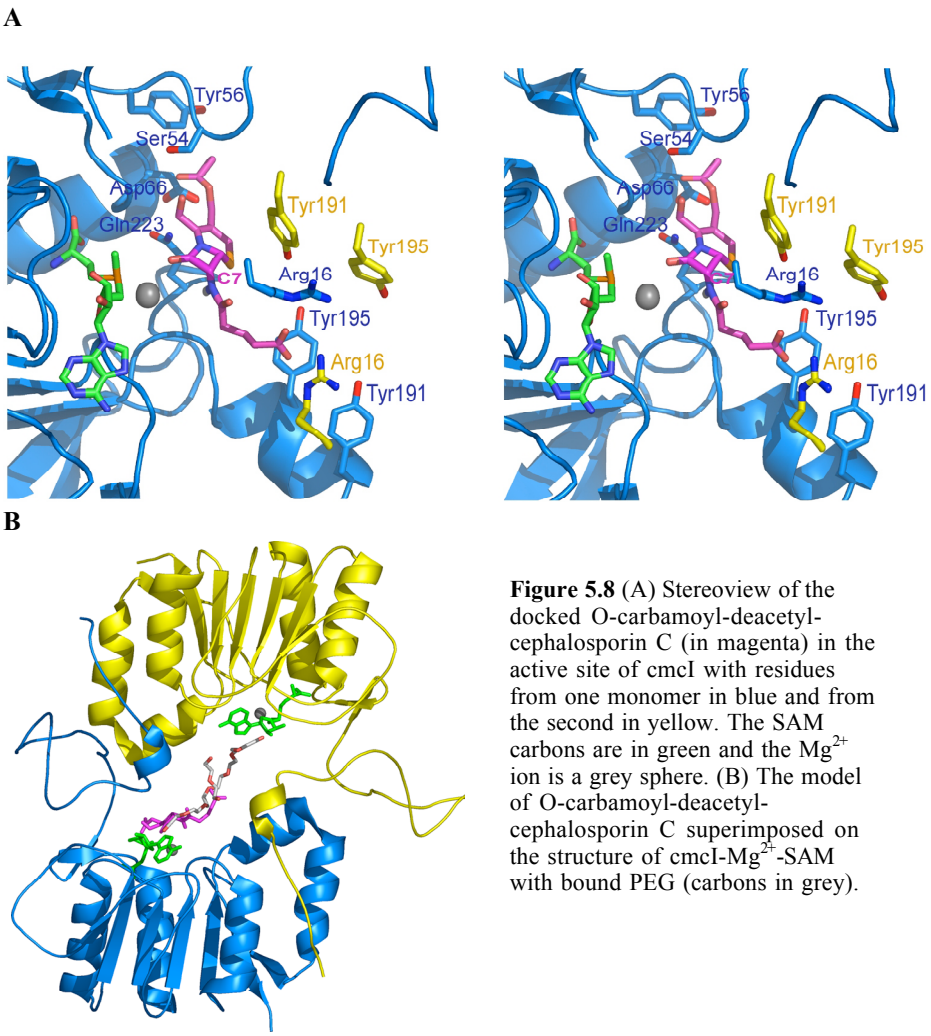

Figure 5.8 (A) Stereoview of the docked O-carbamoyl-deacetyl-cephalosporin C (in magenta) in the active site of cmcI with residues from one monomer in blue and from the second in yellow. The SAM carbons are in green and the Mg^{2+} ion is a grey sphere. (B) The model of O-carbamoyl-deacetyl-cephalosporin C superimposed on the structure of cmcI-Mg^{2+}-SAM with bound PEG (carbons in grey).

5.7 The function of cmcI

From what was known about the methoxylation reaction in the cephamycin biosynthesis before this project was started, three main theories on the function of cmcI and cmcJ could be formulated: (i) CmcI is the cephalosporin hydroxylase using 2-oxoglutarate or NADH as electron donor, and cmcJ is the methyltransferase. (ii) CmcJ is the hydroxylase and cmcI the methyltransferase. (iii) CmcI and cmcJ form a complex and perform both reactions together with the active site on cmcI and cmcJ working as a helper protein. By determining the structure of cmcI alone and in complex with cofactors at least one of these theories can be ruled out and the knowledge about cephamycin biosynthesis and the role of cmcI has increased dramatically.

Since the structure of cmcI bears no similarities to the family of 2-oxoglutarate dependent dioxygenases, neither in the fold nor in the important iron-binding motif, it seems unlikely that cmcI would function as a 2ODD. The structural similarity to NAD-dependent dehydrogenases points to a much more likely mechanism if cmcI is the hydroxylase. However, the unusual and probably unproductive binding of NADH, also rules this function out.

In respect to a hydroxylase function of cmcI, two other more "exotic" mechanisms have been considered. One is based on the structural similarity to aclacinomycin-10-hydroxylase from *S. pururascens* (Paper IV) (Jansson *et al.*, 2003). Aclacinomycin-10-hydroxylase has the fold of a methyltransferase and binds SAM/SAH in the common binding site for these enzymes but the enzyme is a hydroxylase and requires SAM, molecular oxygen and a reducing agent for activity. This mechanism is unlikely in light of the suggested cofactor requirements for 7α-cephem hydroxylation. The other mechanism considered in the aspect of SAM binding to cmcI, is the possibility of cmcI to be a SAM dependent radical enzyme (Layer *et al.*, 2004). These enzymes require an essential Fe_4S_4 cluster coordinated by a characteristic CxxxCxxC sequence motif (C, Cys; x any amino acid). The absence of these features in cmcI argues against this mechanism here. However, other types of reactions with one-electron donors that may lead to radical formation on SAM may be possible.

Based on the structural investigations and activity measurements on cmcI, theory number one can be ruled out. This is a significant finding since cmcI and cmcJ are assigned the respective roles as cephalosporin hydroxylase and methyltransferase in the different protein databases. This could serve as a lesson that the annotations of proteins should not be trusted without further studies confirming the literature. Unfortunately, several recently sequenced proteins from the big genome sequencing projects have been discovered to have putative functions as cephalosporin hydroxylases and methyltransferases based on their sequence similarities to cmcI and cmcJ, respectively. This could add a lot of confusion and unnecessary work in future research and the mistakes could take long to repair.

The structural similarity of cmcI to SAM-dependent methyltransferases, also in respect of SAM/SAH binding, indicates cmcI as the methyltransferase in the cephamycin biosynthesis. The magnesium ion is bound to cmcI at the same position as the crucial magnesium ion in COMT. The docking of the substrate analogue O-carbamoyl-deacetylcephalosporin C to cmcI further supports this theory since the docked conformation of the molecule places the C-7 position of the cephalosporin molecule in a way that would allow a methyltransfer from the bound SAM molecule to the cephalosporin. This could explain the lack of observed activity of cmcI to cephalosporin substrates that lack the 7α-hydroxyl group. There are no 7α-hydroxyl cephalosporins commercially available, these are labile compounds that are very difficult to synthesize. Methyltransferases commonly form oligomeric structures, predominantly dimers and tetramers, however at least two structures of hexameric methyltransferases (Weiss *et al.*, 2000; Tanaka *et al.*, 2004) have been reported. In most small molecule

methyltransferases, the N terminus is involved in the oligomerisation and this is also observed in cmcI.

If cmcI catalyses the methyl transfer reaction, the hydroxylase reaction would be left to be catalysed by cmcJ. Not much is known about cmcJ. It shares a low sequence homology with a gibberellin desaturase that has documented hydroxylase activity (Tudzynski *et al.*, 2003). Both cmcJ and the gibberellin desaturase have a HXD motif, which is part of the characteristic 2-His-1-carboxylate triad motif that ligates the iron in 2ODDs. In addition, the molecular weight of the purified cephalosporin hydroxylase reported by Xiao *et al.*(1991) corresponds to the molecular weight of cmcJ.

Hexameric methyltransferases have been reported, however there is a possibility that the observed hexameric arrangement of cmcI, in solution and in the crystal, may be an artefact caused by the high concentration of cmcI and possibly the lack of cmcJ. The stabilisation gained from the interaction with two adjacent molecules of cmcI in the hexamer, could be provided by cmcJ resulting in a cmcI/cmcJ complex that could catalyse the hydroxylase and methyltransfer reaction in the same active site as suggested (Coque et al., 1995b; Enguita et al., 1996). The monomer of cmcI in the present structure with a long excursion of the N terminal domain is unlikely to exist as a soluble entity on its own. Without further experimental proof, the theory that cmcI and cmcJ form a complex cannot be excluded.

6. Future perspectives

Are there ever any finished projects in science? I don't think so and with respect to the projects discussed in this thesis many questions remain or new ones have emerged from this work. Some of the most important aspects will be discussed below.

6.1 Deacetoxycephalosporin C Synthase

When I started my PhD one of the goals was to catch a 3-D structure of the ferryl intermediate of DAOCS. To do this a non-twinned crystal form was needed! The plan was to expose reduced DAOCS-Fe(II)-2-oxoglutarate crystals with oxygen in a controlled manner. From experience of work in the lab on horseradish peroxidase (Berglund *et al.*, 2002), the crystals were expected to be reduced in the x-ray beam upon data collection and the plan was therefore to merge the data from several different crystals, only collecting a few degrees on each crystal. This would not be possible using the twinned R3 crystal form of DAOCS since each crystal would have a different degree of twinning resulting in large errors on scaling. Unfortunately, the new non-twinned crystal form did not diffract to high enough resolution to be able to detect the small differences caused by ferryl formation at the iron, especially considering the problem of bound cryo-protectant. However, to catch the ferryl intermediate would still be a very interesting project!

With respect to the importance of the C terminus of DAOCS, a complex with bound substrate or product with a visible free C terminal tail would be very interesting to obtain. Co-crystallisation of His-tagged DAOCS in an anaerobic box in the presence of iron, 2-oxoglutarate and substrate might be the way to obtain this. However, crystal form II only grow in the cold-room at a temperature of 4°C. This would introduce the additional problem of how to cool the sample in the anaerobic box.

6.2 ORF7

A key to success in this project seems to be improved purity of ORF7. The purification has been altered in a number of ways. Re-cloning to add a tag to simplify the purification has been attempted. Unfortunately, no expression or insoluble protein was the outcome of these experiments.

ORF15, the homologue to ORF7 with 48 % sequence identity has successfully been crystallised in our lab. Native data has been collected and phase determination is under way. The structure of ORF15 will most probably shed light on the function of both ORF7 and ORF15 in the clavulanic acid biosynthesis. However, since both the *orf7* and *orf15* genes are crucial for clavulanic acid production and cannot replace each other, structural investigations and comparisons of both proteins would be highly interesting.

6.3 CmcI

Many important and basic questions remain to be answered to understand the biosynthesis of cephamycin C. Is cmcI the methyltransferase and cmcJ the hydroxylase and do the enzymes work separately or in a complex? What is the function of the hexamer of cmcI? Do cmcI and cmcJ work together in a complex, performing both reactions?

These questions can best be answered by activity assays of the two enzymes, together and separately. The activity studies so far performed on cmcI failed to detect activity. It is impossible to know if this depends on lack of substrate, lack of cmcJ or inactive cmcI. Attempts are under way to express cmcJ on its own and to co-express cmcI and cmcJ.

7. References

Abraham, E.P. & Chain, E.B. (1940). An enzyme from bacteria able to destroy penicillin. *Nature* **146**, 837.

Alexander, D.C. & Jensen, S.E. (1998). Investigation of the Streptomyces clavuligerus cephamycin C gene cluster and its regulation by the CcaR protein. *J. Bacteriol.* **180**, 4068-79.

Arnold, F.H. (2001). Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253-7.

Arroyo, M., de la Mata, I., Acebal, C. & Pilar Castillon, M. (2002). Biotechnological applications of penicillin acylases: state-of-the-art. *Appl. Microbiol. Biotechnol.* **60**, 507-514.

Bachmann, B.O., Li, R. & Townsend, C.A. (1998). Beta-Lactam synthetase: a new biosynthetic enzyme. *Proc. Natl. Acad. Sci. USA* **95**, 9082-6.

Baggaley, K.H., Brown, A.G. & Schofield, C.J. (1997). Chemistry and biosynthesis of clavulanic acid and other clavams. *Nat. Prod. Rep.* **14**, 309-33.

Baldwin, J.E., Adlington, R.M., Bryans, J.S., Bringhen, A.O., Coates, J.B., Crouch, N.P., Lloyd, M.D. & Schofield, C.J. (1991). Isolation of dihydroclavaminic acid, an intermediate in the biosynthesis of clavulanic acid. *Tetrahedron* **47**, 4089-4100.

Baldwin, J.E., Lloyd, M.D., Wha-Son, B., Schofield, C.J., Elson, S.W., Baggaley, K.H. & Nicholson, N.H. (1993). A substrate analog study on clavaminic acid synthase: possible clues to the biosynthetic origin of proclavaminic acid. *J. Chem. Soc. Chem. Commun.*, 500-502.

Baldwin, J.E. & Schofield, C.J. (1992). The Chemistry of beta-lactams. (Page, M. I.) London, Blackie, 1-78.

Barends, T.R., Yoshida, H. & Dijkstra, B.W. (2004). Three-dimensional structures of enzymes useful for beta-lactam antibiotic production. *Curr. Opin. Biotechnol.* **15**, 356-63.

Barlow, J.N., Zhang, Z., John, P., Baldwin, J.E. & Schofield, C.J. (1997). Inactivation of 1-Aminocyclopropane-1-carboxylate Oxidase Involves Oxidative Modifications. *Biochemistry* **36**, 3563-3569.

Bayles, K.W. (2000). The bactericidal action of penicillin: new clues to an unsolved mystery. *Trends. Microbiol.* **8**, 274-278.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrel, B.G., Parkhill, J. & Hopwood, D.A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147.

Berglund, G.I., Carlsson, G.H., Smith, A.T., Szoke, H., Henriksen, A. & Hajdu, J. (2002). The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417**, 463-8.

Brakhage, A.A. (1998). Molecular Regulation of beta-lactam biosynthesis in filamentous fungi. *Microbiol. Mol. Biol. Rev.* **62**, 547-585.

Breithaupt, H. (1999). The new antibiotics. *Nat. Biotechnol.* **17**, 1165-1169.

Brewer, S.J., Taylor, P.M. & Turner, M.K. (1980). An adenosine triphosphate-dependent carbamoylphosphate-3-hydroxymethylcephem O-carbamoyltransferase from Streptomyces clavuligerus. *Biochem. J.* **185**, 555-64.

Brotzu, G. (1948). Richerche su di un nuovo antibiotico. *Lavori dell'Istituto d'Igiene di Cagliari*, 1-11.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem.* **230**, 788-796.

Burton, H.S. & Abraham, E.P. (1951). Isolation of antibiotics from a species of *Cephalosporium*: cephalosporins P1, P2, P3, P4 and P5. *Biochem. J.* **50**, 168.

Burzlaff, N.I., Rutledge, P.J., Clifton, I.J., Hensgens, C.M.H., Pickford, M., Adlington, R.M., Roach, P.L. & Baldwin, J.E. (1999). The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721-724.

Byford, M.F., Baldwin, J.E., Shiau, C.-Y. & Schofield, C.J. (1997). The mechanism of ACV Synthetase. *Chem. Rev.* **97**, 2631-2649.

Caines, M.E., Elkins, J.M., Hewitson, K.S. & Schofield, C.J. (2004). Crystal structure and mechanistic implications of N2-(2-carboxyethyl)arginine synthase, the first enzyme in the clavulanic acid biosynthesis pathway. *J. Biol. Chem.* **279**, 5685-92.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.

Chin, H.S., Goo, K.S. & Sim, T.S. (2004). A Complete Library of Amino Acid Alterations at N304 in *Streptomyces clavuligerus* Deacetoxycephalosporin C Synthase Elucidates the Basis for Enhanced Penicillin Analogue Conversion. *Appl. Environ. Microbiol.* **70**, 607-609.

Chin, H.S., Sim, J. & Sim, T.S. (2001). Mutation of N304 to Leucine in *Streptomyces clavuligerus* Deacetoxycephalosporin C Synthase Creates an Enzyme with Increased Penicillin Analogue Conversion. *Biochem. Biophys. Res. Commun.* **287**, 507-513.

Chin, H.S. & Sim, T.S. (2002). C-terminus modification of *Streptomyces clavuligerus* deacetoxycephalosporin C synthase improves catalysis with an expanded substrate specificity. *Biochem. Biophys. Res. Commun.* **295**, 55-61.

Cho, H., Adrio, J.L., Luengo, J.M., Wolfe, S., Ocran, S., Hintermann, G., Piret, J.M. & Demain, A.L. (1998). Elucidation of conditions allowing conversion of penicillin G and other penicillins to deacetoxycephalosporins by resting cells and extracts of *Streptomyces clavuligerus* NP1.

Clifton, I.J., Doan, L.X., Sleeman, M.C., Topf, M., Suzuki, H., Wilmouth, R.C. & Schofield, C.J. (2003). Crystal structure of carbapenem synthase (CarC). *J. Biol. Chem.* **278**, 20843-50.

Cohen, M.L. (2000). Changing patterns of infectious disease. *Nature* **406**, 762-767.

Coque, J.J., Perez-Llarena, F.J., Enguita, F.J., Fuente, J.L., Martin, J.F. & Liras, P. (1995a). Characterization of the cmcH genes of Nocardia lactamdurans and Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem O-carbamoyltransferase for cephamycin biosynthesis. *Gene* **162**, 21-7.

Coque, J.J.R., Enguita, F.J., Martin, J.F. & Liras, P. (1995b). A two-protein component 7alpha-Cephem-Methoxylase encoded by two genes of the Cephamycin C cluster converts Cephalosporin C to 7-Methoxycephalosporin C. *J. Bacteriol.* **177**, 2230-2235.

Cowtan, K. (1994). dm: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.

Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-91.

Crawford, L., Stepan, A.M., McAda, P.C., Rambosek, J.A., Conder, M.J., Vinci, V.A. & Reeves, C.D. (1995). Production of cephalosporin intermediates by feeding adipic acid to recombinant *Penicillium chrysogenum* strains expressing ring expansion activity. *Biotechnol.* **13**, 58-62.

Crowfoot, D., Bunn, C.W., Rogers-Low, B. & Turner-Jones, A. (1949). The Chemistry of Penicillin) Princeton, N.J., Princeton University Press.

Deacon, A.M. & Ealick, S.E. (1999). Selenium-based MAD phasing: setting the sites on larger structures. *Structure* **7**, 161-166.

Demain, A.L. & Baez-Vasquez, M.A. (2000). Immobilized *Streptomyces clavuligerus* NP1 cells for biotransformation of penicillin G into deacetoxycephalosporin G. *Appl. Biochem. Biotechnol.* **87**, 135-140.

Demain, A.L. & Fang, A. (2000). The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **69**, 1-39.

Drenth, J. (1994). Principles of Protein X-ray Crystallography. Springer.

Dubus, A., Lloyd, M.D., Lee, H.-J., Shofield, C.J., Baldwin, J.E. & Frere, J.-M. (2001). Probing the penicillin sidechain selectivity of recombinant deacetoxycephalosporin C synthase. *Cell. Mol. Life Sci.* **58**, 835-843.

Elander, R.P. (2003). Industrial production of beta-lactam antibiotics. *Appl. Microbiol. Biotechnol.* **61**, 385-392.

Elkins, J.M., Clifton, I.J., Hernandez, H., Doan, L.X., Robinson, C.V., Schofield, C.J. & Hewitson, K.S. (2002a). Oligomeric structure of proclavaminic acid amidino hydrolase: evolution of a hydrolytic enzyme in clavulanic acid biosynthesis. *Biochem. J.* **366**, 423-34.

Elkins, J.M., Hewitson, K.S., McNeill, L.A., Seibel, J.F., Schlemminger, I., Pugh, C.W., Ratcliffe, P.J. & Schofield, C.J. (2003). Structure of Factor-inhibiting Hypoxia-inducible Factor (HIF) reveals mechanism of oxidative modification of HIF-1a. *J. Biol. Chem.* **278**, 1802-1806.

Elkins, J.M., Ryle, M.J., Clifton, I.J., Dunning Hotopp, J.C., Lloyd, J.S., Burzlaff, N.I., Baldwin, J.E., Hausinger, R.P. & Roach, P.L. (2002b). X-ray crystal structure of Escherichia coli taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrate. *Biochemistry* **16**, 5185-5192.

Enguita, F.J., Liras, P., Leitao, A.L. & Martin, J.F. (1996). Interaction of the two proteins of the methoxylation system involved in Cephamycin C biosynthesis. *J. Biol. Chem.* **271**, 33225-33230.

Fernandez, M.-J., Adrio, J.L., Piret, J.M., Wolfe, S., Ro, S. & Demain, A.L. (1999). Stimulatory effect of growth in the presence of alcohols on biotransformation of penicillin G into cephalosporin-type antibiotics by resting cells of *Streptomyces clavuligerus* NP1. *Appl. Microbiol. Biotechnol.* **52**, 484-488.

Fleming, A. (1929). On the antibacterial action of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. *Brit. J. Exp. Pathol.* **10**, 226-236.

Florey, H.W., Chain, E.B., Heatley, N.G., Jennings, M.A., Sanders, A.G., Abraham, E.P. & Florey, M.E. (1949). Antibiotics) London, Oxford University Press.

Garman, E. & Schneider, T.R. (1997). Macromolecular cryocrystallography. *J. Appl. Cryst.* **30**, 211-237.

Grosse-Kunstleve, R.W. & Adams, P.D. (2003). Substructure search procedures for macromolecular structures. *Acta Crystallogr. D* **59**, 1966-73.

Hausinger, R.P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21-68.

Hayward, S. & Berendsen, H.J.C. (1998). Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144-154.

Hegg, E.L. & Que, L. (1997). The 2-His-1-carboxylate facial triad. *Eur. J. Biochem.* **250**, 625-629.

Hodgson, J.E., Fosberry, A.P., Rawlinson, N.S., Ross, H.N., Neal, R.J., Arnell, J.C., Earl, A.J. & Lawlor, E.J. (1995). Clavulanic acid biosynthesis in Streptomyces clavuligerus: gene cloning and characterization. *Gene* **166**, 49-55.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Holme, E. (1975). A Kinetic Study of Thymine 7-Hydroxylase from *Neurospora crassa*. *Biochemistry* **14**, 4999-5003.

Hood, J.D., Elson, A., Gilpin, M.L. & Brown, A.G. (1983). Identification of 7alpha-hydroxycephalosporin C as an intermediate in the methoxylation of Cephalosporin C by a cell free extract of Streptomyces clavuligerus. *J. Chem. Soc. Chem. Comm.*, 1187-1188.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526-531.

Jansson, A., Niemi, J., Lindqvist, Y., Mäntsälä, P. & Schneider, G. (2003). Crystal Structure of Aclacinomycin-10-hydroxylase, a S-Adenosyl-L-Methionine-dependent Methyltransferase homolog involved in Acrtacycline Biosynthesis in Streptomyces purpurascens. *J. Mol. Biol.* **334**, 269-280.

Jensen, S.E., Elder, K.J., Aidoo, K.A. & Paradkar, A.S. (2000). Enzymes catalysing the early steps of Clavulanic acid biosynthesis are encoded by two sets of paralogous genes in *Streptomyces clavuligerus*. *Antimicrob. Agents Chemother.* **44**, 720-726.

Jensen, S.E., Westlake, D.W. & Wolfe, S. (1983). Partial purification and characterization of isopenicillin N epimerase activity from *Streptomyces clavuligerus*. *Can. J. Microbiol.* **29**, 1526-1531.

Jensen, S.E., Westlake, D.W. & Wolfe, S. (1985). Deacetoxycephalosporin C synthetase and deacetoxycephalosporin C hydroxylase are two separate enzymes in *Streptomyces clavuligerus*. *J. Antibiot.* **38**, 263-265.

Kennedy, J. & Turner, G. (1996). Aminoadipyl-L-cysteinyl-D-valine synthetase is a rate limiting enzyme for penicillin production in Aspergillus nidulans. *Mol. Gen. Genet.* **253**, 189-197.

Khaleeli, N., Li, R. & Townsend, C.A. (1999). Origin of the beta-lactam carbons in clavulanic acid from an unusual thiamine pyrophosphate-mediated reaction. *J. Am. Chem. Soc.* **121**,

Kivirikko, K.I. & Pihlajaniemi, T. (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylase. *Adv. Enzymol. Relat. Areas Mol. Biol.* **2**, 325-400.

La Fortelle, E. & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472-492.
Layer, G., Heinz, D.W., Jahn, D. & Schubert, W.-D. (2004). Structure and function of radical SAM enzymes. *Curr. Opin. Chem. Biol.* **8**, 468-476.
Lee, H.-J., Lloyd, M.D., Harlos, K., Clifton, I.J., Baldwin, J.E. & Schofield, C.J. (2001). Kinetic and Crystallographic Studies on Deacetoxycephalosporin C Synthase (DAOCS). *J. Mol. Biol.* **308**, 937-948.
Lee, H.-J., Schofield, C.J. & Lloyd, M.D. (2002). Active Site Mutations of Recombinant Deacetoxycephalosporin C Synthase. *Biochem. Biophys. Res. Commun.* **292**, 66-70.
Li, R., Khaleeli, N. & Townsend, C.A. (2000). Expansion of the clavulanic acid gene cluster: identification and in vivo functional analysis of three new genes required for biosynthesis of clavulanic acid by Streptomyces clavuligerus. *J. Bacteriol.* **182**, 4087-95.
Lipscomb, S.J., Lee, H.-J., Mukherji, M., Baldwin, J.E., Schofield, C.J. & Lloyd, M.D. (2002). The role of arginine residues in substrate binding and catalysis by deacetoxycephalosporin C synthase. *Eur. J. Biochem.* **269**, 2735-2739.
Lloyd, M.D., Lee, H.-J., Harlos, K., Zhang, Z.-H., Baldwin, J.E., Schofield, C.J., Charnock, J.M., Garner, C.D., Hara, T., Terwisscha van Scheltinga, A.C., Valegård, K., Viklund, J.A.C., Hajdu, J., Andersson, I., Danielsson, Å. & Bhikhabhai, R. (1999). Studies on the Active Site of Deacetoxycephalosporin C Synthase. *J. Mol. Biol.* **287**, 943-960.
Lloyd, M.D., Lipscomb, S.J., Hewitson, K.S., Hensgens, C.M., Baldwin, J.E. & Schofield, C.J. (2004). Controlling the substrate selectivity of deacetoxycephalosporin/deacetylcephalosporin C synthase. *J Biol Chem* **279**, 15420-6.
Lorenzana, L.M., Perez-Redondo, R., Santamarta, I., Martin, J.F. & Liras, P. (2004). Two Oligopeptide-Permease-Encoding Genes in the Clavulanic Acid Cluster of *Streptomyces clavuligerus* are Essential for Production of the beta-lactamase Inhibitor. *J. Bacteriol.* **186**, 3431-3438.
Luengo, J.M. (1995). Enzymatic synthesis of hydrophobic penicillins. *J. Antibiot.* **48**, 1195-1212.
Martin, J.L. & McMillan, F. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783-793.
Matagne, A., Dubus, A., Galleni, M. & Frere, J.-M. (1999). The beta-lactamase cycle: a tale of selective pressure and bacterial ingenuity. *Nat. Prod. Rep.* **16**, 1-19.
Matthews, B.W. (1968). *J. Mol. Biol.* **33**, 491-497.
Mellado, E., Lorenzana, L.M., Rodriguez-Saiz, M., Diez, B., Liras, P. & Barredo, J.L. (2002). The clavulanic acid biosynthetic cluster of Streptomyces clavuligerus: genetic organization of the region upstream of the car gene. *Microbiology* **148**, 1427-38.
Miller, I.M., Stapley, E.O. & Chaiet, L. (1962). Production of synnematin B by a member of the genus *Streptomyces*. *Bacteriol. Proc.* **49**, 32.
Miller, M.T., Bachmann, B.O., Townsend, C.A. & Rosenzweig, A.C. (2001). Structure of beta-lactam synthetase reveals how to synthesize antibiotics instead of asparagine. *Nat. Struct. Biol.* **8**, 684-9.
Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. & Olsson, A.J. (1998). Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Computational Chemistry* **19**, 1639-1662.

Muller, I., Kahnert, A., Pape, T., Sheldrick, G.M., Meyer-Klaucke, W., Dierks, T., Kertesz, M. & Uson, I. (2004). Crystal Structure of the Alkylsulfatase AtsK: Insights into the Catalytic Mechanism of the Fe(II) alpha-ketoglutarate-Dependent Dioxygenase superfamily. *Biochemistry* **43**, 3075-3088.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D* **53**, 240-255.

Nagarajan, R., Boeck, L.D., Gorman, M., Hamill, R.L., Higgens, C.H., Hoehn, M.M., Stark, W.M. & Whitney, J.G. (1971). Beta-lactam antibiotics from *Streptomyces*. *J. Amer. Chem. Soc.* **93**, 2308-2310.

Nicholson, N.H., Baggaley, K.H., Cassels, R., Davison, M., Elson, S.W., Fulston, M., Tyler, J.W. & Woroniecki, S.R. (1994). Evidence that the immediate biosynthetic precursor of clavulanic acid is its N-aldehyde analog. *J. Chem. Soc. Chem. Commun.* **11**, 1281-1282.

O'Sullivan, J. & Abraham, E.P. (1980). The conversion of cephalosporins to 7alpha-methoxycephalosporins by cell-free extracts of Streptomyces clavuligerus. *Biochem. J.* **186**, 613-616.

Poole, K. (2004). Resistance to beta-lactam antibiotics. *Cell. Mol. Life Sci.* **61**, 2200-2223.

Prescott, A.G. (1993). A dilemma of dioxygenases (or where biochemistry and molecular biology fail to meet). *J. Exp. Bot.* **44**, 849-861.

Prescott, A.G. & Lloyd, M.D. (2000). The iron(II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. *Nat. Prod. Rep.* **17**, 367-83.

Price, J.C., Barr, E.W., Tirupati, B., Bollinger, J.M., Jr. & Krebs, C. (2003). The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **42**, 7497-508.

Proshlyakov, D.A., Henshaw, T.F., Monterosso, G.R., Ryle, M.J. & Hausinger, R.P. (2004). Direct detection of oxygen intermediates in the non-heme Fe enzyme taurine/alpha-ketoglutarate dioxygenase. *J. Am. Chem. Soc.* **126**, 1022-3.

Roach, P.L., Clifton, I.J., Fulöp, V., Harlos, K., Barton, G.J., Hajdu, J., Andersson, I., Schofield, C.J. & Baldwin, J.E. (1995). Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **375**, 700-704.

Roach, P.L., Clifton, I.J., Hensgens, C.M.H., Schofield, C.J., Hajdu, J. & Baldwin, J.E. (1997). Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827-830.

Rossmann, M.G., Moras, D. & Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250**, 194-199.

Salowe, S.P., Krol, W.J., Iwata-Reuyl, D. & Townsend, C.A. (1991). Elucidation of the order of oxidations and identification of an intermediate in the multistep clavaminate synthase reaction. *Biochemistry* **30**, 2281-2292.

Sami, M., Brown, T.J.N., Roach, P.L., Schofield, C.J. & Baldwin, J.E. (1997). Glutamine-330 is not essential for activity in isopenicillin N synthase from Aspergillus nidulans. *FEBS Lett.* **405**, 191-194.

Schneider, T.R. & Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772-1779.

Schofield, C.J. & Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722-31.

Sharff, A.J., Koronakis, E., Luisi, B. & Koronakis, V. (2000). Oxidation of selenomethionine: some MADness in the method! *Acta Crystallogr. D* **56**, 785-788.

Sio, C.F. & Quax, W.J. (2004). Improved beta-lactam acylases and their use as industrial biocatalysts. *Curr. Opin. Biotechnol.* **15**, 349-55.

Skatrud, P.L., Tietz, A.J., Ingolia, T.D., Cantwell, C.A., Fisher, D.L., Chapman, J.L. & Queener, S.W. (1989). Use of recombinant DNA to improve production of cephalosporin C by *Cephalosporium acremonium*. *Bio/Technology* **7**, 477-485.

Smith, G.D., Nagar, B., Rini, J.M., Hauptman, H.A. & Blessing, R.H. (1998). The Use of SnB to Determine an Anomalous Scattering Substructure. *Acta Crystallogr. D* **54**, 799-804.

Stapley, E.O., Birnbaum, J., Miller, A.K., Wallick, H., Hendlin, D. & Woodruff, H.B. (1979). Cefoxitin and cephamycins: microbiological studies. *Rev. Inf. Dis.* **1**, 73-87.

Stapley, E.O., Jackson, M., Hernandez, S., Zimmermann, S.B., Currie, S.A., Mochalis, S., Mahta, J.M., Woodruff, H.B. & Hendlin, D. (1972). Cephamycins, a new family of beta-lactam antibiotics. 1. Production of actinomycetes, including *Streptomycetes lactamdurans* sp.n. *Antimicrob. Agents Chemother.* **2**, 122-131.

Tanaka, Y., Tsumoto, K., Yasutake, Y., Umetsu, M., Yao, M., Fukada, H., Tanaka, I. & Kumagai, I. (2004). How Oligomerization Contributes to the Thermostability of an Archaeon Protein. *J. Biol. Chem.* **279**, 32957-32967.

Terwilliger, T.C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849-861.

Terwisscha van Scheltinga, A.C., Valegård, K., Ramaswamy, S., Hajdu, J. & Andersson, I. (2001). Multiple isomorphous replacement on merohedral twins: structure determination of deacetoxycephalosporin C synthase. *Acta Crystallogr. D* **57**, 1776-1785.

Thomazeau, K., Gilles, C., Thompson, A., Dumas, R. & Biou, V. (2001). MAD on threonine synthase: the phasing power of oxidized selenomethionine. *Acta Crystallogr. D* **57**, 1337-1340.

Thykaer, J. & Nielsen, J. (2003). Metabolic engineering of beta-lactam production. *Metab. Eng.* **5**, 56-69.

Tudzynski, B., Mihlan, M., Rojas, M.C., Linnemannstöns, P., Gaskin, P. & Hedden, P. (2003). Characterization of the final two genes of the Gibberellin biosynthesis gene cluster of Gibberella fujikuroi. *J. Biol. Chem.* **278**, 28635-28643.

Valegård, K., Terwisscha van Scheltinga, A.C., Lloyd, M.D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H.-J., Baldwin, J.E., Schofield, C.J., Hajdu, J. & Andersson, I. (1998). Structure of a cephalosporin synthase. *Nature* **394**, 805-809.

Velasco, J., Adrio, J.L., Moreno, M.A., Diez, B., Soler, G. & Barredo, J.L. (2000). Environmentally safe production of 7-aminodeacetoxycephalosporanic acid (7-ADCA) using recombinant strains of *Acremonium chrysogenum*. *Nat. Biotechnol.* **18**, 857-861.

Vidgren, J., Svensson, A.L. & Liljas, A. (1994). Crystal structure of cathecol O-methyltransferase. *Nature* **368**, 354-358.

Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781.

Ward, J.M. & Hodgson, J.E. (1993). The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three Streptomyces. *FEMS Microbiol. Lett.* **110**, 239-42.

Wei, C.-L., Yang, Y.-B., Wang, W.-C., Liu, W.-C., Hsu, J.-S. & Tsai, Y.-C. (2003). Engineering Streptomyces clavuligerus Deacetoxycephalosporin C Synthase for Optimal Ring Expansion Activity toward Penicillin G. *Appl. Environ. Microbiol.* **69**, 2306-2312.

Weiss, V.H., McBride, A.E., Soriano, M.A., Filman, D.J., Silver, P.A. & Hogle, J.M. (2000). The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. *Nat. Struct. Biol.* **7**, 1165-1171.

Wilmouth, R.C., Turnbull, J.J., Welford, R.W.D., Clifton, I.J., Prescott, A.G. & Schofield, C.J. (2002). Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis thaliana. *Structure* **10**, 93-103.

Xiao, X., Wolfe, S. & Demain, A.L. (1991). Purification and characterisation of cephalosporin 7alpha-hydroxylase from Streptomyces clavuligerus. *Biochem. J.* **280**, 471-474.

Zhang, Z., Ren, J., Harlos, K., McKinnon, C.H., Clifton, I.J. & Schofield, C.J. (2002). Crystal structure of clavaminate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centred rearrangements. *FEBS Lett.* **517**, 7-12.

Zhang, Z., Ren, J., Stammers, D.K., Baldwin, J.E., Harlos, K. & Schofield, C.J. (2000). Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7**, 127-133.

Zhou, J., Kelly, W.L., Bachmann, B.O., Gunsior, M., Townsend, C.A. & Solomon, E.I. (2001). Spectroscopic studies of substrate interactions with clavaminate synthase 2, a multifunctional alpha-KG-dependent non-heme iron enzyme: correlation with mechanisms and reactivities. *J. Am. Chem. Soc.* **123**, 7388-98.

8. Acknowledgements

First of all I would like to thank my supervisor Inger. Thanks for your never-ending enthusiasm, for always believing in me and taking time to give me support and guidance. Thanks for being such a nice person and always caring what is best for all of us in the group, not only from a scientific point of view.

Anke, my co-supervisor, thanks for all your help and support, especially for your patience and enthusiasm in the early days when you introduced me to the world of crystallography. The lab has gone much more quite and boring since you left, I miss your laughs! Karin, not officially my supervisor but you have always felt like one and also a very good friend. Thanks for all your support and help, I will miss our coffee breaks, lunches and discussions about all and nothing.

This work was part of a EU-collaboration with the groups of Chris Schofield from Oxford and Jean-Marie Frère in Liege, Belgium. I would like to thank all people involved for great meetings, discussions and collaborations. Special thanks to Alain Dubus for taking care of me during my time in Liege and showing me the DAOCS assay. Many thanks also to Kirsty Hewitson for nice collaboration with ORF7 and guidance and help during my visit in Oxford.

Big thanks to all the present and former members in the structural biology unit at BMC! It is a great place to work at and mostly thanks to the wonderful people around. Thanks everyone for nice discussions and company over coffee breaks, lunches, parties and other occasions!

Janos, thanks for spirit, inspiration and new ideas, your ability to always see things from new perspectives is amazing. Many thanks to Diane – it has been nice to work with you, thanks for sharing all of your knowledge in molecular biology. Also many thanks to Martin for nice collaboration and travel-company at the EU meetings. Saeid, thanks for encouragement and pep-talks. Al, for many great and crazy ideas about everything from science to ice climbing. Ola, thanks for trying to keep order in the lab. Jenny, thanks for being such a great person, it was always fun to share the office with you and I miss you and our chats a lot. Anton, my other roommate, thanks for nice discussions and for being a real scientist. Tom, for playing wonderful music and for proofreading this thesis. Stefan, thanks for taking good care of the department. Tove, thanks for being a nice roommate in a dark and crowded room. Sara and Alexandra, for bringing some well needed girl-power to Janos group! Nic, thanks for your brilliant way of explaining physics stuff. Gunilla, for always being so friendly.

David, Erling and Remco, deserves a million thanks for invaluable help with computers. Special thanks to Alwyn for help with O during the building of cmcI, it was a lot of fun!

Malin, Rosie, Martin, Urszula, Susanna, Nisse, Andreas, Kenth - thanks for great company at the synchrotron trips and all the fun in Grenoble!

Alla@xray: Emma, Andrea, Ulla, Talal, Hasse, Tex, Karl, Seved, Anette, Jimmy, Fredrik, Louise, Gösta, Christer, Ellinor, Patrik, Margareta, Mats, Gunnar, Daniel, Anna, Lena, Jeff, Alina, Isabella, Calle, Torsten, Nina, Fredrik, Magnus, EvaLena, Jerry, Lars…

Tack till glada pub-gänget i Uppsala, jag kommer verkligen sakna våra kvällar på diverse nationer och andra påhitt. Gamla tjejkompisarna från Mariestad: Helena, Sofia, Helena, Åsa och Anna, det är alltid lika kul att träffa er.

Stort tack till hela tjocka släkten, för att ni bryr er och undrar hur det går med den där forskningen. Nu får ni gå på disputationsfest!

Eriks familj: Thomas, Ingrid, Helena & Jerker! Tack för att ni alltid får mig att känna mig som hemma.

Mamma, Pappa och Lillebror Jonas! Tack för att ni alltid finns där och tror på mig vad jag än tar mig för.

Erik, du är underbar! Tack för allt ditt stöd, uppmuntran och omtanke. Du är min stora glädje. ♥

Linda

Appendix A

Data collection and refinement statistics for the cmcI crystal structures described in chapter 5 and not included in paper IV. Quality checks performed on the structures using PROCHECK show good geometry and good agreement between model and data. The structures features the same cis-peptide and Ramachandran outliers as the structures presented in paper IV.

	cmcI-N160 NADH	**cmcI-D160 NADH**
Data collection		
Beam line of ESRF	ID14EH2	ID14EH2
Wavelength (Å)	0.933	0.933
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	93.7, 102.8, 182.9	95.1, 103.6, 183.0
Resolution range (Å)	91.3-2.50	91.3-2.60
outer shell (Å)	2.59-2.50	2.29-2.60
No. of observations	1543023	1010907
No. of unique reflections	62266	56812
Redundancy	10	7
$I/\sigma(I)$[a]	34 (2.8)	18 (2.4)
R-merge (%)[b]	7.4 (29)	8.0 (31.5)
Completeness (%)	98.5 (88.4)	98.6 (89.5)
Refinement		
Residues in model	1388	1385
Number of solvent molecules	713	263
Wilson B-factor ($Å^2$)	64	56
R_{cryst}[c]	21.1	22.3
R_{free}[c]	25.5	27.1
r.m.s.d from ideal geometry		
bond lengths (Å)	0.0010	0.009
bond angles (°)	1.23	1.20

[a] Number in parentheses is for the outer resolution shell.
[b] R-merge = $\Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I_i(hkl) \rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation and $\langle I_i(hkl) \rangle$ is the mean intensity of reflection *hkl*, respectively.
[c] $R = \Sigma_{hkl} | |F_o| - |F_c| | / \Sigma_{hkl} |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.